Sensors and Signals

Dr Graham Brooker

Graham Brooker

Sensors and Signals

Copyright © 2007

Australian Centre for Field Robotics

Rose St Building (J04)

University of Sydney, 2006

Australia

gbrooker@acfr.usyd.edu.au

All rights reserved. No part of this book may be reproduced or utilised in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system without permission from the author, except for bona fide academic purposes

Table of Contents

Sensors and Signals ..... **i**

Chapter 1.Introduction **1**
1.1. Sensors and Transducers 1
1.2. Frequency Band Allocations for the Electromagnetic Spectrum 3
1.3. Frequency Band Allocations for the Acoustic Spectrum 6
References 6

Chapter 2.Signal Processing and Modulation **7**
2.1. The Nature of Electronic Signals 7
2.2. Sinusoidal Signals 7
2.3. The Fourier Series 9
2.4. Sampled Signals 11
2.5. Filtering 14
2.6. Analog Modulation and Demodulation 19
2.7. Frequency Modulation (FM) 21
2.8. Linear Frequency Modulation 25
2.9. Pulse Coded Modulation Techniques 27
2.10. Frequency Shift Keying (FSK) 30
2.11. Phase Shift Keying (PSK) 32
2.12. Stepped Frequency 33
2.13. Convolution 34
2.14. References 39

Chapter 3.Radiometers and Image Intensifiers **41**
3.1. Introduction 41
3.2. Thermal Emission 41
3.3. Emissivity and Reflectivity 45
3.4. Detecting Thermal Radiation 48
3.5. Heating 51
3.6. Performance Criteria for Detectors 51
3.7. Noise Processes and Effects 54
3.8. Applications 55
3.9. Introduction to Thermal Imaging Systems 58
3.10. Scanning Mechanisms and Arrays 60
3.11. Key Optical Parameters 62
3.12. System Performance Measures 64
3.13. Target Detection and Recognition 65
3.14. Example of FLIR Detection 66
3.15. Thermal Images 68
3.16. Image Intensifiers 69
3.17. References 75

Chapter 4.Millimetre Wave Radiometers **77**
4.1. Thermal Emission 77
4.2. Antenna Power Temperature Correspondence 77
4.3. Brightness Temperature 79
4.4. Apparent Temperature 80

4.5. Atmospheric Effects....81
4.6. Terrain Brightness....83
4.7. Example....84
4.8. Antenna Considerations....85
4.9. Receiver Considerations....86
4.10. The System Noise Temperature....88
4.11. Radiometer Temperature Sensitivity....89
4.12. Radiometer Implementation....92
4.13. IF and Video Gains....94
4.14. Target Seeker Design Example....94
4.15. Airborne Push-Broom Scanner....100
4.16. Radiometric Applications....102
4.17. References....110

Chapter 5.Active Ranging Sensors....111
5.1. Overview....111
5.2. Principle of Operation....112
5.3. Pulsed Range Measurement....123
5.4. Other Methods to Measure Range....129
5.5. The Radar Range Equation....135
5.6. Issues with TOF Measurement....140
5.7. Examples of Range Measurement Radar....141
5.8. References....145

Chapter 6.Active Imaging Sensors....147
6.1. Imaging Techniques....147
6.2. Laser Radar Performance....150
6.3. Digital Terrain Models....158
6.4. Laser Based Sea-Bottom Profiling....161
6.5. 2D Laser Scanners....162
6.6. 3D laser Scanners....163
6.7. 3D Millimetre Wave Radar Scanners....164
6.8. Scanning Acoustic Microscopes....165
6.9. References....169

Chapter 7.Range Measurement Applications....171
7.1. Introduction....172
7.2. Acoustic Level Measurement....172
7.3. Acoustic Systems....177
7.4. Short Range Radar Level Measurement....178
7.5. Long Range Radar Level Measurement....183
7.6. Laser Level Measurement....184
7.7. Industrial Laser Ranging Systems....189
7.8. Recreational Laser Ranging Systems....190
7.9. Selection of the Correct Sensor....191
7.10. Short Range Sensors....192
7.11. Orepass Radar Development: Case Study....197
7.12. Prototype Build and Test....206
7.13. References....208

Chapter 8.Signal Propagation....209
8.1. The Sensing Environment....209

8.2. Propagation of Electromagnetic Waves 211
8.3. Attenuation of Electromagnetic Waves 212
8.4. Refraction of Electromagnetic Waves 223
8.5. Acoustics and Vibration 224
8.6. Attenuation of Acoustic Waves in Water 233
8.7. Reflection and Refraction of Acoustic Waves 235
8.8. Multipath Effects 239
8.9. References 246

Chapter 9. Target and Clutter Characteristics 247
9.1. Target Cross Section 247
9.2. Radar Cross Sections (RCS) 249
9.3. RCS of Simple Shapes 250
9.4. RCS of Complex Targets 254
9.5. RCS of Living Creatures 258
9.6. Fluctuations in RCS 260
9.7. Radar Stealth 262
9.8. Stealth and Frequency Allocation 265
9.9. Complex Targets at IR and Visible Frequencies 266
9.10. Complex Acoustic Targets 268
9.11. Clutter 271
9.12. Surface Clutter 271
9.13. Calculating Surface Clutter Backscatter 275
9.14. Calculating Volume Backscatter 276
9.15. Estimating the Backscatter of Radar Signals 280
9.16. References 282

Chapter 10. Detection of Targets in Noise 283
10.1. Receiver Noise 283
10.2. Noise Probability Density Functions 284
10.3. Probability of False Alarm 285
10.4. Probability of Detection 286
10.5. Detector Loss Relative to an Ideal System 288
10.6. The Matched Filter 289
10.7. Integration of Pulse Trains 290
10.8. Detection of Fluctuating Signals 293
10.9. Constant False Alarm Rate (CFAR) Processors 295
10.10. Air Traffic Control Radar Performance 298
10.11. References 302

Chapter 11. High Range Resolution Techniques 303
11.1. Classical Modulation Techniques 303
11.2. Amplitude Modulation 303
11.3. Frequency & Phase Modulation 305
11.4. Phase-Coded Pulse Compression 307
11.5. SAW Based Pulse Compression 314
11.6. Step Frequency 317
11.7. Frequency Modulated Continuous Wave Radar 317
11.8. Stretch 329
11.9. Interrupted FMCW 329
11.10. Sidelobes and Weighting for Linear FM Systems 332
11.11. FMCW Radar Systems 334

11.12. Application: Brimstone Antitank Missile 336
11.13. Application: Ground Penetrating Radar 347
11.14. Application: 2D Medical Ultrasound 353
11.15. References 356

Chapter 12. High Angular Resolution Techniques 357
12.1. Introduction 357
12.2. Phased Array Synthesis 357
12.3. The Radiation Pattern 363
12.4. Beam Steering 366
12.5. Array Characteristics 368
12.6. Applications 370
12.7. Phased Array Application: Side-Scan Sonar 372
12.8. Other Applications 383
12.9. Phased Array Image Gallery 384
12.10. Doppler Beam-Sharpening 385
12.11. Operational Principles of Synthetic Aperture 387
12.12. Range and Cross Range Resolution 388
12.13. Example: Synthetic Aperture Sonar 393
12.14. Radar Image Quality Issues 397
12.15. Airborne SAR 399
12.16. Space Based SAR 403
12.17. Magellan SAR Map of Venus 407
12.18. References 407

Chapter 13. Range and Angle Tracking 409
13.1. Introduction 409
13.2. Range Estimation and Tracking 409
13.3. Range Tracking Systems 419
13.4. Angle Measurement 422
13.5. Angle Tracking Principles 423
13.6. Lobe Switching (Sequential Lobing) 424
13.7. Conical Scan 425
13.8. Amplitude Comparison Monopulse 431
13.9. Comparison between Conscan and Monopulse 434
13.10. Angle Tracking Loops 435
13.11. Angle Estimation and Tracking Applications 435
13.12. Triangulation 439
13.13. References 441

Chapter 14. Doppler Measurement 443
14.1. The Doppler Shift 443
14.2. Doppler Geometry 445
14.3. Doppler Frequency Extraction 447
14.4. Pulsed Doppler 452
14.5. Doppler Sensors 454
14.6. References 461

Chapter 15. Tracking Moving Targets 463
15.1. Track While Scan (TWS) 463
15.2. The Coherent Pulsed Tracking Radar 465
15.3. Limitations to MTI Performance 470

15.4. Range Gated Pulsed Doppler Tracking....470
15.5. Co-ordinate Frames....472
15.6. Antenna Mounts and Servo Systems....473
15.7. Fire Control Radar Design....480
15.8. References....499

Chapter 16.Radio Tags and Transponders....501
16.1. Principle of Operation....501
16.2. History and Applications....501
16.3. RFID Systems....502
16.4. Technical Challenges....510
16.5. Battlefield Combat ID System (BCIS)....510
16.6. References....511

Chapter 17.Tomography and 3D Imaging....513
17.1. Principle of Operation....513
17.2. CT Imaging....514
17.3. MRI-Magnetic Resonance Imaging....521
17.4. MRI Images....525
17.5. Functional MRI (fMRI) Investigations of Brain Function....526
17.6. Positron Emission Tomography (PET)....527
17.7. 3D Ultrasound Imaging....529
17.8. 3D Sonar Imaging....533
17.9. 3D Ground Penetrating Radar....535
17.10. References....536

Chapter 1.
Introduction

1.1. Sensors and Transducers

Sensors in the natural world include those which equip us with our five senses – sight, hearing, smell, taste and touch. These convert the various and diverse inputs to electrochemical signals that can be used to inform or control the living organism.

In a similar way, in man-made devices, sensors are also used to measure various stimuli. However, because of the broad range of potential inputs and outputs, the accepted definition of a sensor is refined. In this definition, all devices that convert input energy into output energy are referred to as *transducers*, and sensors form a small subset of the group as defined below:

"A sensor is a transducer that receives a input signal or stimulus and responds with an electrical signal bearing a known relationship to the input " [2]

Many measuring and sensing devices, as well as loudspeakers, thermocouples, microphones, and phonograph pickups, may be termed transducers.

Systems of sensors and transducers are constructed for a variety of applications including surveillance, imaging, mapping and target tracking. In some cases, the sensors provide their own source of *illumination* and are referred to as active sensors. Passive sensors, on the other hand, do not provide illumination and depend on variations of natural conditions for detection.

1.1.1. Active Sensors

Active sensors require the application of external power for their operation. This *excitation signal* is modified by the sensor to produce an output signal. They are restricted to frequencies that can be generated and radiated fairly easily. This excludes part of the far infrared (above 3×10^{12}Hz), parts of the ultraviolet band and the gamma ray region. However, inroads are being made into these regions with the development of Terahertz sources and detectors based on artificial photonic crystals.

Semi-active sensors use a man-made excitation signal generated by (or radiated from) a source that is not coincident with the sensor. Such systems include multistatic radar that

uses mobile phone or commercial broadcasts as the source and are proving to be particularly effective in detecting stealth aircraft.

1.1.2. Passive Sensors

Passive sensors directly generate an electric signal in response to a stimulus. That is, the input stimulus is converted by the sensor into output energy without the need for an additional source of power to illuminate the environment. The salient characteristics of most passive sensors are as follows:

- Do not emit radiation (or an excitation signal) as part of the measurement process.
- Rely on a locally generated or natural source of radiation (light from the sun), an available energy field (gravity) or even a chemical gradient
- Passive sensors can exploit EM radiation of any frequency in which some natural phenomenon radiates and for which a detection mechanism exists. This can extend from ELF (below 3×10^{3}Hz) up to gamma rays (above 3×10^{19}Hz).
- They can exploit acoustic energy (vibration) from infrasound frequencies <1Hz from earthquakes or explosions up to the ultrasound

Table 1.1:Advantages of active and passive sensors

PASSIVE	ACTIVE
Low power requirement (long life) Cannot be detected (covert) Simple principles (sometimes) Good reliability due to simplicity Field of view constrains observation High angular resolution possible Large variety	Measurement requirements matched to transmitter characteristics Range measurement by temporal correlation Radiation pattern constrains observation High range and angle resolutions possible Long range operation possible

Table 1.2: Disadvantages of active and passive sensors

PASSIVE	ACTIVE
Targets of interest must radiate or modify the field (electrical, gravity) Prone to feature ambiguity and errors of scale (eg angle only) Typically short range, though not always (astronomy etc) Availability not guaranteed (no light, contrast etc)	Large power requirement Easy to detect (not covert) Complex Rx and Tx processes Reliability determined by two elements, Rx and Tx

1.1.3. Sensor Systems

Sensors do not operate in isolation, they are always part of a larger system which may include other transducers, signal processing networks and often some from of actuation or data recording capability. An autonomous vacuum cleaner is a good example of a typical system.

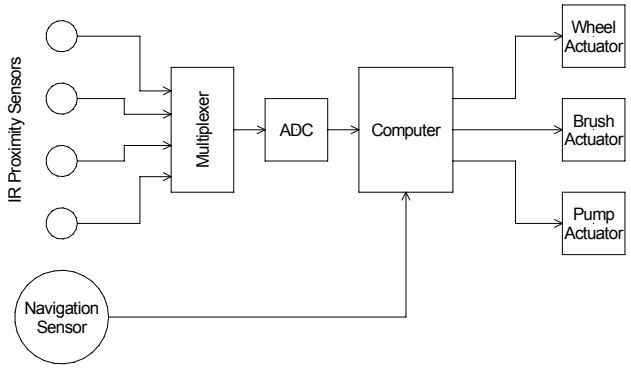


Figure 1.1: Sensors and actuators in an autonomous vacuum cleaner. The IR proximity sensors are active, while the navigation sensor is passive and relies on the earth's magnetic and gravity fields

1.2. Frequency Band Allocations for the Electromagnetic Spectrum

Sensors can be made to operate over a very broad band of frequencies for both electromagnetic (EM) and acoustic applications.

In theory, though the EM spectrum extends below one Hz, in practise, radiation, detection and data-rate limitations at such low frequencies is impractical. An existing extra low frequency (ELF) application is the communication with submerged submarines at about 80Hz (a wavelength of 3750km).

At the other extreme, Positron Emission Tomography (PET) scanners generate and detect gamma rays with an energy of 511keV, equivalent to a frequency of 1.23×10^{20} Hz or a wavelength of only 2.43×10^{-12} m.

As a matter of convenience, the EM spectrum is broken up into overlapping bands.

- Gamma rays
- X-rays
- UV
- Visible
- Infra red
- Microwave
- Radio

Many of these general bands are further subdivided to convey more information, as with near and far infrared, or from an historical motive as with the radar band designations.

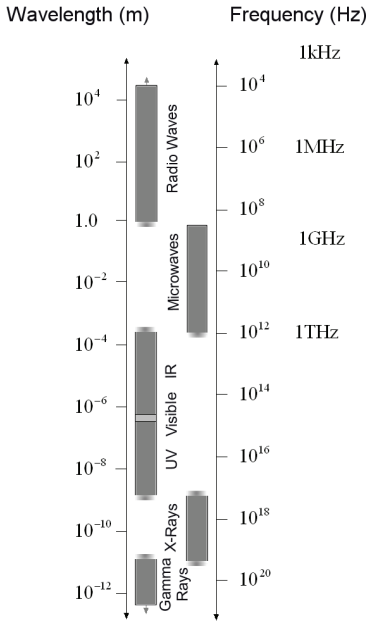


Figure 1.2: The Electromagnetic Spectrum

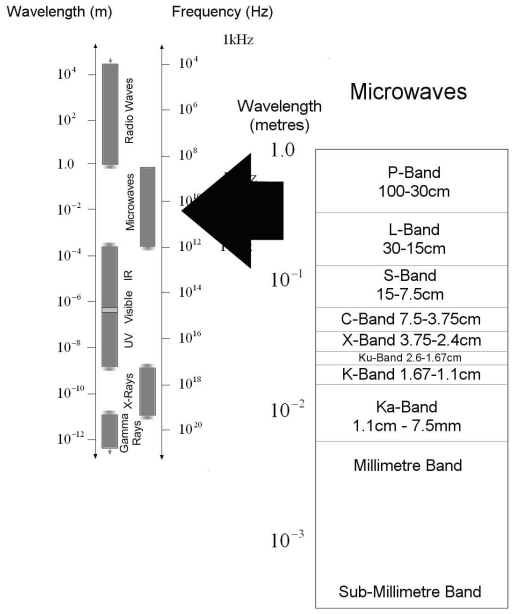


Figure 1.3: The Microwave Band

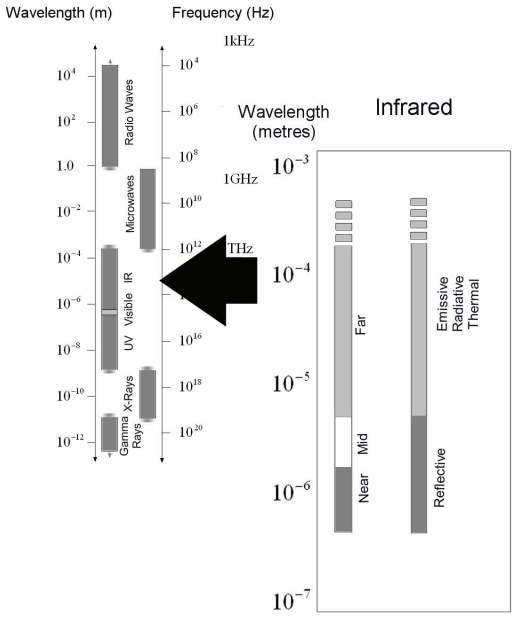


Figure 1.4: The Infrared Band

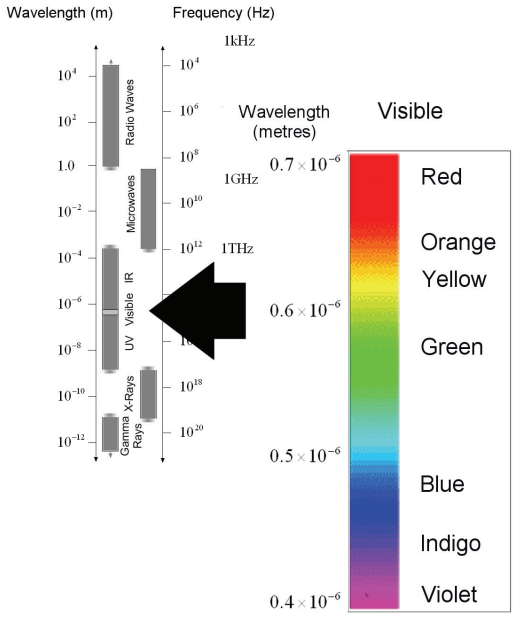


Figure 1.5: The Visible Band

The subdivision of the frequency allocations in the microwave band can be confusing as there are various different standards in use. In this book, the United States Microwave and radar nomenclature is used.

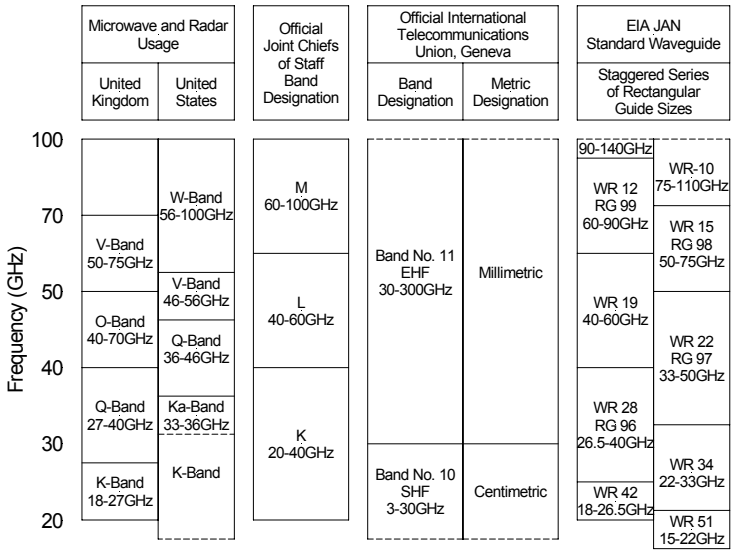


Figure 1.6: Nomenclature for Electromagnetic Spectrum Designation [3]

Useful EM radiation spans 16 orders of magnitude, from the ELF band below 3×10^{3}Hz up to Gamma Rays above 3×10^{19}Hz.

Table 1.3: Frequency bands and typical applications for electromagnetic systems

Band	**Frequency**	**Wavelength**	**Applications**
VHF	30-300MHz	1m-10m	Over the Horizon
UHF	300-1000MHz	30-100cm	Ground penetrating
L	1-2GHz	15-30cm	Ground Surveillance, Astronomy
S	2-4GHz	7.5-15cm	Ground Surveillance
C	4-8GHz	3.75-7.5cm	Space SAR
X	8-12.5GHz	2.4-3.75cm	Fire Control , Proximity, Airborne/Space SAR
Ku	12.5-18GHz	16.7-24mm	Collision Avoidance, Speed Traps
K	18-26.5GHz	11.3-16.7mm	Fire Control Radar
Ka	26.5-40GHz	7.5-11.3mm	FCR, Surveillance
Millimetre	30-300GHz	1-10mm	Imaging & astronomy, Collision Avoidance
Sub-mm		50μm-1mm	Astronomy
Far infrared		14-50μm	Properties of molecules
Longwave IR		8-14μm	Laser radar, Forward looking IR
Near IR		1-3μm	Personnel Detection
Very near IR		0.76-1	Imaging, Laser ranging (industrial)
Visible		380-760nm	Imaging, Astronomy
UV		100-380nm	Missile plume detection, Gas fire detection

1.3. Frequency Band Allocations for the Acoustic Spectrum

Useful acoustic signals span eight orders of magnitude from the lower infrasound at below 1Hz for earthquake and structural vibration measurements up to extreme ultrasound at about 1GHz for microscopy.

Table 1.4: Frequency bands and typical applications for acoustic systems

Designation	Frequency Range	Applications
Infrasound	Below 50Hz	Weather monitoring, earthquakes, nuclear explosion detection
Audio	50Hz to 20kHz	Audio communications, acoustic modem, range measurement, hydro-acoustic positioning, seismic prospecting
Ultrasound	20kHz to 200kHz 200kHz to 500kHz ≈1GHz	Range measurement, Imaging Underwater Sonar Microscopy

1.4. References

[1] Joseph J. Carr, *Electronic Circuit Guidebook: Sensors,* Prompt, 1997
[2] Jacob Fraden, *Handbook of Modern Sensors 2nd ed.,* Air Press, 1996
[3] *Jane's Radar and Electronic Warfare Systems, 1995-1996*

Chapter 2.

Signal Processing and Modulation

2.1. The Nature of Electronic Signals

2.1.1. Static and Quasi-Static Signals

Static signals are by definition unchanging over a long period of time. Such signals are essentially DC levels, while quasi-static signals are those that change very slowly such as the drift on a sensor.

2.1.2. Periodic and Repetitive Signals

Periodic signals are those that repeat themselves on a regular basis. These include sine, square and sawtooth waves. Their nature is such that each waveform is identical.

Repetitive signals are periodic in nature, but the exact shape may change slightly with time. ECG signals are an example of this type.

2.1.3. Transient and Quasi Transient Signals

Transient signals are either one time only signals while quasi-transient signals are those which are periodic but with a duration which is very short compared to the period of the waveform. Pulsed radar signals are good examples of these.

2.2. Sinusoidal Signals

Most acoustic and electromagnetic sensors exploit the properties of sinusoidal signals

In the time domain, such signals are constructed of sinusoidally varying voltages or currents constrained within wires,

$$v_c(t) = A_c \cos\omega_c t = A_c \cos 2\pi f_c t \,, \tag{2.1}$$

where $v_c(t)$ – Signal,
A_c – Signal amplitude (V),
ω_c – Frequency (rad/s),
f_c – Frequency (Hz),
t – Time (s).

Sinusoidal electrical signals can be generated by the appropriate frequency selective feedback (shown below) or by feedback across an inductive-capacitive tank circuit.

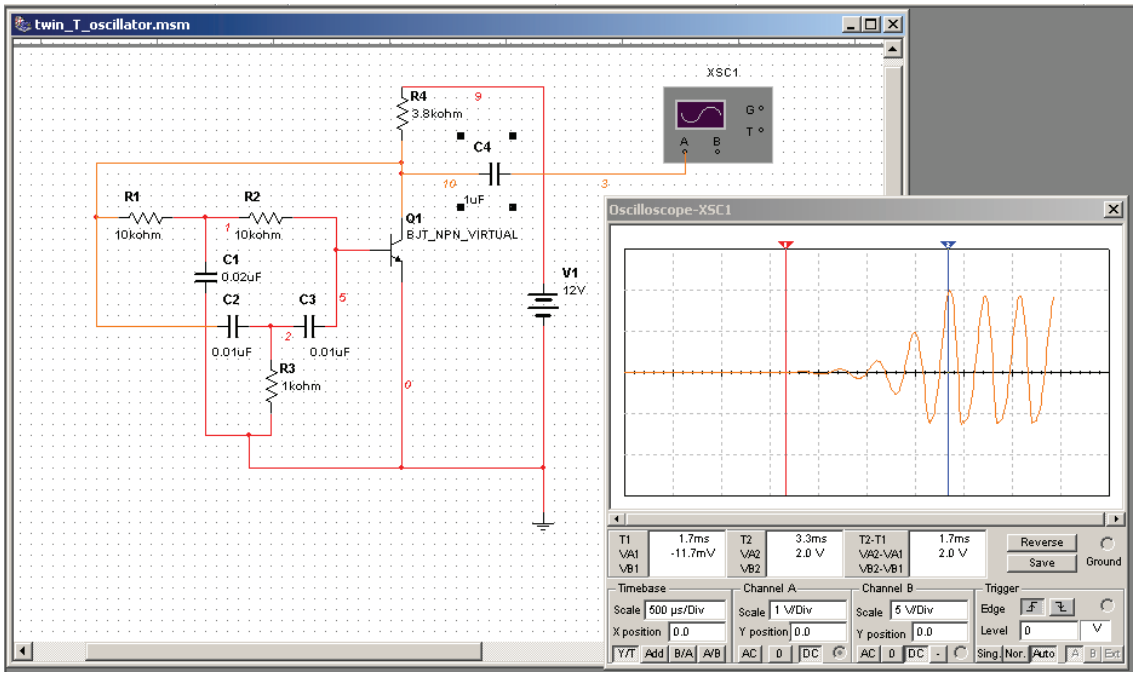


Figure 2.1: Sinusoidal voltage signal generated by an oscillator

In the frequency domain, a continuous sinusoidal signal of infinite duration can be represented in terms of its position on the frequency continuum and its amplitude only.

Most practical signals are not of infinite duration and so there is some uncertainty in the measured frequency which is represented in the frequency domain by a finite spectral width.

From a mathematical perspective, this is equivalent to windowing the continuous sinusoidal signal using a rectangular pulse of duration τ. Because windowing, or multiplication, in the time domain becomes convolution in the frequency domain, the continuous signal spectrum must be convolved by the spectrum, or Fourier transform, of a rectangular pulse to obtain the spectrum of the windowed signal.

As this Fourier transform is the Sync function

$$F(\omega) = \tau \frac{\sin(\omega\tau/2)}{\omega\tau/2} \tag{2.2}$$

and the spectrum of a continuous sinusoidal signal is an impulse $\delta(\omega)$, the resultant convolution is just the Sync function.

It can be seen from the equation for the Sync function that as the duration of the signal decreases, $\tau \to 0$, its spectral width increases until, in the limit, when the signal can be represented by an impulse $\delta(t)$, the spectral width is infinite. This relationship is shown graphically in the following figure.

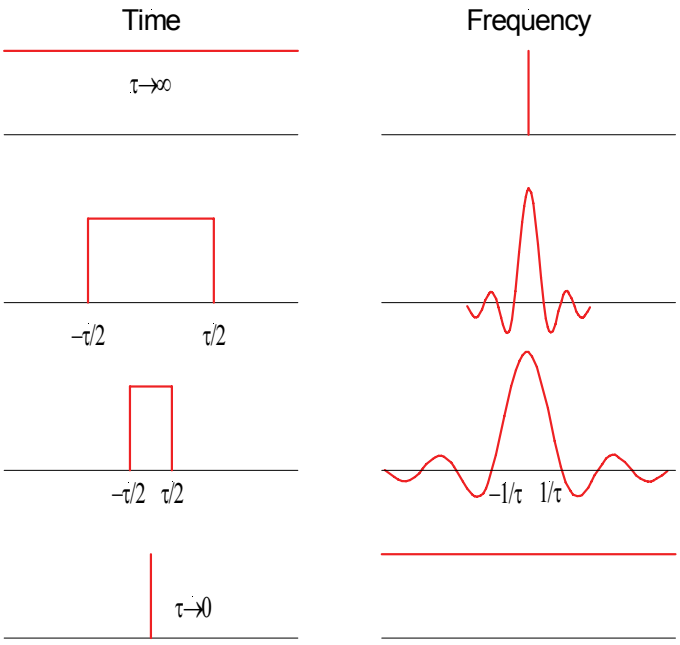


Figure 2.2: Mapping the relationship between the duration of a pulse and its spectrum

2.3. The Fourier Series

All continuous periodic signals can be represented by a fundamental frequency sine wave and a collection of sine and/or cosine harmonics of that fundamental sine wave.

The Fourier series for any waveform can be expressed as

$$v(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(n\omega t) + \sum_{n=1}^{\infty} b_n \sin(n\omega t), \tag{2.3}$$

where: a_n b_n - Amplitudes of the harmonics (can be zero),
n - Integer.

The amplitude coefficients can be calculated as follows,

$$\begin{aligned} a_n &= \frac{2}{T}\int_0^t v(t)\cos(n\omega t)dt \\ b_n &= \frac{2}{T}\int_0^t v(t)\sin(n\omega t)dt \end{aligned} \tag{2.4}$$

The amplitude terms are non-zero at specific frequencies determined by the Fourier series. Because only certain frequencies, determined by integer n, are allowed, the spectrum is discrete.

The term $a_0/2$ is the average value of $v(t)$ over a complete cycle.

In general, though the harmonic series is infinite, the coefficients become so small that their contribution is considered to be negligible.

An ECG trace, for example with a fundamental frequency of about 1.2Hz can be reproduced with 70 to 80 harmonics (a bandwidth of about 100Hz)

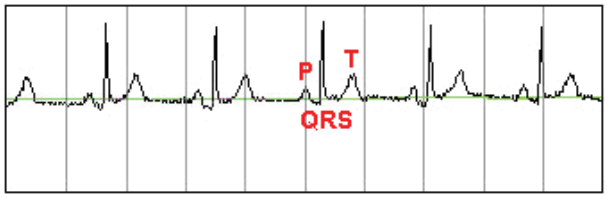


Figure 2.3: Typical ECG trace

A square wave on the other hand may require up to 1000 harmonics to reproduce the sharp transitions as shown in the figure below, depending on the application.

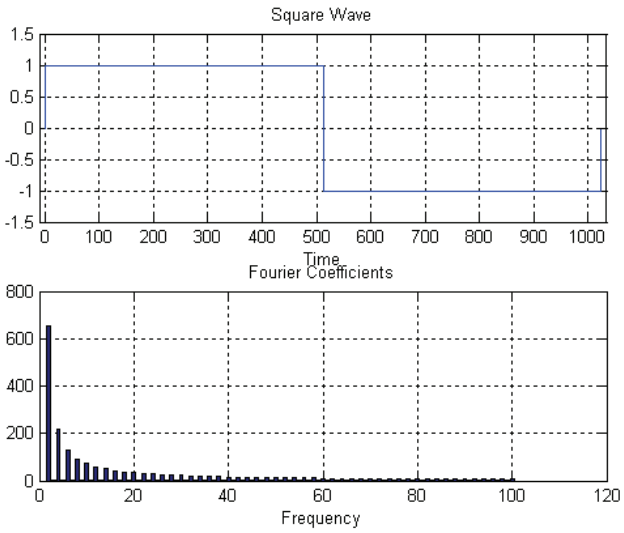


Figure 2.4: Amplitudes of Fourier coefficients to produce a square wave

If insufficient Fourier coefficients are used in the reconstruction of the signal it will be distorted as shown in the following example.

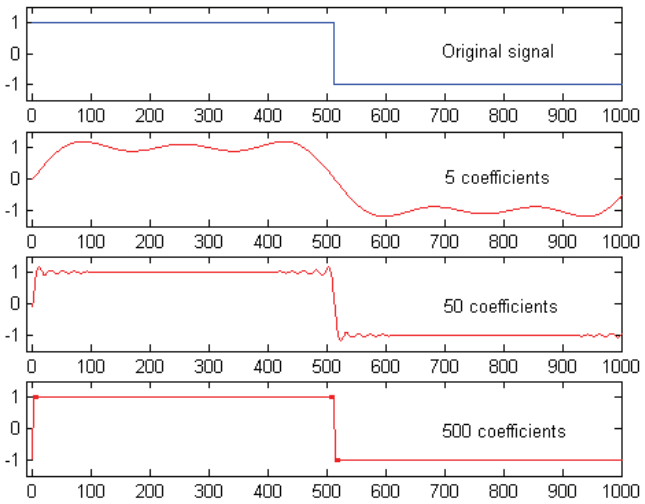


Figure 2.5: Effect on the reconstructed signal of limiting the number of Fourier coefficients

2.4. Sampled Signals

To process signals within a computer (Digital Signal Processing) requires that they be sampled periodically and then converted to a digital representation using an Analog to Digital Converter (ADC).

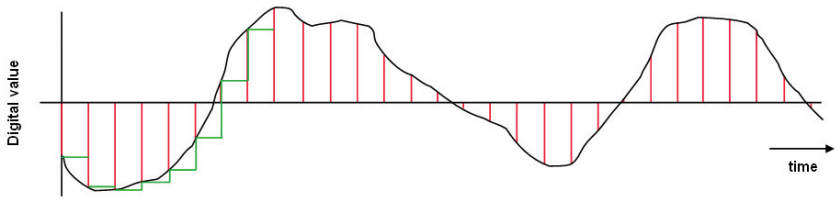


Figure 2.6: Digitising a signal

To ensure accurate representation the signal must be sampled at a rate which is at least double the highest significant frequency component of the signal. This is known as the Nyquist rate.

In addition, the number of discrete levels to which the signal is quantised must also be sufficient to represent variations in the amplitude to the required accuracy. Most ADCs quantise to 12 or 16 bits which represent 2^{12} = 4096 or 2^{16} = 65536 discrete levels.

After the signal has been processed, it is often necessary to generate an analog output. This function is performed by a Digital to Analog Converter (DAC)

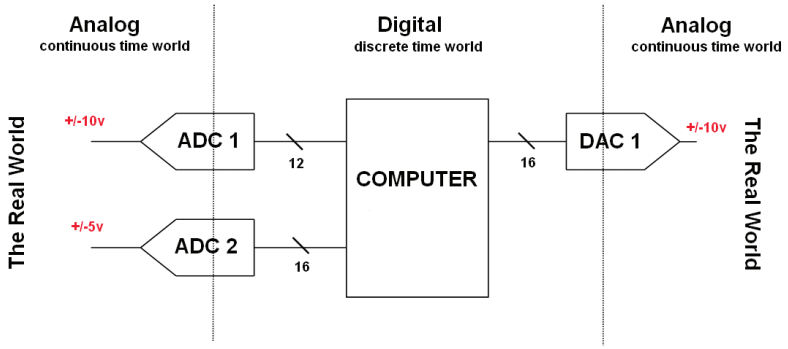


Figure 2.7: Typical Configuration for a digital signal processing application

The reconstruction process generally involves holding the signal constant (zero order hold) during the period between samples as shown in the following figure. This signal is then cleaned up by passing it through low-pass filter to remove high frequency components generated by the sampling process.

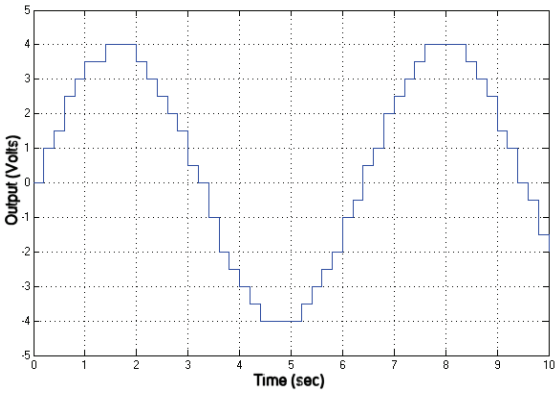


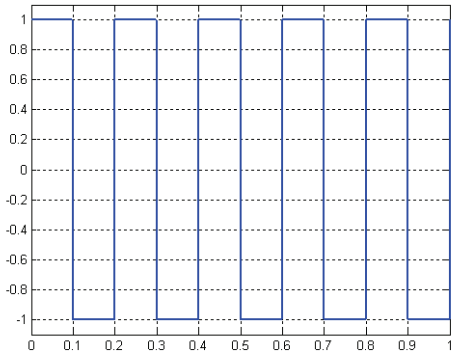
Figure 2.8: Analog Reconstruction of a sampled signal using a zero order hold

2.4.1. Generating Signals in MATLAB

MATLAB includes a number of built-in functions that make it easy to generate both periodic and aperiodic signals.

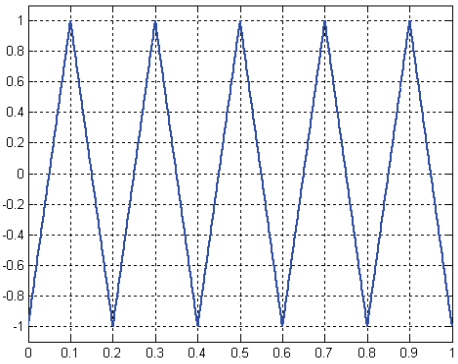
Generation of a square wave with amplitude A, fundamental frequency w0 (rad/s) and duty cycle rho as a percentage.

```
% generate a square wave
A = 1;
w0 = 10*pi;
rho = 50;
t = 0:0.001:1;
sq = A*square(w0*t, rho);
plot(t,sq)
axis([0,1,-1.1,1.1]);
```



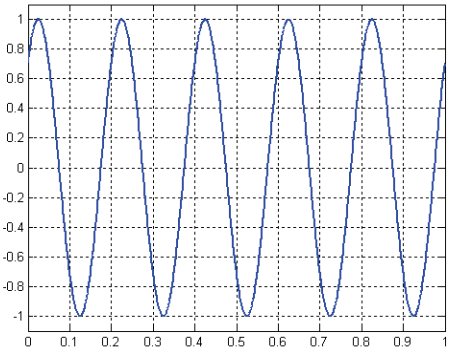
Generation of a triangular wave with amplitude A, fundamental frequency w0 (rad/s) and width W.

```
% generate a triangular wave
A = 1;
w0 = 10*pi;
W = 0.5;
t = 0:0.001:1;
tri = A*sawtooth(w0*t, W);
plot(t,tri)
grid
axis([0,1,-1.1,1.1]);
```



Generation of a sine wave with amplitude A, fundamental frequency w0 (rad/s) and start phase angle phi (rad).

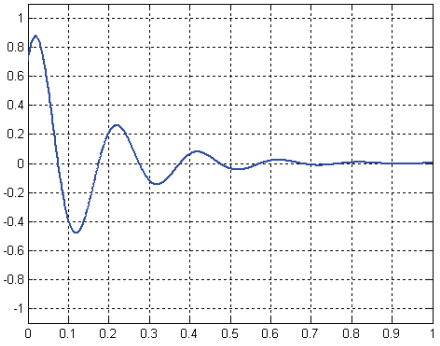
```
% generate a sine wave
A = 1;
w0 = 10*pi;
phi = pi/4;
t = 0:0.001:1;
sine = A*sin(w0*t + phi);
plot(t,sine)
grid
axis([0,1,-1.1,1.1]);
```



An exponentially damped sine wave is easily generated by taking the product of an exponential function and a sine wave.

$$x(t) = A\sin(\omega_o t + \phi)e^{-at} \tag{2.5}$$

```
% generate a exponentially
% decaying sine wave
A = 1;
w0 = 10*pi;
phi = pi/4;
a = 6;
t = 0:0.001:1;
expsine = A*sin(w0*t + phi).*exp(-a*t);
plot(t,expsine)
grid
axis([0,1,-1.1,1.1]);
```



Other useful MATLAB functions include the following:, COS, CHIRP, DIRAC, GAUSPULS, PULSTRAN, RECTPULS, SINC and TRIPULS.

2.4.2. Aliasing

If the analog signal is not sampled at at least twice the frequency of the highest frequency component, then these high frequency signals are “aliased” down to a lower frequency as shown in the following figure.

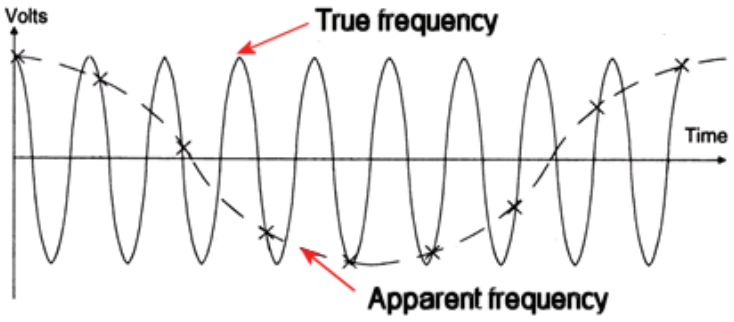


Figure 2.9: Interpretation of aliasing effects in the time domain

In the frequency domain, a generic analog signal may be represented in terms of its amplitude and total bandwidth as shown in the figure below.

A sampled version of the same signal can be represented by a repeated sequence spaced at the sample frequency (generally denoted f_s).

If the sample rate is not sufficiently high, then the sequences will overlap, and high frequency components will appear at a lower frequency (albeit with reduced amplitude) as shown in the figure.

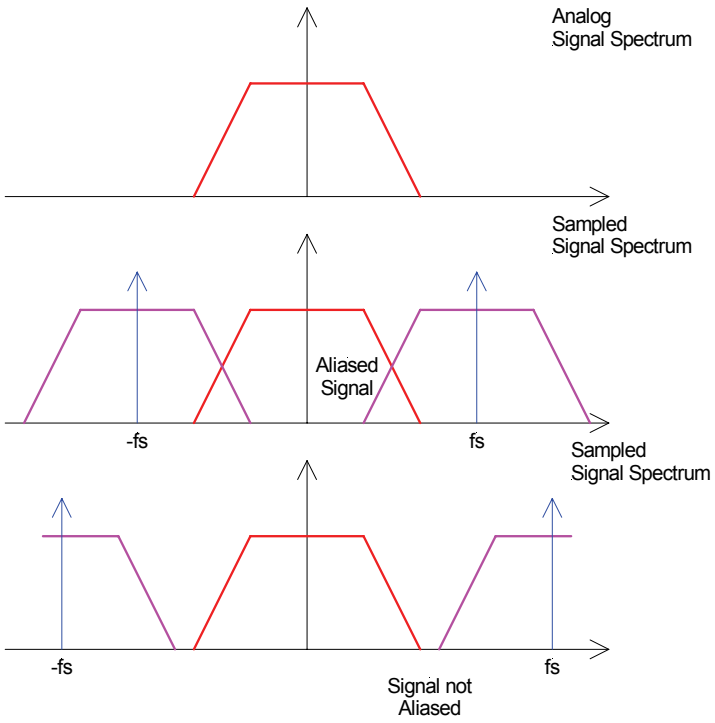


Figure 2.10: Interpretation of aliasing effects in the frequency domain

In most applications, an anti aliasing (low pass) filter ensures that the high frequency signals are sufficiently attenuated prior to sampling. A typical filter will attenuate these unwanted signals by between 40 and 60dB (1/100 to 1/1000) in voltage.

2.5. Filtering

A filter is a frequency selective network that passes certain frequencies of an input signal and attenuates others

The three common types of filter are:

- High Pass
- Low Pass
- Band Pass

High pass filter blocks signals below its cutoff frequency and passes those above.

Low pass filter passes signals below its cutoff frequency and attenuates those above.

Band pass filter passes a range of frequencies while attenuating those both above and below that range.

A fourth, less common configuration, is a band-stop or notch filter that attenuates signals at a specific frequency or over a narrow range of frequencies and passes all other frequencies.

Filters can be characterised by their impulse responses in the time domain, but are usually characterised by their frequency domain amplitude response $|H(\omega)|$ or Gain which presents the ratio of the output to input voltage over the frequency range of interest.

In this course we will only use sampled data filters that can be synthesized in MATLAB as shown in the figure.

```
% Banspass filter
fs = 200e+03;
ts = 1/fs;
fmat = 40.0e+03;
bmat = 10.0e+03;
wl=2*ts*(fmat-bmat/2);% lower band
wh=2*ts*(fmat+bmat/2);% upper band
wn=[wl,wh];
% 6th order Butterworth filter
[B,A]=butter(3,wn);
[h,w]=freqz(B,A,1024);
freq=(0:1023)/(2000*ts*1024);
%semilogx(freq,20*log10(abs(h)));
plot(freq,abs(h));
grid
title('Bandpass Filter Transfer Fn')
xlabel('Frequency (kHz)');
ylabel('Gain')
```

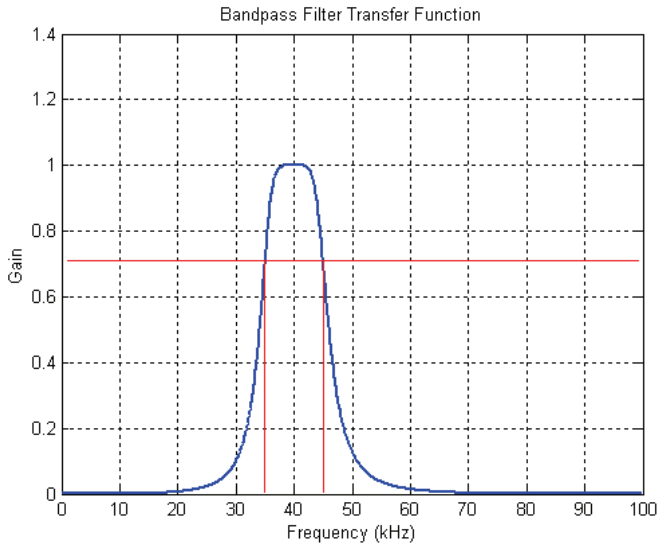


Figure 2.11: Butterworth bandpass filter transfer function generated by MATLAB

Note that the upper and lower limits of the pass band represent the half power points (0.707 of the peak voltage gain)

If the gain was plotted in dB then it would be calculated using $20*\log_{10}(\text{Gain})$ to convert to power.

MATLAB implements filters as the "Direct Form II Transposed" of the standard difference equation

$$a(1)*y(n) = b(1)*x(n) + b(2)*x(n-1) + ... + b(n_b+1)*x(n-n_b)$$
$$-a(2)*y(n-1) - ... - a(n_a+1)*y(n-n_a)$$

where the coefficients A =[a(1), a(2)…a(n_a+1)] and B = [b(1), b(2) …b(n_b+1)] are generated when the filter is synthesised

2.5.1. Filter Categories

The major filter categories are as follows:

- Butterworth (maximally flat).
- Chebyshev (equi ripple).
- Bessel (linear phase).
- Elliptical

Butterworth

This approximation to an ideal low pass filter is based on the assumption that a flat response at zero-frequency is most important. The transfer function is an all-pole type with roots that fall on the unit circle.

It exhibits fairly good amplitude and transient characteristics.

Chebyshev

The transfer function is also all-pole, but with roots that fall on an ellipse. This results in a series of equi amplitude ripples in the pass band and a sharper cutoff than the Butterworth.

It exhibits good selectivity but poor transient behaviour.

Bessel

Optimised to obtain a linear phase response which results in a step response with no overshoot or ringing and an impulse response with no oscillatory behaviour.

Poor frequency selectivity compared to the other response types.

Elliptic

These filters have zeros as well as poles which create equi-ripple behaviour in the pass band similar to Chebyshev filters.

Zeros in the stop band reduce the transition region so that extremely sharp roll-off characteristics can be achieved, for that reason they are commonly used in anti-aliasing filters.

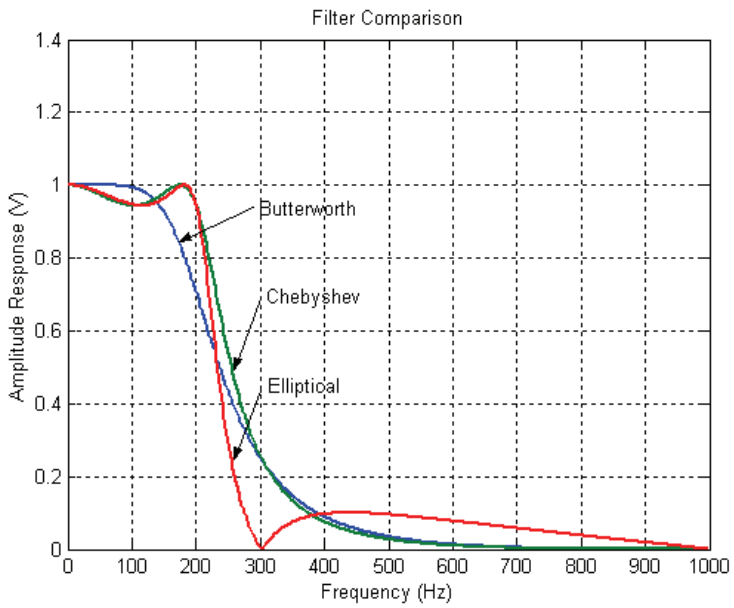


Figure 2.12: Lowpass filter comparison

Note that the cutoff frequency (200Hz in this case) specified in MATLAB is equal to the 3dB point for the Butterworth filter and to the passband ripple for the Chebyshev and Elliptic filters

The rate at which the signal is attenuated as a function of frequency is proportional to the order of the filter. The following figure shows the roll-off for Butterworth filters. The theoretical slope is 6n dB/octave.

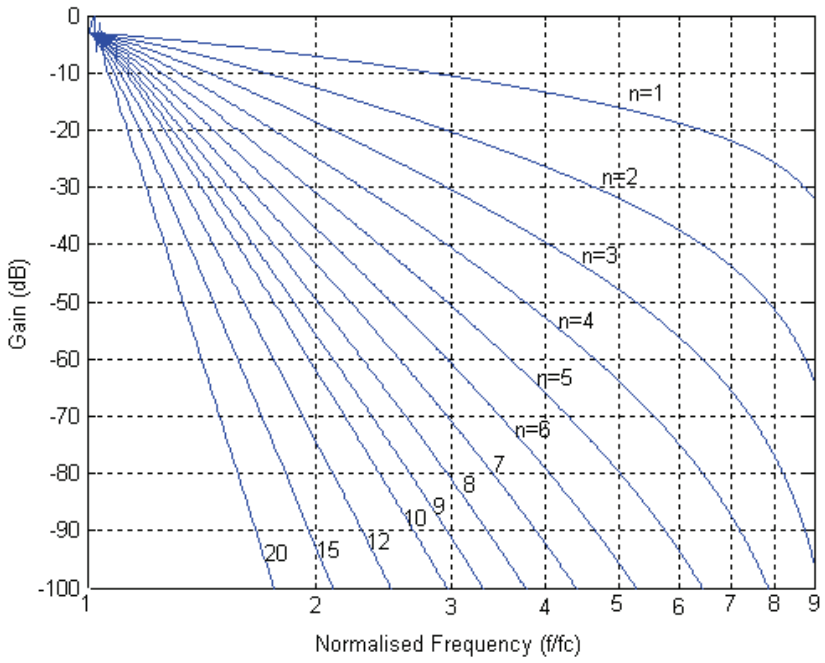


Figure 2.13: Roll off for Butterworth lowpass filters

2.5.2. The Ear as a Filter Bank

In the ear, sound waves are transmitted into the cochlea which tapers in size like a cone. Through the middle stretches the basilar membrane which gets wider as the cochlea gets narrower.

The vibratory movement is transmitted as a standing wave in the basilar membrane (much like a snapping rope) with the amplitude reaching a peak at a location dependent on frequency due to the varying resonant characteristics of the membrane as shown schematically in the following figure.

High frequency peaks occur toward the base (where the membrane is stiffest and narrowest) while the low frequency peaks occur towards the apex. Hair cells rest on the basilar membrane and convert these vibrations to electro-chemical signals which are transmitted to the brain for interpretation.

The base resonates at about 20kHz while the apex resonates at about 20Hz

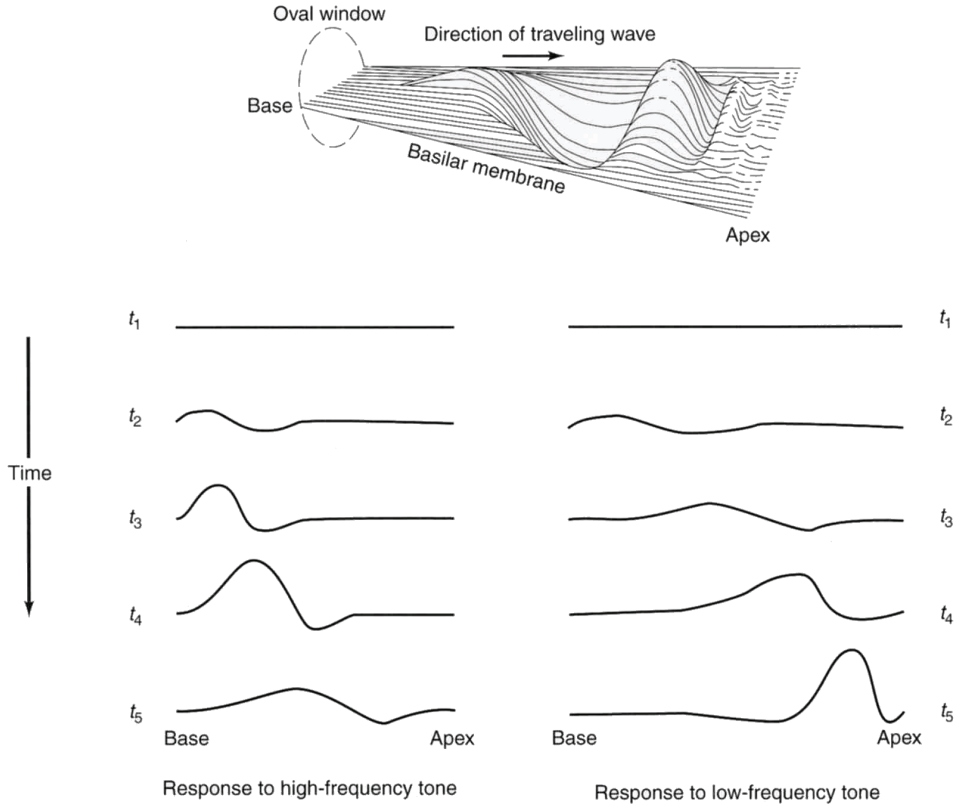


Figure 2.14: The Basilar Membrane of the Cochlea depicted uncoiled and flattened showing the resonance for travelling waves of different frequencies

The cochlear implant consists of a string of electrodes, each excited by a narrow band of frequencies, which stimulate the hair cell nerves directly. This allows some hearing to be restored in the case when either the cilia or basilar membrane has been damaged.

2.6. Analog Modulation and Demodulation

A continuous unmodulated signal cannot be used to measure range as there is no way of determining when the signal was transmitted.

For most sensor applications, the transmitted signal is "marked" in some way either by altering its amplitude or frequency.

The round trip time from the moment that the "mark" is transmitted to when it is received can then be used to determine the range to the target if the speed of propagation is known.

Marking the transmitted signal is known as modulation.

2.6.1. Amplitude Modulation (AM)

AM is defined as a modulation technique in which the amplitude of the carrier is varied in accordance with some characteristic of the baseband modulating signal.

It is the most common form of modulation because of the ease with which the baseband signal can be recovered from the transmitted signal.

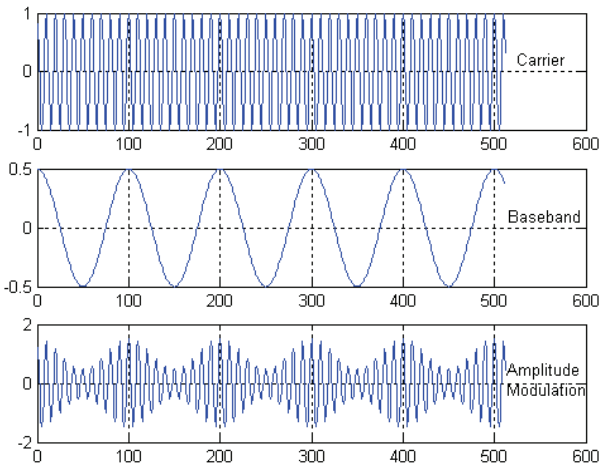


Figure 2.15:Time domain representation of amplitude modulation

For an unmodulated carrier described earlier and for a baseband signal $v_b(t)$, the AM signal $v_{am}(t)$ is described by the following equation

$$v_{am}(t) = A_c\left[1 + v_b(t)\right]\cos 2\pi f_c t\,. \tag{2.6}$$

For the example shown above, where the modulating signal is a sinusoid with a frequency f_a and amplitude A_{am} then the revised formula,

$$v_{am}(t) = A_c\left[1 + A_{am}\cos 2\pi f_a t\right]\cos 2\pi f_c t\,. \tag{2.7}$$

In general A_{am}<1 otherwise a phase reversal occurs and demodulation becomes more difficult.

The extent to which the carrier has been amplitude modulated is expressed in terms of a *percentage modulation* which is just calculated by multiplying A_{am} by 100.

To determine the characteristics of the signal in the frequency domain, it can be rewritten in the following form (using a trig identity),

$$v_{am}(t) = A_c \cos 2\pi f_c t + \frac{A_c A_{am}}{2}\left[\cos 2\pi (f_c - f_a)t + \cos 2\pi (f_c + f_a)t\right]. \qquad (2.8)$$

It can be seen that this is made up from three independent frequencies:

- The original carrier at a frequency of f_c.
- A frequency at the difference between the carrier and the baseband signal
- A frequency at the sum of the carrier and the baseband signal

The spectrum is shown in the figure.

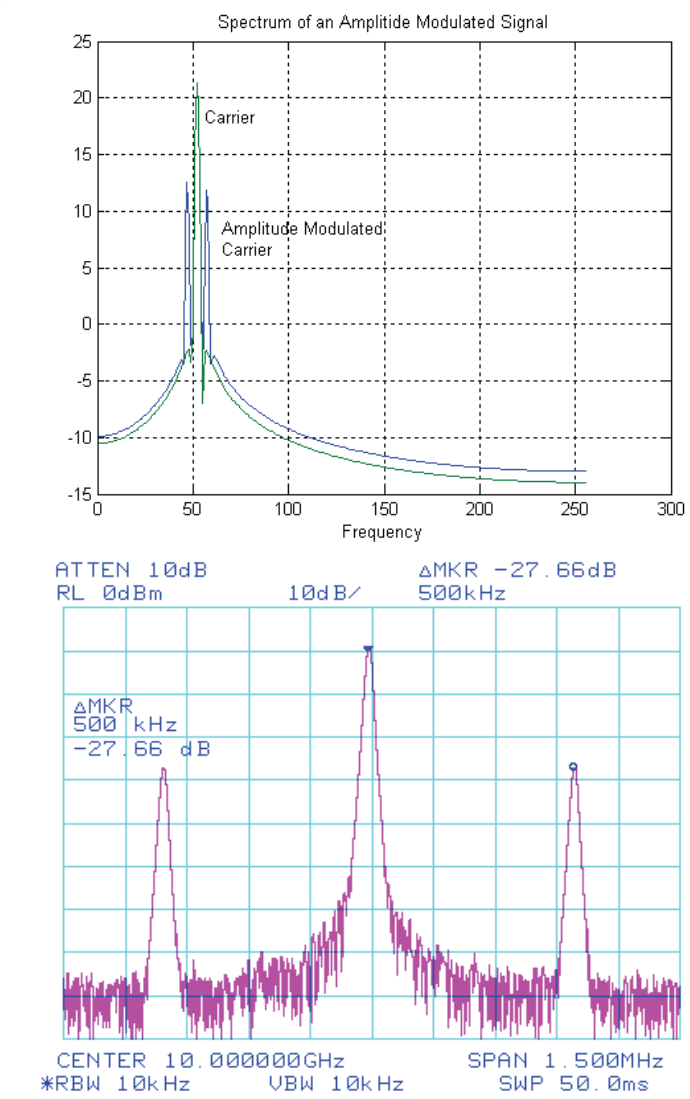


Figure 2.16: Simulated and measured frequency domain representation of amplitude modulation

The "tank" circuit in the crystal radio shown in the figure below is tuned to resonate at the carrier frequency and so operates to select only a single radio program which passes into the detection stage.

Demodulation is then achieved with the use of a simple diode rectifier (crystal) and lowpass filter (capacitor) and a pair of high-impedance headphones converts the resultant envelope current to an audio signal that accurately reproduces the original modulation.

From the figure below it can be seen how simple the demodulation of an AM signal can be.

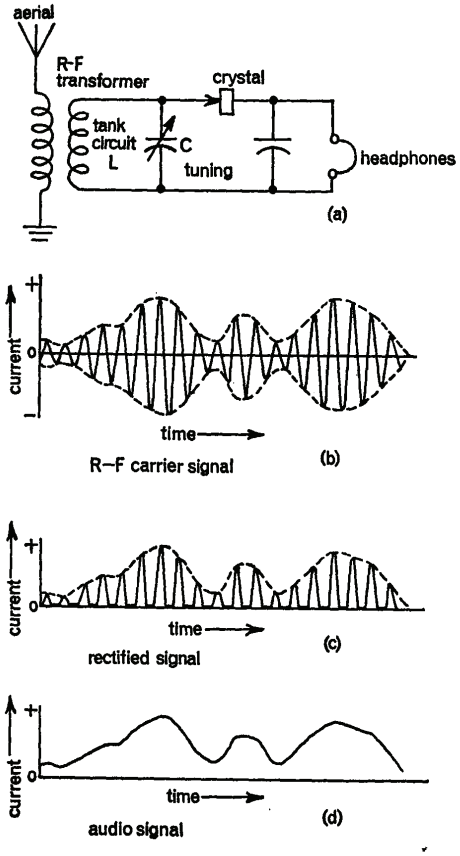


Figure 2.17: Demodulation of an AM signal by a crystal radio

Note that there is sufficient energy in the RF signal (with the appropriate long wire antenna) to drive a set of high-impedance headphones directly without any amplification (hence no battery).

2.7. Frequency Modulation (FM)

FM is a modulation technique in which the frequency of the carrier is varied in accordance with some characteristic of the baseband modulating signal.

$$v_{fm}(t) = A_c \cos\left[\omega_c t + k\int_{-\infty}^{t} v_b(t)dt\right] \tag{2.9}$$

The reason that the modulating signal is integrated is because variations in the modulating term equate to variations in the carrier phase.

The instantaneous angular frequency can be obtained by differentiating the instantaneous phase,

$$\omega = \frac{d}{dt}\left[\omega_c t + k\int_{-\infty}^{t} v_b(t)dt\right] = \omega_c + kv_b(t)\,. \tag{2.10}$$

The deviation of the instantaneous frequency from the carrier frequency, $\omega_c/2\pi$, *is*

$$\delta f = f - f_c = \frac{k}{2\pi}v_b(t) \tag{2.11}$$

This shows that the deviation of the instantaneous frequency is directly proportional to the amplitude of the modulating signal. Hence the combination of an integrator and phase modulator produces frequency modulation.

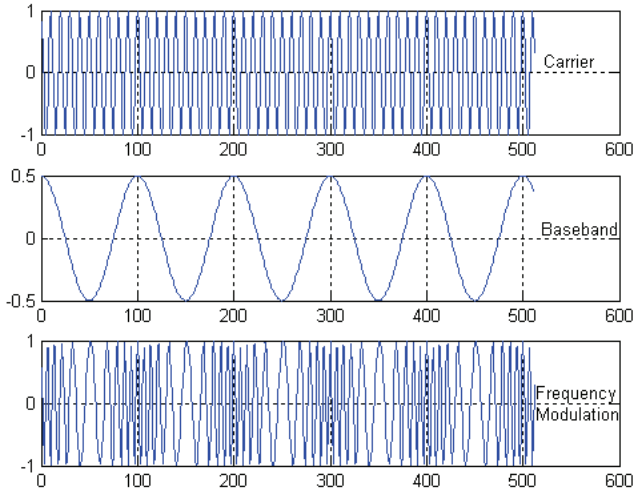


Figure 2.18: Time domain representation of frequency modulation

For sinusoidal modulation, the formula for FM is

$$v_{fm}(t) = A_c \cos\left[\omega_c t + \beta \sin \omega_a t\right], \tag{2.12}$$

where β, which is the maximum phase deviation, is usually referred to as the modulation index.

The instantaneous frequency in this case is

$$\begin{aligned} f &= \frac{\omega_c}{2\pi} + \frac{\beta\omega_a}{2\pi}\cos\omega_a t \\ f &= f_c + \beta f_a \cos\omega_a t \end{aligned} \tag{2.13}$$

So the maximum frequency deviation defined as Δf, when $\cos\omega_a t = 0$, is

$$\Delta f = \beta f_a . \tag{2.14}$$

Even though the instantaneous frequency lies within the range f_c+/-Δf, the spectral components of the signal don't lie within this range.

Some manipulation of the formula for $v_{fm}(t)$ shows that the spectrum comprises a carrier with amplitude $J_o(\beta)$ with sidebands spaced symmetrically on either side of the carrier at offsets of ω_a, $2\omega_a$, $3\omega_a$, ….as shown in the figure below.

Theoretically, the bandwidth is infinite, however, for any β, most of the power is confined within a finite bandwidth.

As a rule of thumb (Carson's Rule), the bandwidth is twice the sum of the maximum frequency deviation plus the modulating frequency.

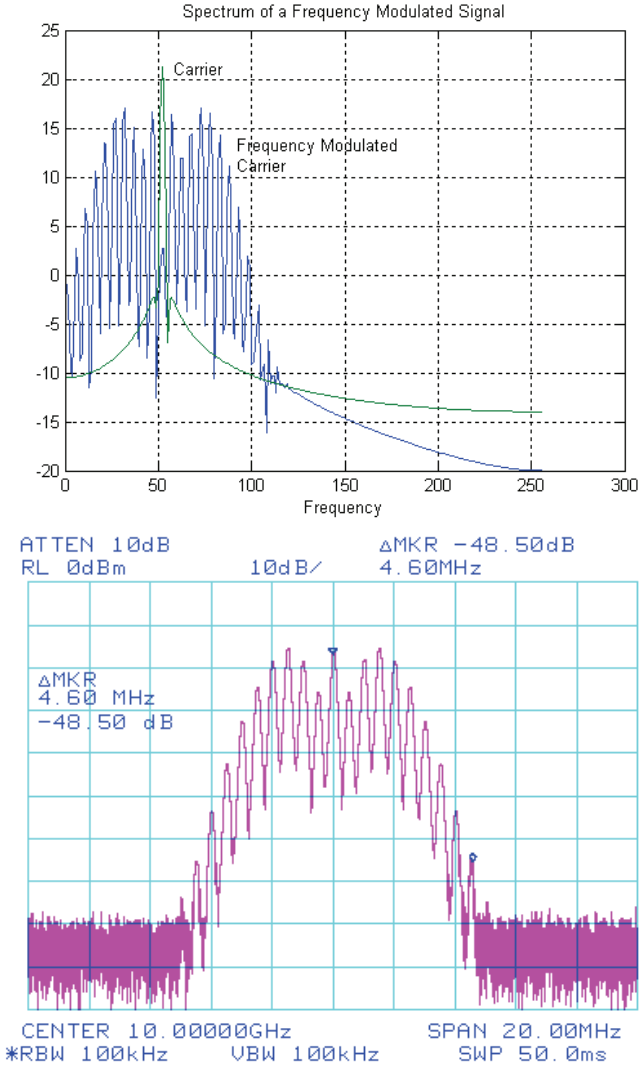


Figure 2.19: Simulated and measured frequency domain representation of frequency modulation

Demodulation of an FM signal is commonly achieved by converting it into AM and then envelope detecting it.

The simplest method to perform the conversion to AM is to pass the signal through a frequency sensitive circuit such as a low-pass or bandpass filter. The circuitry to perform this function is known as a discriminator

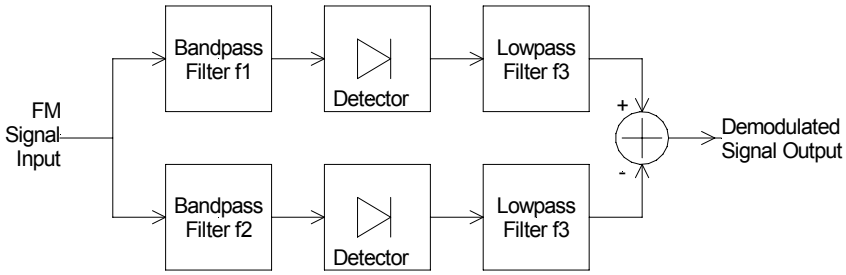


Figure 2.20: A discriminator converts the FM signal into an amplitude variation and envelope detects the resulting AM signal

In this example the FM signal is split and passed through two bandpass filters with centre frequencies just below, and just above the carrier frequency. The two signals are then detected and filtered to remove the residual carrier before the difference is taken. The transfer function for this process is shown in the figure below.

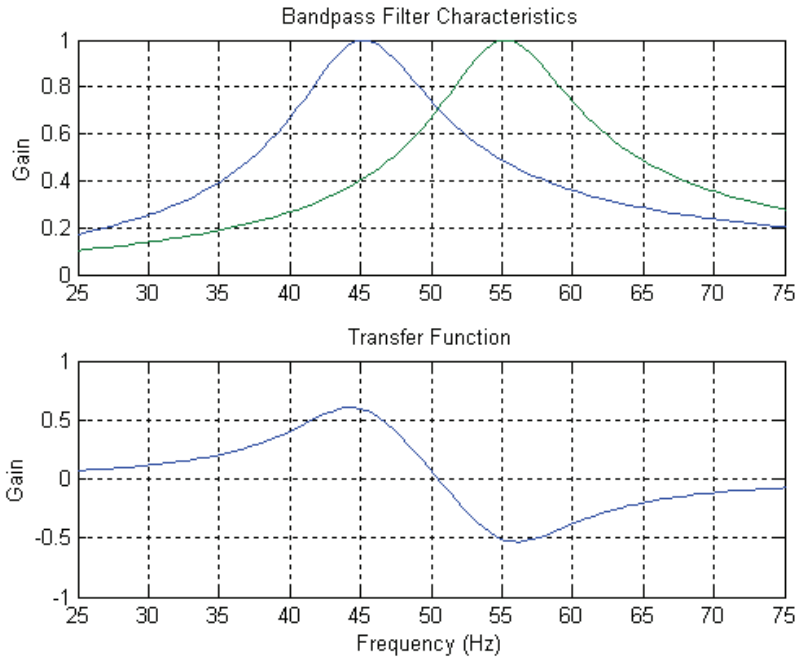


Figure 2.21: The difference signal from a pair of offset bandpass filters produces a symmetrical transfer function to convert variations in frequency to variations in amplitude

In the time domain, the FM signal after detection and filtering produces two symmetrical demodulated signals with DC offsets. The difference between these reproduces the original baseband signal as shown in the figure below.

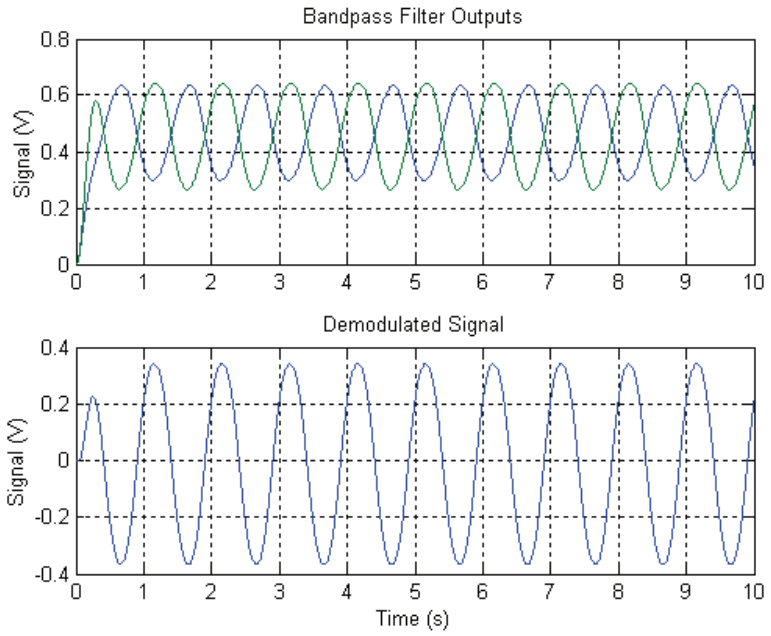


Figure 2.22: Stages of FM demodulation showing (a) the detected and filtered outputs from the bandpass filters and (b) the final demodulated output obtained by taking the difference between these signals

Alternative techniques used in most modern radio receivers use either phase-locked loop or quadrature detection techniques to perform this function.

2.8. Linear Frequency Modulation

In most active sensors that operate using the Frequency Modulated Continuous Wave (FMCW) principle, the frequency is not modulated sinusoidally, but in a linear manner with time.

$$\omega_b = A_b t \,. \tag{2.15}$$

Substituting into the standard equation for FM, we obtain the following result

$$\begin{aligned} v_{fm}(t) &= A_c \cos\left[\omega_c t + A_b \int_{-\infty}^{t} t dt\right] \\ v_{fm}(t) &= A_c \cos\left[\omega_c t + \frac{A_b}{2} t^2\right] \end{aligned} \tag{2.16}$$

Note that the phase modulation follows a quadratic function.

In general it is not possible to continue to increase the frequency indefinitely, so the modulation often follows a sawtooth or triangular function.

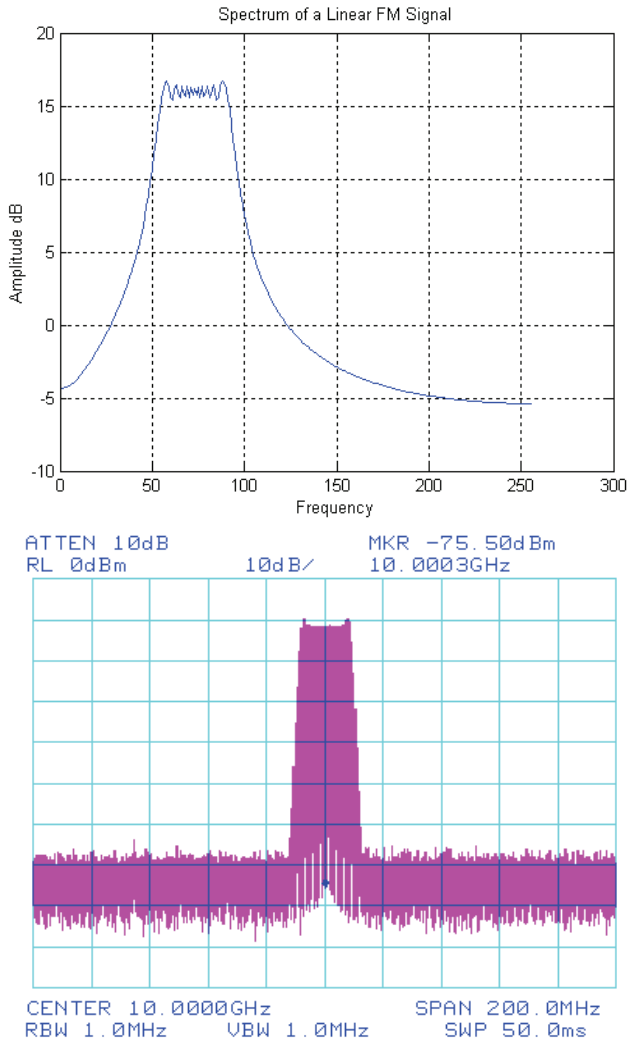


Figure 2.23: Simulated and measured frequency domain representation of linear FM

In FMCW systems, a portion of the transmitted signal is mixed with (multiplied by) the returned echo as discussed in Chapter 11.

The transmit signal will be shifted from that of the received signal because of the round trip time, τ,

$$v_{fm}(t-\tau) = A_c \cos\left[\omega_c(t-\tau) + \frac{A_b}{2}(t-\tau)^2\right]. \tag{2.17}$$

Calculating the product of the transmitted signal and the delayed echo

$$v_{fm}(t-\tau)v_{fm}(t) = A_c^2 \cos\left[\omega_c t + \frac{A_b}{2}t^2\right]\cos\left[\omega_c(t-\tau) + \frac{A_b}{2}(t-\tau)^2\right] \tag{2.18}$$

Equating using the trig identity $\cos A \cos B = 0.5[\cos(A+B)+\cos(A-B)]$

$$v_{out}(t) = \frac{1}{2}\left[\begin{array}{l}\cos\left\{\left(2\omega_c - A_b\tau\right)t + A_b t^2 + \left(\frac{A_b}{2}\tau^2 - \omega_c\tau\right)\right\} \\ + \cos\left\{A_b\tau.t + \left(\omega_c\tau - \frac{A_b}{2}\tau^2\right)\right\}\end{array}\right] \tag{2.19}$$

The first cosine term describes a linearly increasing FM signal (chirp) at about twice the carrier frequency with a phase shift that is proportional to the delay time τ. This term is generally filtered out.

The second cosine term describes a beat signal at a fixed frequency,

$$f_{beat} = \frac{A_b}{2\pi}\tau\,. \tag{2.20}$$

It can be seen that the signal frequency is directly proportional to the delay time τ, and hence is directly proportional to the round trip time to the target.

The spectrum of this output signal is shown in the following figure.

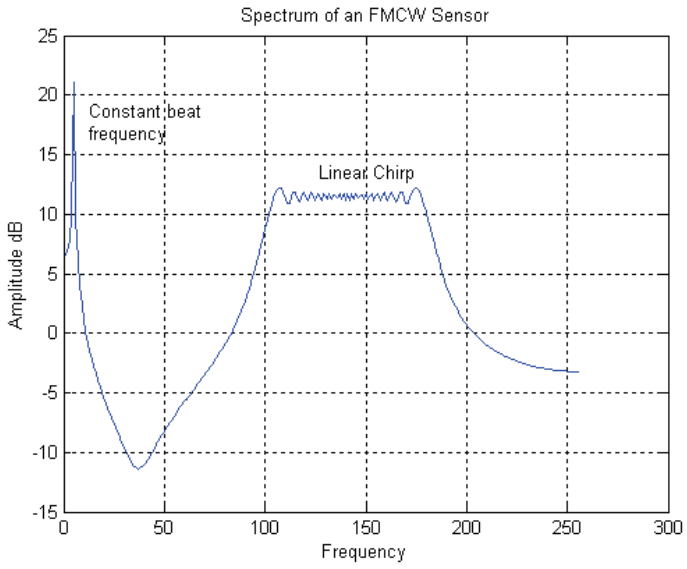


Figure 2.24: Frequency domain representation of FMCW sensor output

2.9. Pulse Coded Modulation Techniques

2.9.1. Pulse Amplitude Modulation (PAM)

Modulation in which the amplitude of individual, regularly spaced pulses in a pulse train is varied in accordance with some characteristic of the modulating signal. For time-of-flight sensors, the amplitude is generally constant.

The ability of a pulsed sensor to resolve two closely spaced targets is determined by the pulse width as shown in more detail in Chapter 11.

The following figure shows the relationship between the width of a single pulse and its spectral content.

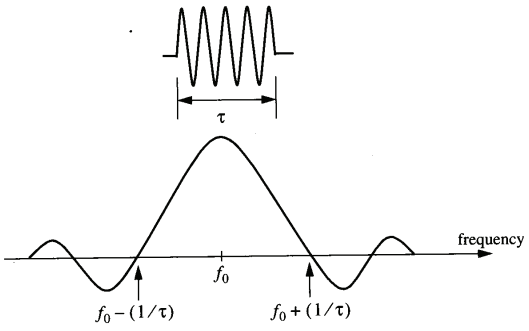


Figure 2.25: Relationship between pulsewidth and frequency for a single pulse

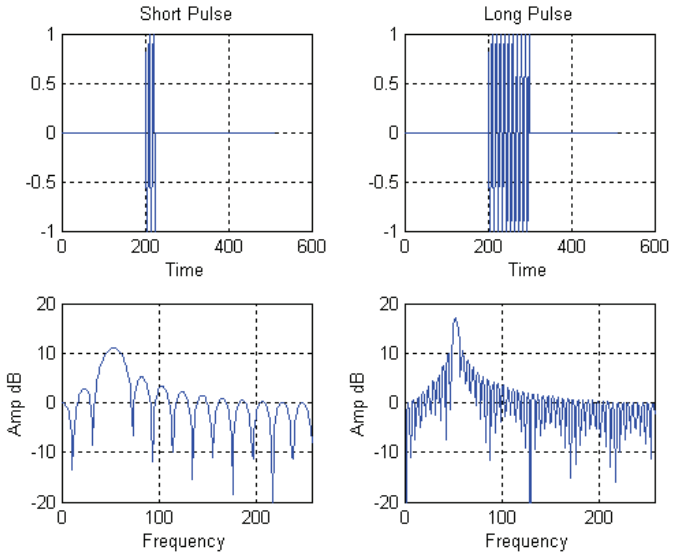


Figure 2.26: Relationship between the duration of a pulse and its spectrum for pulsed amplitude modulation

Note that the width of the pulse spectrum in the frequency domain increases as the pulse width decreases with the following relationship,

$$\beta \approx \frac{1}{\tau}, \tag{2.21}$$

where β is the spectral width (Hz) between the half power (3dB) points and τ is the pulsewidth (sec). In a time-of-flight sensor, this relationship determines the bandwidth required by a receiver to receive pulses of a specific duration. A filter that conforms to this relationship is known as a "matched" filter. It is formally defined in Chapter 11.

In a repeated sequence of pulses, the fine structure of the spectrum is determined by the total length of the observed sequence as shown in the figure below.

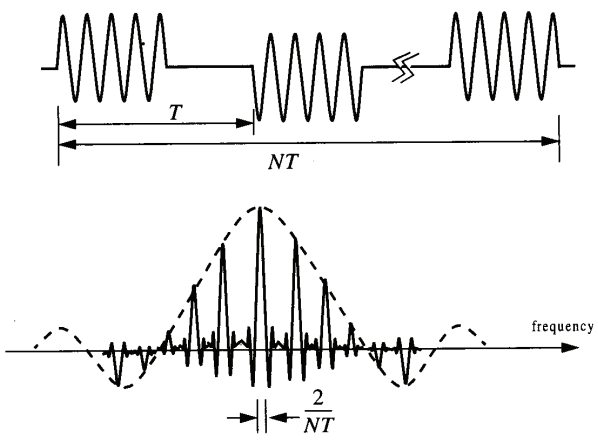
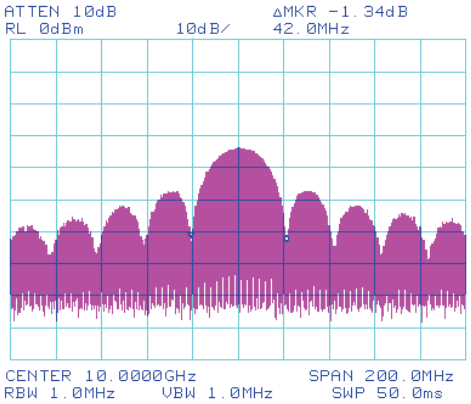
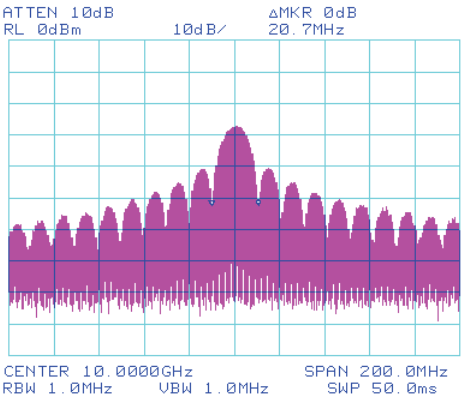


Figure 2.27: Relationship between pulsewidth and frequency for a finite length sequence of pulses

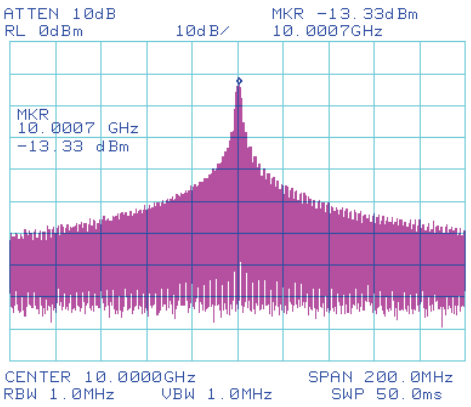
The following figures show the measured spectra of a number of continuous and pulsed signals with different periods



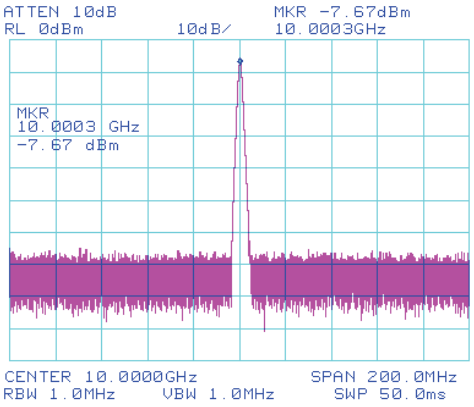
(a)



(b)



(c)



(d)

Figure 2.28: Spectra of pulsed signals with durations of (a) 50ns, (b) 100ns, (c) 500ns, (d) continuous

2.10.Frequency Shift Keying (FSK)

This modulation technique is the digital equivalent of linear FM where only two different frequencies are utilised. A single bit can be represented by a single cycle of the carrier, or if the data rate is not critical, then multiple cycles can be used.

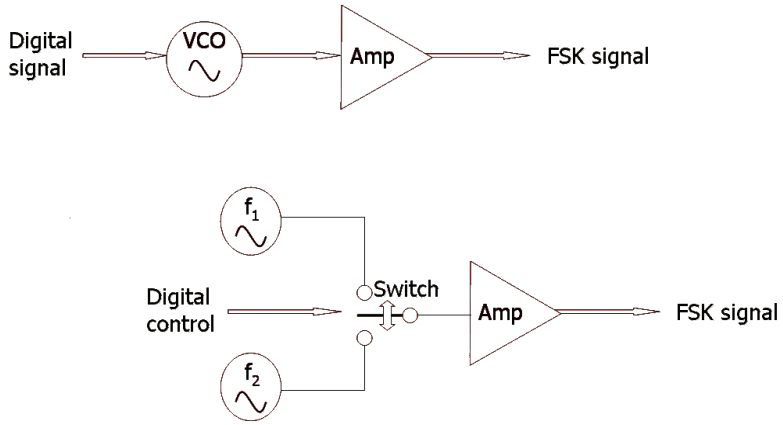


Figure 2.29: Two common methods of generating FSK

Demodulation can be achieved by detecting the outputs of a pair of filters centred at the two modulation frequencies, f_1 and f_2 as shown, or by using a phase locked loop.

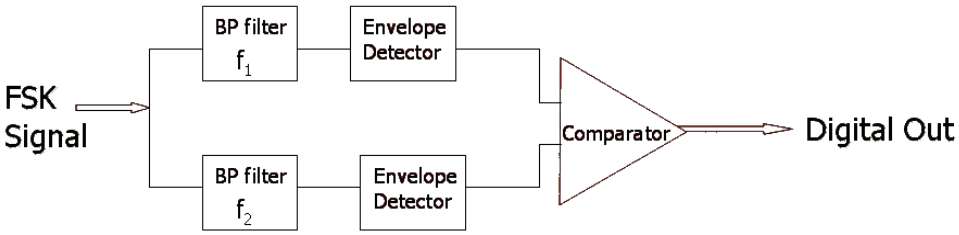


Figure 2.30: Band pass filter method of demodulating FSK signals

The following figure shows an example of FSK modulation with multiple cycles per bit.

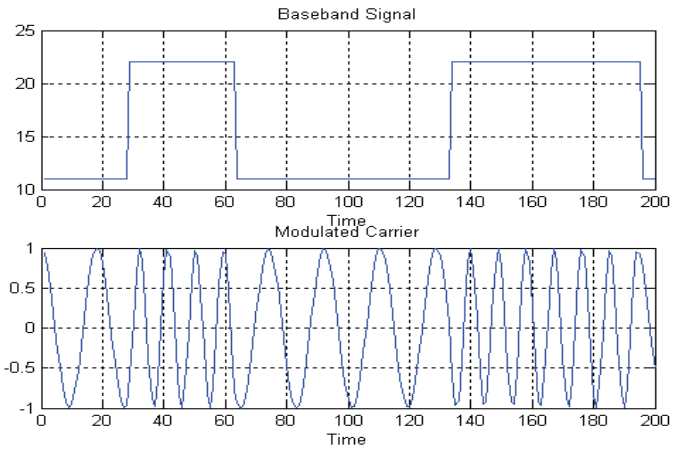


Figure 2.31: Example of frequency shift keying (FSK)

In coherent FSK, the bit transition is synchronised with the carrier frequency so that there are no phase discontinuities. In the non-coherent option there is no synchronisation and large phase discontinuities can occur

As the number of cycles per bit decreases, the spectrum shifts from being two distinct peaks centred t the two frequencies to a flatter, broad almost uniform peak as shown in the figures below.

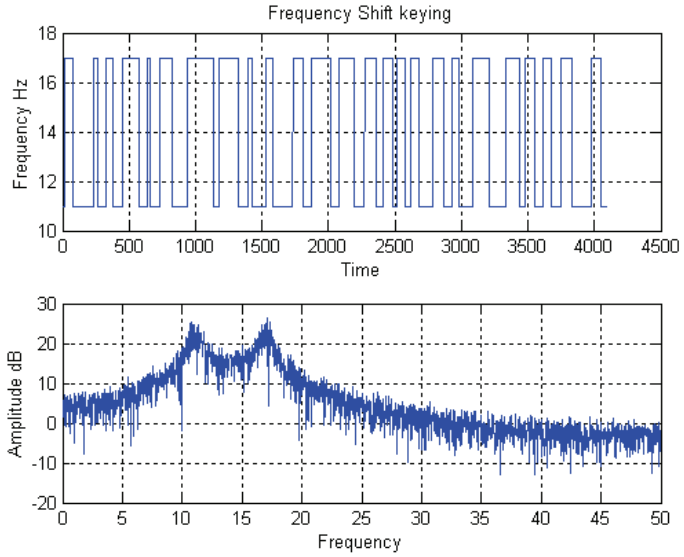


Figure 2.32: FSK spectrum, 5 cycles per bit

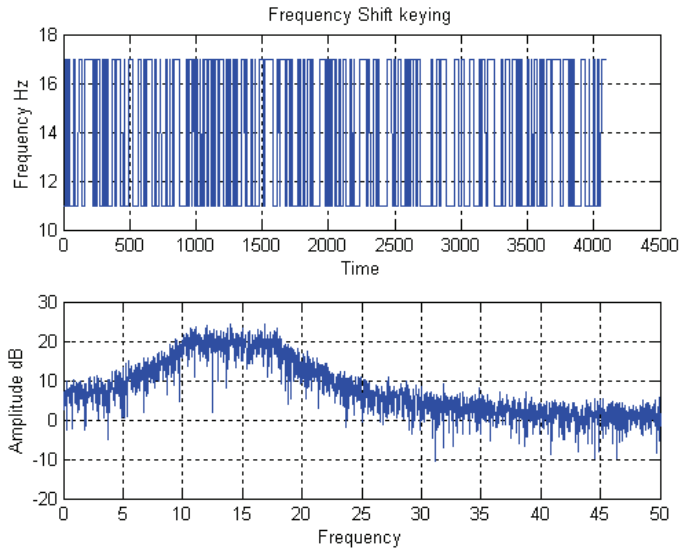


Figure 2.33: FSK spectrum, 1 cycle per Bit

FSK is a very simple modulation technique and is still extremely popular. It was originally used by teleprinters which operated at about 45bps, and was introduced in 1962 for a Bell modem which operated at up to 300bps. Early PC's used a Kansas City Interface which used FSK to store software on audio cassettes at up to 1200bps. It is now used in touch-tone phones and a myriad of other communications systems, operating at speeds in excess of 1Mbps.

2.11.Phase Shift Keying (PSK)

The most common form of PSK is binary phase coding. The carrier, f_o is switched between +/-180° according to a digital base band, f_m sequence.

This modulation technique can be implemented quite easily using a balanced mixer as shown in the figure, or with a dedicated BPSK modulator.

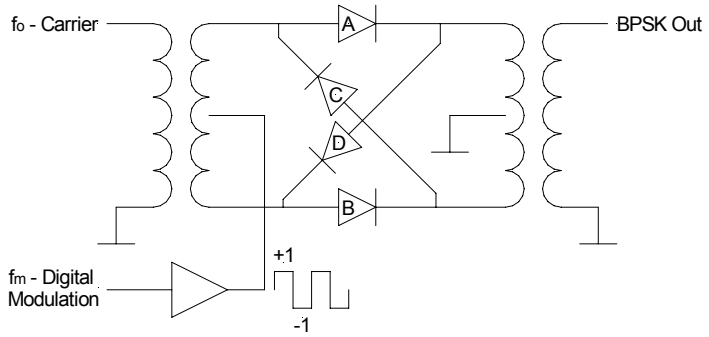


Figure 2.34: Implementing BPSK using a balanced mixer

When the modulation signal f_m is high (+1), diodes A and B are forward biased and the carrier f_o is coupled to the directly to the output transformer. However, when f_m is low (-1), then diodes C and D are forward biased and f_o is coupled to the to the opposite terminals of the output transformer which results in a reversal of the phase.

The modulation and output waveforms are shown in the following figure.

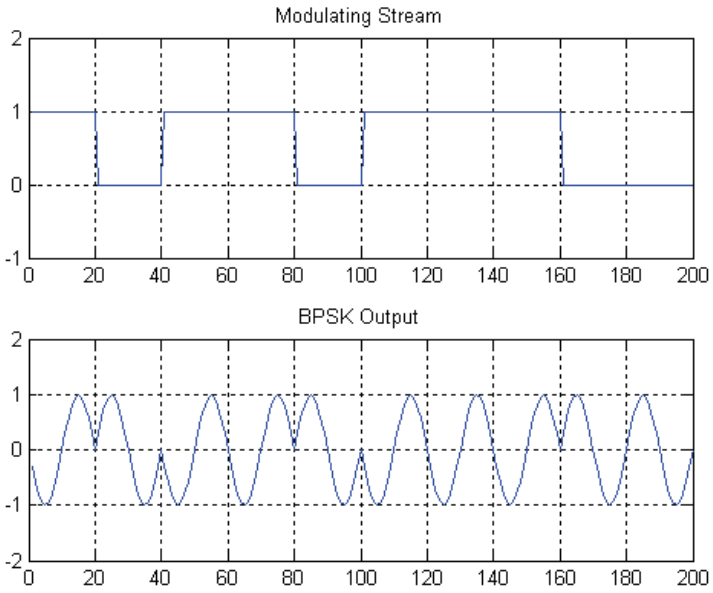


Figure 2.35: Example of binary phase shift keying with one cycle per bit

Demodulation is achieved by multiplying the modulated signal by a coherent carrier (a carrier that is identical in frequency and phase to the carrier that originally modulated the BPSK signal).

This produces the original BPSK signal plus a signal at twice the carrier which can be filtered out.

The spectral width of the carrier is widened by the modulation process as shown in the figure below. This is one of the modulation techniques that is used for "broadband" communications for obvious reasons.

Note in this example that the amplitude of the transmitted signal has been reduced by 30dB because the power has been spread, even though the total power transmitted (and hence the range performance) remains unaltered (see Chapter 11).

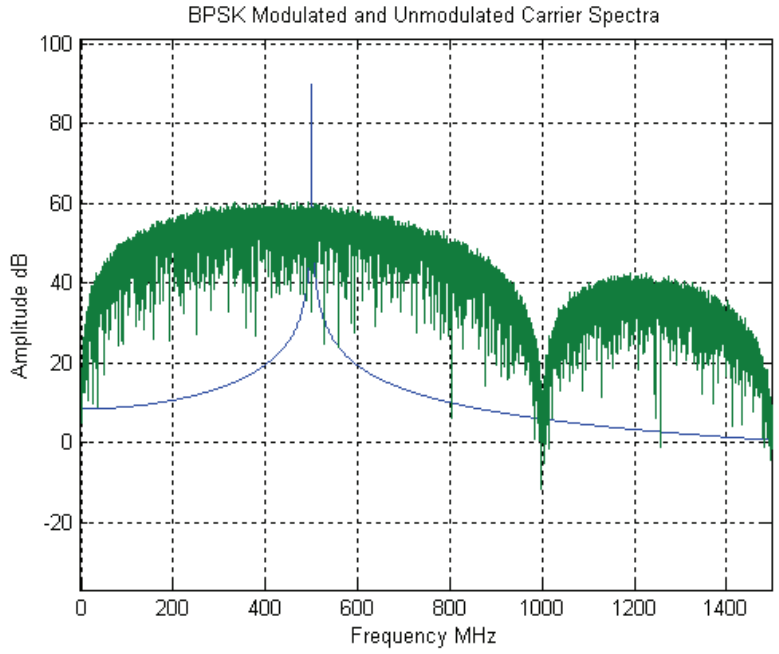


Figure 2.36: PSK Spectrum for a 500MHz carrier and 1 cycle per bit modulation

2.12.Stepped Frequency

In this modulation technique, a sequence of pulses are transmitted each at a slightly different frequency.

The pulse width (in the radar case) is made sufficiently wide to span the region of interest, but because it is so wide, it cannot resolve individual targets within that region. However, if all of the pulses are processed together, the effective resolution is improved because the total bandwidth is widened by the total frequency deviation of the sequence of pulses as discussed below.

If a signal is transmitted at a frequency ω_c and it reflects off a target at a range R, then the round trip delay time is

$$\tau = \frac{2R}{c}. \tag{2.22}$$

The phase difference between the transmitted signal and the echo can be determined by mixing a portion of the transmitted signal with the echo as follows and filtering out the high frequency components,

$$\begin{aligned} v_{out}(t) &= \cos\omega_c t.\cos\omega_c(t-\tau) \\ v_{out}(t) &= \frac{1}{2}\left[\cos\omega_c(2t-\tau)+\cos\omega_c\tau\right] \end{aligned} \tag{2.23}$$

Substituting for τ and filtering the high frequency components leaves the phase term only,

$$v_{out}(t) = \frac{1}{2}\cos\frac{2\omega_c R}{c}. \tag{2.24}$$

The phase shift, ϕ, can be obtained by replacing ω_c by $2\pi f_c$

$$\phi = \frac{4\pi f_c R}{c}. \tag{2.25}$$

It can be seen that the phase shift will change with each new frequency f_c

This phase change is sinusoidal in nature and can be considered to be a synthetic Doppler frequency and can be processed in the frequency domain to determine the target range more accurately.

If a number of closely spaced reflectors are present they will each have their own unique Doppler frequency and so can each be resolved as a separate target.

2.13.Convolution

2.13.1. Linear Time Invariant Systems

It is convenient to describe the relationship between the input $x(t)$ and output $y(t)$ signals of Linear Time Invariant (LTI) systems in terms of the impulse response $h(t)$ of the system. It can be shown that an LTI system is completely characterised by its impulse response $x(t) = \delta(t)$ applied at time $t = 0$ (or $n = 0$).

Applying an impulse to the input of an unknown LTI system is therefore a good method of determining its characteristics. This is easily achieved in the discrete time case where the input is set equal to an impulse $\delta(n)$. However, in the continuous time case, it is not possible to produce a true impulse with zero width and infinite amplitude and an approximation is used.

As an arbitrary input signal can expressed as the weighted superposition of time shifted impulses $x(t) = x(\tau)\delta(t-\tau)$, the system output is just the weighted superposition of time-shifted impulse responses.

$$y(t) = \int_{\tau=-\infty}^{t} h(t-\tau)x(\tau)d\tau = \int_{0}^{\infty} h(\tau)x(t-\tau)d\tau \tag{2.26}$$

This weighted superposition is termed the *convolution integral* for continuous time systems and the *convolution sum* for discrete time systems.

Given a system defined by its impulse response $h(t)$, the output, $y(t)$ is found by convolving the input $x(t)$ with this function.

If the Laplace transform is taken of this convolution integral, it can be shown that

$$Y(s) = H(s)X(s) \tag{2.27}$$

That is, convolution in the time domain is equivalent to multiplication in the Laplace domain. Additionally, if s is replaced by $j\omega$, then the equation can be replaced by

$$Y(\omega) = H(\omega)X(\omega) \tag{2.28}$$

which describes the frequency response of the system. It is the magnitude of this transfer function, $|H(\omega)|$, that is considered when the characteristics of various filters are considered in the following section.

In this case, the mapping between the time domain and the frequency domain is through the Fourier Transform. In the following example, the impulse response of a 3rd order Butterworth bandpass filter (normalised passband 0.2 to 0.3) is generated. As this characterises the filter completely, its Fourier Transform should reproduce the frequency response accurately.

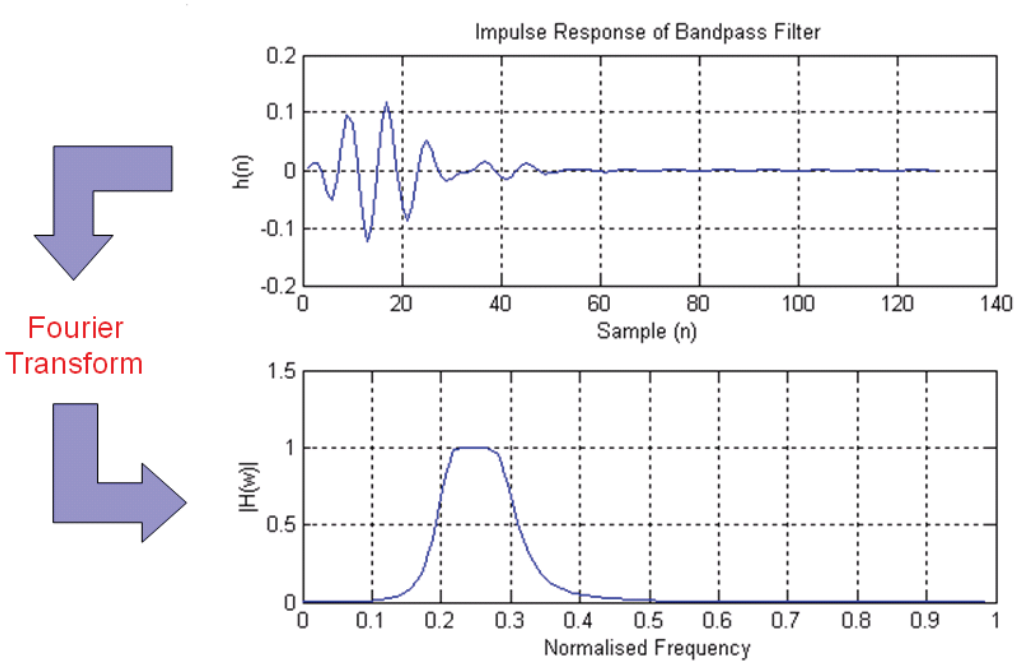


Figure: MATLAB example showing the relationship between the impulse response of a bandpass filter and its Fourier Transform

```
% relationship between filters in the time and frequency domain
%
% generate a Butterworth bandpass filter
wn=[0.2,0.3];
[b,a]=butter(3,wn)

% generate the impulse response of the filter
x=zeros(1,128);
x(1)=1;
y=filter(b,a,x);
subplot(211), plot(y),grid, xlabel('Sample (n)'), ylabel('h(n)')
title('Impulse Response of Bandpass Filter')

% take the Fourier transform of the impulse response
f=fft(y);
freq=(0:63)*1/64;
subplot(212), plot(freq, abs(f(1:64))), grid, xlabel('Normalised
Frequency'),ylabel('|H(w)|')
%title('Frequency Response')
```

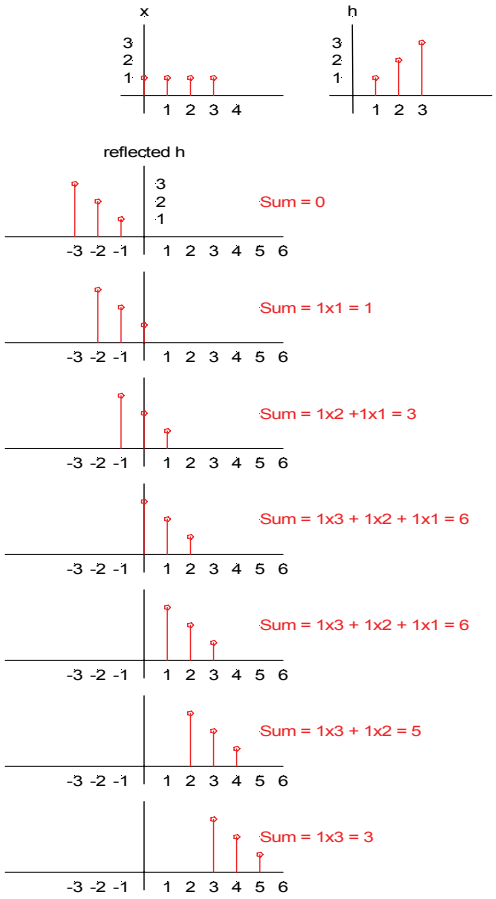
2.13.2. The Convolution Sum

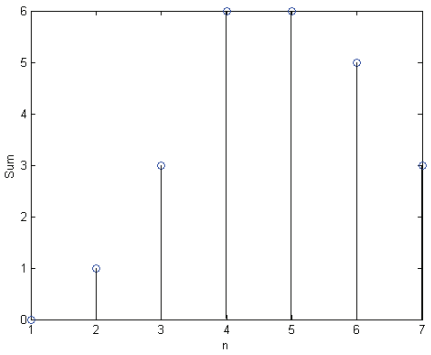
Most modern systems are digital which requires that the input and output data be sampled. In this case the *convolution sum* replaces the convolution integral and is defined by the following equation where $y(n)$ is the output, $x(n)$ the input and $h(n)$ the system impulse response

$$y(n) = \sum_{k=-\infty}^{\infty} x(k)h(n-k) \tag{2.29}$$

However, to be tractable $x(n)$ and $h(n)$ are nonzero only over a finite interval. In the example below $N_x = 4$ and $N_h = 3$ which makes the total duration of the convolved sequence $N_x + N_h - 1 = 6$

To perform this function manually, the order of $h(n)$ is first reversed before being shifted across $x(n)$ one sample at a time. The output $y(n)$ is equal to the sum of the products of each of the aligned terms






```
% convolution example
% convsum.m
x = ([1,1,1,1]);
h = ([0,1,2,3]);
y = conv(x,h);
plot(y,'o')
```

2.13.3. Convolution Example

An unusual application for the convolution process is to use it to model the propagation of a pulsed time-of-flight sensor.

In this application the impulse response of the round-trip propagation to the point target is a signal that is delayed in time and attenuated in amplitude $h(t) = a\delta(T-\tau)$ where a represents the attenuation and τ the round trip time to the target and back.

Consider an aircraft flying towards a radar system as shown in the following figure:



Figure 2.37: Radar illuminating an aircraft target

Assume that the main points of reflection from the aircraft are listed in the following table:

Table 2.1: Reflection contributions from various parts of the aircraft

Contribution	Distance from Datum (m)	Magnitude
Nose	2	1
Wing	6	1
Engine	8	1
Tail	14	1

Determine the magnitude of the echo signal if a pulse with a length of 3m is transmitted.

- The transmit signal is treated as a sequence of impulses separated by the sample interval of 10cm covering a total of 3m in range
- The target is treated as a sequence of four impulses with separations as listed in the table
- The reflected signal received back at the radar is well described by the convolution of the transmitted pulse and the array of point reflectors as shown in the figure.

```
% convolution demo
%
% generate the transmit pulse as a sequence of impulses 10cm apart
a=([zeros(1,70),ones(1,30),zeros(1,70)]);

% generate the point target reflectors
b=zeros(1,170);
b(20)=1;                        % Nose
b(60)=1;                        % Wings
b(80)=1;                        % Engine
b(140)=1;                       % Tail

% take the convolution to determine the return from all of the reflectors
c=conv(a,b);

% plot the results
x=(1:170)/10;
subplot(211) ,plot(x,a,x,a,'o'), grid;
title('Transmit Pulse'),ylabel('amplitude');
subplot(212), plot(x,b,x,b,'o'),grid;
title('Target Reflectors'),xlabel('range(m)'),ylabel('amplitude');
figure(2)
xx=(1:339)/10;
plot(xx,c,xx,c,'+');
grid
title('Echo Pulse')
xlabel('range(m)')
ylabel('amplitude')
```

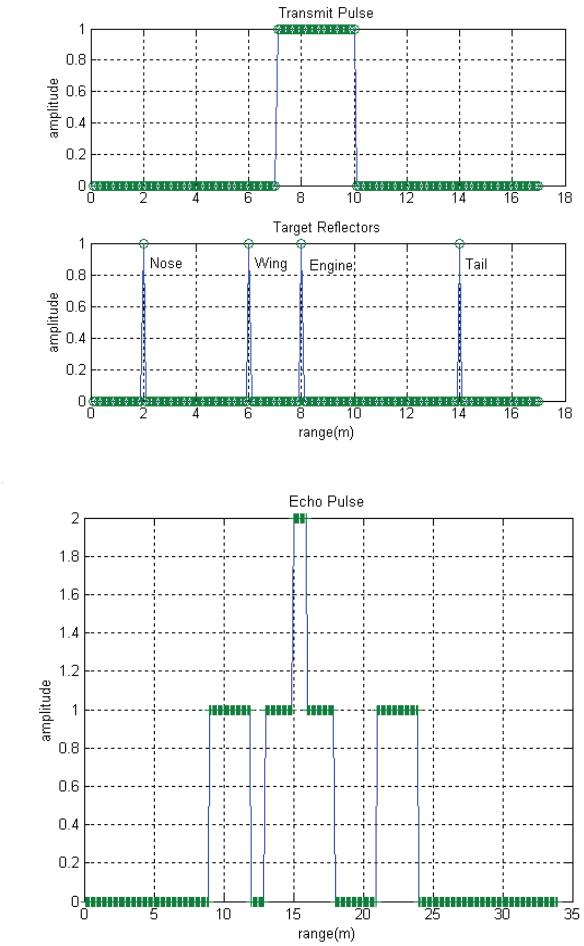


Figure 2.38: Matlab simulation output showing transmit signal, target structure and echo signal

This example does not consider that the transmitted signal is in fact sinusoidal in nature and hence the convolution should include both amplitude and phase effects. Nor does not consider the effects of the round-trip distance which doubles the effective spacing between the reflectors as explained in Chapter 5.

2.14.References

[1] N.Taub & D.Schilling, *Principles of Communication Systems.* McGraw Hill. 1971

[2] B.Mahafza, *Radar Systems Analysis and Design using MATLAB.* Chapman & Hall/CRC. 2000

[3] N.Currie, *Radar Reflectivity Measurement: Techniques and Applications.* Artech House, 1989

[4] M.Skolnik, *Radar Handbook 2nd Ed.* McGraw Hill. 1990

[5] N.Currie & C.Brown. *Principles and Applications of Millimeter-Wave Radar.* Artech House. 1987

[6] T.Duncan, *Electronics and Nuclear Physics.* John Murray. 1973

[7] J.Carr, *Electronic Circuit Guidebook: Sensors.* Prompt. 1997

[8] H.Jacobovitz, *Electronics Made Simple*. W.H.Allen. 1972

[9] *The ARRL Handbook for Radio Amateurs*: 2000

[10] S.Haykin & B.van Veen, *Signals and Systems.* John Wiley. 2003

[11] H. Durrant-Whyte, *Signal Analysis and Transform Methods,* Course Notes

Chapter 3.
Radiometers & Image Intensifiers

3.1. Introduction

Radiometers are instruments for detecting or measuring radiant energy. The term is generally, though not always, applied to devices used to measure infrared radiation.

The concept of blackbody radiation is introduced in terms of its temperature dependent spectral characteristics as described by the Planck Function. It is shown that the total radiated (or absorbed) power density can be determined by integration and that if the bandwidth is sufficiently wide that this is proportional to the temperature of the body (in Kelvin) raised to the 4^{th} power as described by Boltzmann's Law. This relationship can be exploited to measure temperature, and some of the simplest radiometers do perform this function.

For example, a non-contact thermometer consists of a single radiation-sensitive element called a bolometer, and some form of lens or antenna to constrain its field of view. The radiated energy from the scene heats the element which changes its electrical characteristics which can be measured and the temperature determined.

Early thermal imaging systems scanned a single element over the focal plane in a raster fashion to build up an image. However, more modern systems use a two dimensional array of temperature sensitive elements to produce an image similar to those produced by digital cameras. However, because the amount of energy that illuminates each element is miniscule these arrays are generally made from particularly sensitive materials which are often cooled cryogenically.

The principles used by image intensifiers or "night vision" cameras are completely different to those used by thermal imagers. In such systems an image is projected onto a photo-cathode which generates up to one electron per photon which can be accelerated by an electric field. When the high-energy electron strikes a phosphor it generates many photons and so the image is intensified. These devices can be made sensitive to the infrared by the selection of the appropriate photo-cathodes.

3.2. Thermal Emission

In any object, every atom and every molecule vibrates. The average kinetic energy of the vibrating particles is represented by the absolute temperature (Kelvin). According to the laws of electrodynamics a moving electric charge is associated with a variable

electric field that produces an alternating magnetic field. In essence, this interaction produces an electromagnetic wave that radiates from the body at the speed of light.

3.2.1. Blackbody Radiation

A blackbody is defined as an object which absorbs all radiation that impinges on it at any wavelength. The apparent misnomer is explained by Kirchhoff's law which states that a body that is capable of absorbing all radiation is equally capable of the emission of radiation.

A practical example of the principle is a box that is light proof except for a small hole. Any radiation that enters the hole is scattered and absorbed by repeated reflections from the walls so that almost no energy escapes. If such a box is heated, it becomes what is known as a "cavity radiator", and it radiates energy the characteristics of which are determined only by the temperature.

Classical theory stated that as the temperature was increased, more modes of vibration should be possible and so any black body should radiate huge amounts of energy in the blue and ultraviolet. This was contrary to measured data, and was known as "the ultraviolet catastrophe".

3.2.2. The Planck Function

Planck suggested that the radiation should somehow be constrained so that they could not continuously emit radiation, but should only emit in quanta of a definite magnitude, the size of which would increase with increasing frequency

$$E = hf = hc/\lambda \tag{3.1}$$

where: E – Energy (J)
h – Planck's constant (6.625×10^{-34} Js)
f - Frequency (Hz)
c – Speed of Light (m/s)
λ - Wavelength (m)

Planck went on to produce a formula that fitted the measured relationship between radiated wavelength and temperature even though he did not at first have any theoretical justification for it.

In this version of the formula $B_\lambda(T)$ is the energy (J) emitted per second per unit wavelength from one square meter of a perfect blackbody at a temperature T.

$$B_\lambda(T) = \frac{2\pi hc^2/\lambda^5}{e^{hc/\lambda kT} - 1} \tag{3.2}$$

where: h – Planck's Constant (6.625×10^{-34} Js)
k – Boltzmann's Constant (1.3804×10^{-23} J/K)
λ – Wavelength (m)
c – Speed of light (3×10^8 m/s)
T – Temperature (K)

This spectrum is shown graphically in the following figure:

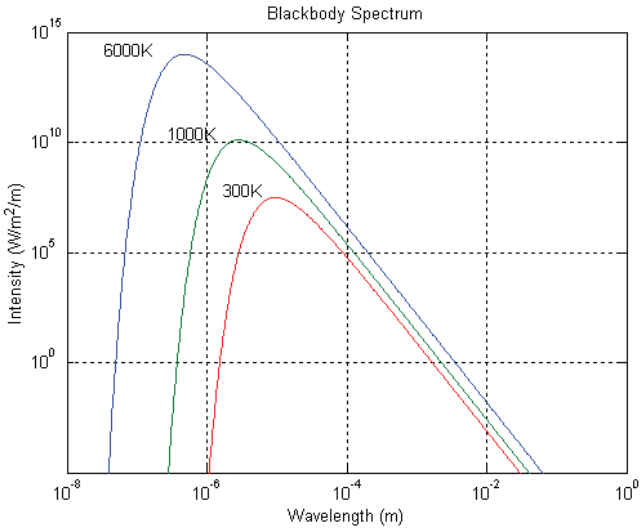


Figure 3.1: Blackbody spectra at different temperatures

Properties of the Planck Function

Because the radiation process is statistical in nature and a large number of particles are involved, a broad range of wavelengths is radiated.

By equating to zero the first derivative with respect to the wavelength, λ, of Planck's Function, the wavelength around which most power is radiated is:

$$\lambda_{max} = \frac{2898}{T} \ \mu m \tag{3.3}$$

This is known as Wien's Displacement Law which states in simple terms that high temperature sources emit most power at short wavelengths.

At long wavelengths ($hc<<\lambda kT$) the power density per unit wavelength becomes proportional to temperature. This is known as the Rayleigh-Jean Law.

$$B_\lambda \approx \frac{2\pi kcT}{\lambda^4} \ W/m^2/\text{unit wavelength} \tag{3.4}$$

This formula is obtained by using a Taylor expansion on the denominator of the Planck function and setting the higher order terms to zero

The total power density, Φ^0 in W/m^2 within a particular bandwidth is determined by integrating B_λ over that bandwidth. This is normally solved numerically or by approximation as there is no closed form.

$$\Phi^0 = \int_{\lambda_1}^{\lambda_2} \frac{2\pi hc^2/\lambda^5}{e^{hc/\lambda kT} - 1} d\lambda \ \ W/m^2 \tag{3.5}$$

If the bandwidth includes much more than 50% of the radiated power then the Stefan-Boltzmann law can be used to approximate the value

$$\Phi^0 \approx \sigma T^4 \quad \text{W/m}^2, \tag{3.6}$$

where σ is the Stefan-Boltzmann constant (5.67×10^{-8} $\text{Wm}^{-2}\text{K}^{-4}$).

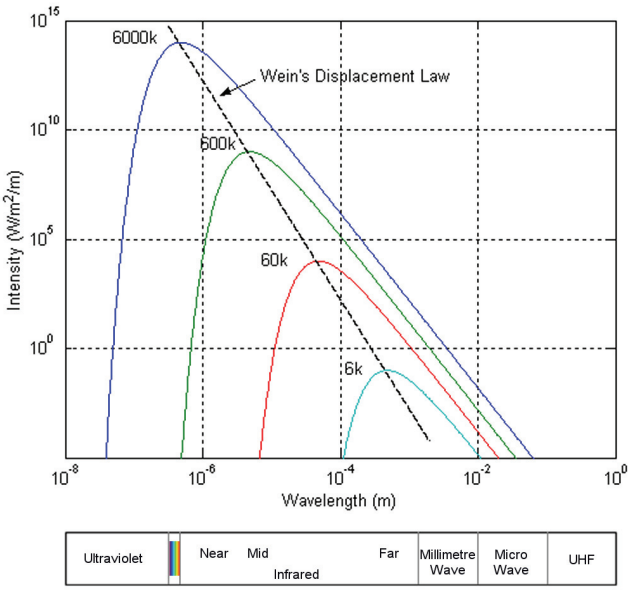


Figure 3.2: The peak of the black body spectrum is a function of the temperature, a relationship known as Wein's Displacement Law

3.2.3. Example

As an example, the total radiated power density (W/m^2) is determined by integrating numerically over the band from 10^{-8} to 10^{-1} m (10nm to 10cm) at a number of different temperatures The same curves are also integrated over the microwave and millimetre wave band from 10^{-3} to 10^{-1} m (1mm to 10cm).

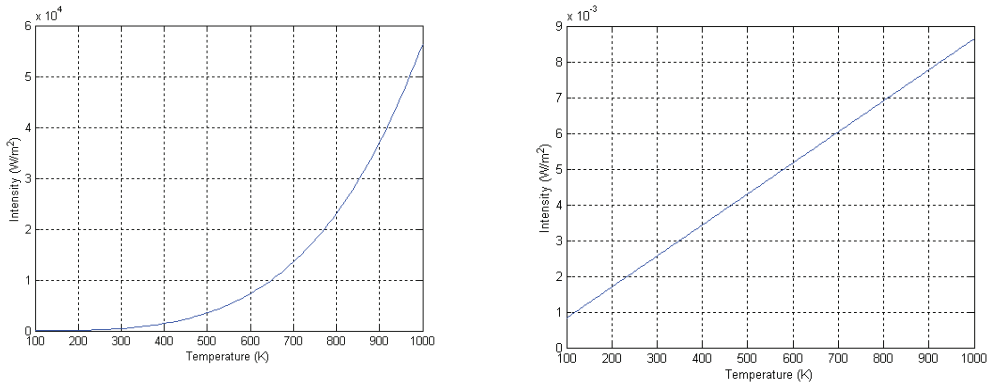


Figure 3.3: Total radiated power density with temperature (a) over the full bandwidth and (b) in the microwave region

From the shapes of the two graphs it can be seen that the total power density in W/m^2 is indeed a function of T^4 over the full band, and a function of T over the millimetre and microwave band.

3.3. Emissivity and Reflectivity

Different materials absorb radiation in different ways. Metals reflect most of the energy away so also do not absorb much radiation, while a black material (like soot) absorbs most of the incident energy and gets warm. To understand how this occurs, it is easiest to consider the interaction of the incident electromagnetic radiation with the charges (electrons) in the material.

Metal has electrons that are free to move through the entire solid, that is why they can conduct electricity. These free electrons also oscillate together in response to the electric field of an incoming light wave. By oscillating, they radiate electromagnetically, just like a current in an antenna. The radiation from the oscillating electrons is the "reflected" light. In this situation, little of the incoming radiant energy is absorbed; it is just re-radiated (reflected).

Soot will conduct electricity but not so well as a metal will. There are unattached electrons which can move about in the solid, but they keep bumping into things (they have a short mean free path). When they bump, they cause a vibration and so give up energy into heat. These free electrons are effective intermediaries in transferring energy from the radiation into heat. Soot will therefore have a lower reflectivity than metal.

Emissivity (ε): A measure of the ability of a body to radiate heat given by the ratio of the power radiated by the body per unit area, to the power radiated per unit area of a blackbody at the same temperature.

Reflectivity (ρ): The ratio of power reflected per unit area to the power incident per unit area. It is related to the emissivity by Kirchoff's law as follows $\rho = (1-\varepsilon)$.

Table 3.1: IR emissivity for various materials

Material	Emissivity ε
Skin*	0.98
Wet soil	0.95
Paint	0.94
Heavy vegetation	0.93
Dry soil	0.92
Dry grass	0.91
Sand	0.90
Dry snow	0.88
Asphalt	0.83
Oxidised Steel	0.79
Concrete	0.76
Polished Steel	0.07

* There is no reference to the skin colour

The primary difference between operating a radiometer in the IR band where the radiated power is proportional to T^4 and operating in the millimetre wave band where the power is proportional to T is that in the former, temperature variations across the target are dominant whereas in the latter, variations in emissivity and reflectivity are more important.

The Stefan-Boltzmann law is generally written to include the surface area of the object, A, and its emissivity, ε, to produce a value for the total radiated power (or flux)

$$\Phi = A\varepsilon\sigma T^4 \quad \text{W}, \tag{3.7}$$

where Φ – Total power radiated (W),
A – Surface area (m^2),
ε - Emissivity,
σ - Stefan-Boltzmann Constant (5.67×10^{-8} $Wm^{-2}K^{-4}$),
T – Temperature (K).

However, this formula specifies the power that would radiate from a body with a temperature T (Kelvin) towards an infinitely cold space.(0K).This, for most practical purposes is irrelevant, as most objects radiate into environments where the temperature is much higher than this, and the net power radiated by the body is reduced.

If a sensor with a temperature T_s and an emissivity ε_s faces an object with a temperature T_o and an emissivity ε_o, then the net power that flows into the sensor (which can be measured) is determined as follows

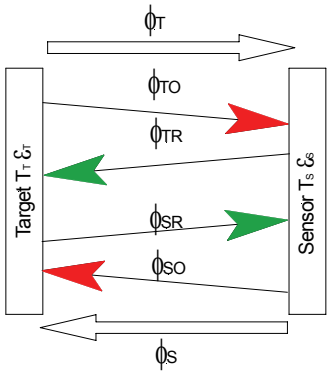


Figure 3.4: Flux balance between sensor and target

The flux radiated from the target towards the sensor is Φ_{TO}.

Some of this flux will be reflected $\Phi_{TR} = -\Phi_{TO}(1-\varepsilon_s)$

Leaving a net flux $\Phi_T = \Phi_{TO} + \Phi_{TR} = \Phi_{TO} - \Phi_{TO}(1-\varepsilon_s) = \Phi_{TO}\varepsilon_s$

Equating Φ_{TO} in terms of the emissivity, temperature and area of the target

$$\Phi_T = A_T\sigma\varepsilon_T T_T^4\varepsilon_S \quad \text{W} \tag{3.8}$$

Similarly, the net flux radiating from the sensor towards the target will be

$$\Phi_S = \Phi_{SO}\varepsilon_T \tag{3.10}$$

$$\Phi_S = A_S\sigma\varepsilon_S T_S^4\varepsilon_T \quad \text{W} \tag{3.11}$$

Because the two fluxes propagate in opposite directions, they are combined into a final net flux that flows between the two surfaces

$$\Phi = \Phi_T + \Phi_S = A\varepsilon_T\varepsilon_S\sigma\left(T_T^4 - T_S^4\right) \quad \text{W} \tag{3.12}$$

3.3.1. Example

Calculate the following

- The total power radiated by a human being radiating into cold space.
- The excess power radiated by a human being in a room at 20°C
- The power radiated by a human being over a 2GHz band at 100GHz

To determine the total power radiated by a human being over all wavelengths we can use the Stefan Boltzmann approximation

$$\Phi = A\varepsilon\sigma T^4 \quad \text{W} \tag{3.13}$$

The temperature of a healthy human being is 37°C (310K)

The body surface area (BSA) of a human being is generally determined empirically from their height and mass using Gehan's Formula.

Gehan's Formula
References :
Estimation of human body surface area from height and weight. Gehan E.A., Georges S.L. Cancer Chemotherapy Report;54:225-235 (1970)
Population used :
a total of 401 subjects from a previous study:
- 130 subjects were more than 20 years old
- 42 subjects were between 5 and 20 years old
- 229 subjects were less than 5 years old.

Method :
Measurement of BSA was realized by coating, surface integrator and triangulation. The model is identical in form to one proposed in 1916 by Dubois & Dubois ($BSA = a0H^{a1}W^{a2}$). Determination of coefficients was computed with a least square procedure.
Formula :
$BSA = 0.02350 \times height^{0.42246} \times weight^{0.51456}$
Units :
BSA (m^2), height (cm), weight (kg)

Using a height of 180cm and a mass of 80kg produces a BSA of $2m^2$

The emissivity is 0.98 as tabulated above

$\Phi = 2\times0.98\times5.67\times10^{-8}\times310^4 = 1026\text{W}$

This power loss could not be sustained if it were not for the compensating absorption of radiation from surrounding surfaces at room temperature which do not vary too much from the temperature of the body

If the surroundings are painted surfaces with an emissivity $\varepsilon = 0.94$ and a temperature of 20°C, then the net power radiated is calculated as follows:

$$\Phi = A\varepsilon_P\varepsilon_W\sigma\left(T_P^4 - T_W^4\right) \text{ W}$$

$$\Phi = 2\times0.980\times.945\times.671\times0^{-8}(\times310^4 - 293^4) = 195\text{W}$$

To determine the power radiated in the millimetre wave band, the Rayleigh-Jean approximation can be used, and integrated over the band of interest

$$\phi \approx A\varepsilon_P\int_{\lambda_1}^{\lambda_2}\frac{2\pi kcT}{\lambda^4}d\lambda = \frac{2\pi A\varepsilon_P kcT}{3}\left(\frac{1}{\lambda_1^3} - \frac{1}{\lambda_2^3}\right) \tag{3.15}$$

For $\lambda_1 = 3.031\times0^{-3}$ m and $\lambda_2 = 2.97\times10^{-3}$, the total radiated power equates to 11.7×10^{-6} W.

This result can be verified by writing a MATLAB function to perform the integration graphically and then scaling by the area and the emissivity of the body.

3.4. Detecting Thermal Radiation

There are three different basic interactions that can be exploited to detect thermal radiation. They are the following:

- External photoeffect
- Internal photoeffect
- Heating

Most IR detectors operate using quantum mechanical interaction between incident photons and electrons and the detector material. Photoconductive detectors absorb photons to elevate an electron from the valence band to the conduction band of the material, changing the conductivity of the detector. For this to happen, a photon must be sufficiently energetic to excite an electron. Photovoltaic detectors absorb photons to create an electron hole pair across a p-n junction which can produce a small current. Such devices can be manufactured as part of an array that includes a capacitor that stores a charge proportional to the incident radiation in a manner similar to that of a CCD.

Detectors with band gap energies small enough to respond to the longer wave IR radiation (8-12μm region) must be cooled to cryogenic temperatures between 77 and 25K to eliminate thermally generated carriers. To maintain these temperatures, the detector must be enclosed in a vacuum housing called a Dewar with a suitable window transparent at the required IR wavelengths.

Golay cells and capacitor microphones are pneumatic detectors. In Golay cells, the sealed xenon gas expands when it is warmed by incident infrared radiation. The resultant variation of pressure shifts a mirror located between a light source and a

photocell, varying the amount of light entering the photocell and thus changing the output of the photocell. In capacitor microphones, the varying expansion of the gas affects the capacitor film, which in turn produces the variation in the electrostatic capacity. These are extremely sensitive and broadband devices.

Micro bolometers absorb thermal energy over all wavelengths and do not require cryogenic cooling even when operating in the far IR. They are often used for low-cost commercial applications. Their main disadvantage is a very poor detectivity as can be seen from the figure below.

3.4.1. External Photoeffect

If light with a photon energy hf (in Joules) that exceeds the work function, W, falls on a metal surface, some of the incident photons will transfer their energy to electrons, which will then be ejected from the metal. Since hf is greater than W, the excess energy hf - W transferred to the electrons will be observed as their kinetic energy outside the metal. The relation between electron kinetic energy, E, and the frequency (that is, $E = hf - W$) is known as the Einstein relation, and its experimental verification helped to establish the validity of quantum theory. The energy of the electrons depends on the frequency of the light, while the intensity of the light determines the rate of photoelectric emission.

Work Function: Energy that must be supplied to the free electrons in the metal to enable them to escape from the metal

Table 3.2: Work functions of some common metals

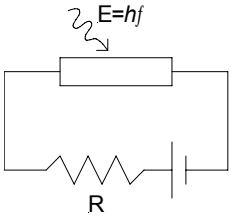
Metal	Work Function (eV)
Barium	2.5
Caesium	1.9
Copper	4.5
Potassium	2.2
Silver	4.6
Sodium	2.3
Tungsten	4.5

$1eV = 1.61 \times 0^{-19}$ J

3.4.2. Internal Photoeffect

The photon has sufficient energy to create a free electron, free hole or both in the material (usually a semiconductor)

Photoconductive Detectors



- Monitor voltage change across R
- Monitor current change through sample

In a semiconductor the valence-band of energy-levels is almost completely full while the conduction band is almost empty. The conductivity of the material derives from the few holes present in the valence band and the few electrons in the conduction band. Electrons can be excited from the valence to the conduction band by light photons having an energy h that is larger than energy gap E_g between the bands. The process is an internal photoelectric effect. The value of E_g varies from semiconductor to semiconductor.

The cutoff wavelength beyond which no emission will occur is

$$\lambda_c = \frac{hc}{E_g} \tag{3.16}$$

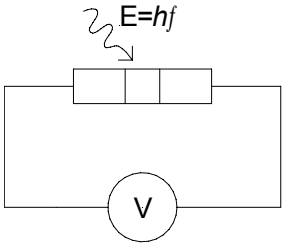
Table 3.3: Band gap and long wave cutoff of some materials

Material	Band Gap Energy E_g (eV)	Long Wave Cutoff λ_c (μm)	Operating Temp (K)
Si	1.09	1.1	300
Ge	0.81	1.4	300
PbS	0.49	2.5	77
InSb	0.22	5.5	77
HgCdTe	0.25	22	77
Ge:Hg	0.087	14	
Si:Ge	0.065	17	

$1eV = 1.6\times10^{-19}$ J

Visible radiation produces electron transitions with almost unity quantum efficiency ($\eta = 1$) in silicon. Each transition yields a hole-electron pair (i.e., two carriers) that contributes to electric conductivity. For example, if one milliwatt of light strikes a sample of pure silicon in the form of a thin plate one square centimetre in area and 0.03 centimetre thick, (which is thick enough to absorb all incident light), the resistance of the plate will be decreased by a factor of about 1,000. In practice, photoconductive effects are not usually as large as this, but this example indicates that appreciable changes in conductivity can occur even with low illumination.

Photovoltaic Detectors



- Measure open circuit voltage or short circuit current
- Sensitive region at junction
- Homojunctions: PN semiconductor
- Heterojunctions: Different materials but similar lattice spacing
- Metal – semiconductor interfaces include Schottky barrier detectors and GaAs photodiodes

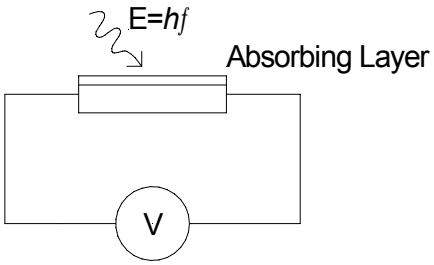
The photovoltaic effect consists in the generation of an electromotive force as a consequence of the absorption of radiation; that is to say, a current will flow across the junction of two dissimilar materials when light falls upon it. The primary effect is photo-ionisation; i.e., the production of equal numbers of positive and negative

charges. One or both charges can then migrate to a region in which charge separation can occur. This charge separation happens normally at a potential barrier between two layers of the solid material.

3.5. Heating

By lattice absorption resulting in increased vibration energy:

- Bolometers: Metal types have a +ve temp coefficient of resistance
- Semiconductor (thermistor) types have a –ve temp coefficient of resistance
- Pyroelectric Effect: Change in electrical polarisation



Pyroelectric materials are crystalline substances capable of generating an electrical charge in response to heat flow. They only operate in response to a change in temperature.

3.5.1. Difference between Photon and Thermal Detectors

Photon detectors measure the rate at which quanta are absorbed, they are sensitive to the frequency of the photons as can be seen in the figure showing D*. Thermal detectors measure the rate at which energy is absorbed and are insensitive to frequency over a wide range. However, they are generally less efficient than photon detectors.

3.6. Performance Criteria for Detectors

The key detector performance parameters are:

- Responsivity (R)
- Signal to Noise characteristics or noise equivalent power (NEP)
- Specific Detectivity (D^*)

3.6.1. Responsivity

If the probability that a photon with energy $E=hf$ will produce an electron is η, (quantum efficiency) then the average rate of production of electrons $\{r\}$ for an incident beam of optical power, P, is

$$\{r\} = \frac{\eta P}{hf}, \tag{3.17}$$

and the current in amps, will be the product of this and the electron charge, e,

$$i = \{r\}e = \frac{\eta eP}{hf} \quad \text{A} \tag{3.18}$$

In theory the output current is proportional to the input power making this a "square-law" detector. This is exploited when the diode is connected as a current-to-voltage converter, where the relationship between P and i tracks along the current axis ($V = 0$)

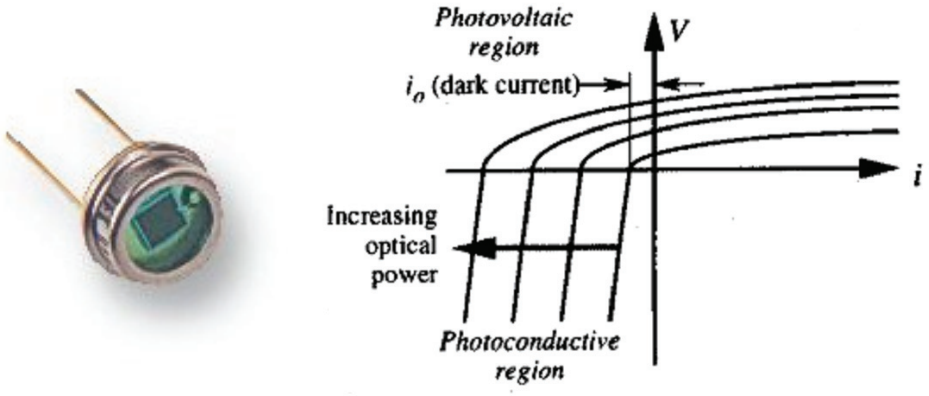


Figure 3.5: Photograph of a photodiode and its VI characteristic

The term i_o refers to the "dark current" which is a current that flows in the absence of light and is attributed to thermal generation of hole-electron pairs.

The Responsivity (R) is the ratio of the photo current (into a short circuit) generated for each watt of incident power. This can be derived from the equation above

$$R = \frac{i}{P} = \frac{\eta e}{hf} = \frac{\eta e \lambda}{hc} \quad \text{A/W} \tag{3.19}$$

Responsivity has an added wavelength dependence because the quantum efficiency is also a function of wavelength. At some frequencies, where the material (silicon) does not absorb strongly, photons may pass through the material or penetrate too deeply to produce photo ionisation close enough to the junction to be detected.

In more general terms, the responsivity is a measure of how well a detector reacts to incident radiation. Generally measured in amps per watt (or sometimes volts per watt)

$$R = \frac{I_{sig}}{P(\lambda) A_{\text{det}}} \quad \text{A/W} \tag{3.20}$$

where: R – Responsivity (A/W),
I_{sig} – Signal Level (A),
$P(\lambda)$ – Incident power (W),
A_{det} – Detector area (m^2).

The following figure shows the responsivity of a typical Silicon photodiode.

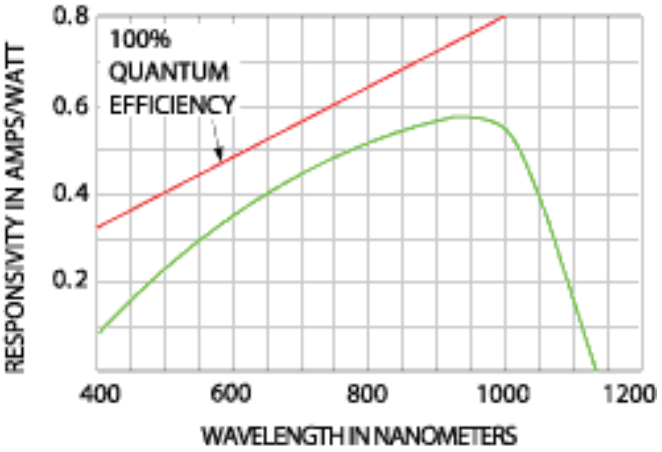


Figure 3.6: Responsivity of a silicon photodiode

3.6.2. Noise Equivalent Power (*NEP*)

This is defined as the power input to the detector that will create a signal to noise ratio of one. Noise is the random fluctuations in photocurrent generated as carriers are created and recombined. Dividing the RMS noise current of the detector by the responsivity gives the *NEP*.

$$NEP = \frac{I_{noise}}{R} \quad \text{W} \tag{3.21}$$

3.6.3. Detectivity

Reciprocal of *NEP*, $D = 1/NEP$ (W^{-1})

3.6.4. Specific Detectivity (*D**)

Because different detectors have different areas and signal bandwidths, a term *D** (dee star) was derived. It refers to a detectivity referred to a bandwidth of 1Hz and a detector area of $1cm^2$. A high detectivity value indicates a low-noise photodiode or detection system.

$$D^* = \frac{\sqrt{A_{\text{det}}\Delta f}}{NEP} \quad \text{cmHz}^{1/2}\text{W}^{-1}. \tag{3.22}$$

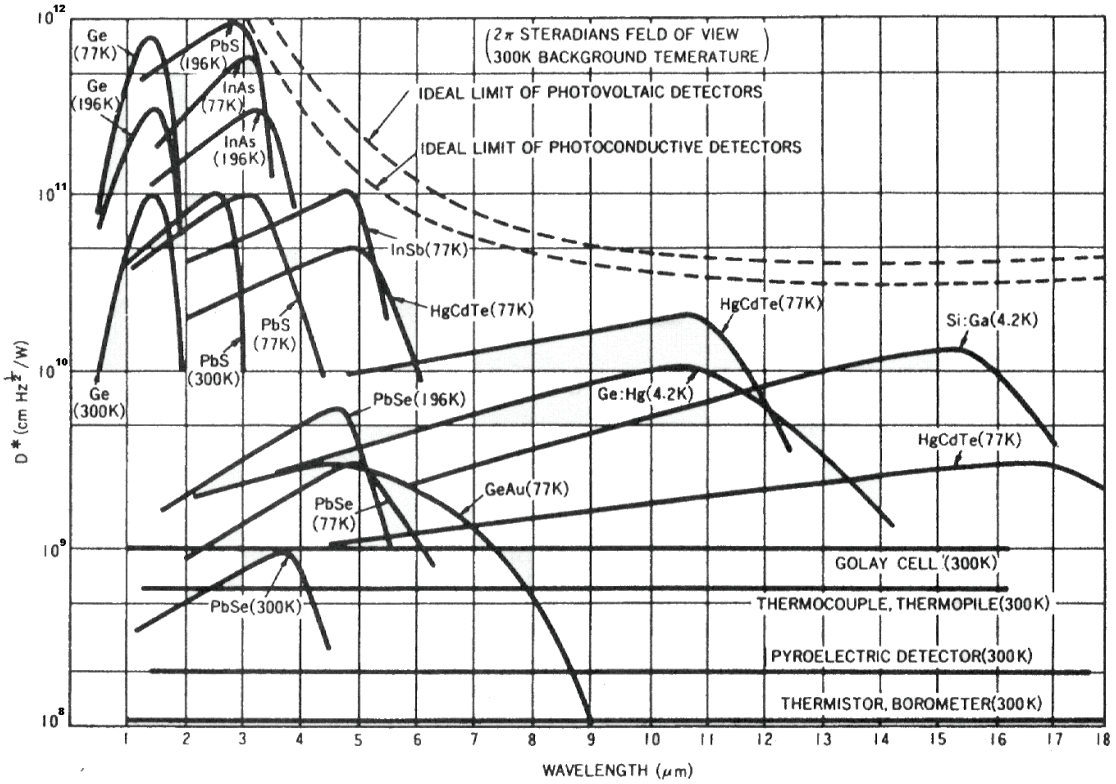


Figure 3.7: Operating ranges for some infrared detectors

3.7. Noise Processes and Effects

The following are the most common sources of noise that affect IR systems

- Johnson Noise: Random thermal motion of carriers. It is broadband and therefore expressed as noise per unit bandwidth

$$I_{johnson} = \sqrt{\frac{4kT}{R_{shunt}}} \quad \text{A.Hz}^{-1/2}\text{W}^{-1} \tag{3.23}$$

  where k is Boltzmann's constant, T the temperature in Kelvin and R_{shunt} is the shunt resistance

- Shot Noise: Fluctuations in the rates of thermal generation and recombination of carriers

$$I_{shot} = \sqrt{2e(I_{photo} + I_{dark})} \quad \text{A.Hz}^{-1/2} \tag{3.24}$$

  where I_{photo} and I_{dark} are the photo and dark currents respectively and e is the charge on an electron

- Surface and contact effects "1/f noise". This is not well understood and is empirically determined for individual detector families. It is only dominant at frequencies below 100Hz

Thermal detectors also exhibit temperature noise due to random temperature changes that arise from fluctuations of the rate of heat transfer from the detector to the surroundings and fixed pattern noise arising from element to element differences

There is a fundamental limit to a detectors performance due to "radiation" noise caused by statistical fluctuations in the incident background radiation.

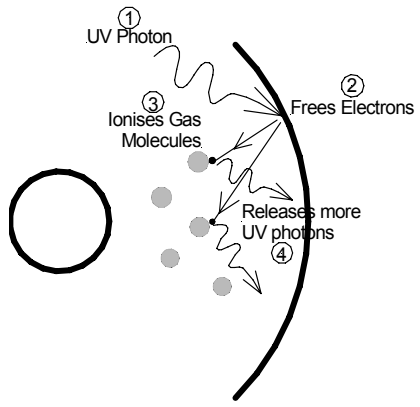
At room temperature, a signal would be swamped by thermally generated carriers in the detector. This can only be overcome by mounting the detector on the cold finger of a Dewar assembly. 8-12μm systems require cooling to 77K and 3-5μm systems require cooling to 193K

3.8. Applications

3.8.1. Passive Ultraviolet Sensor (External Photoeffect)

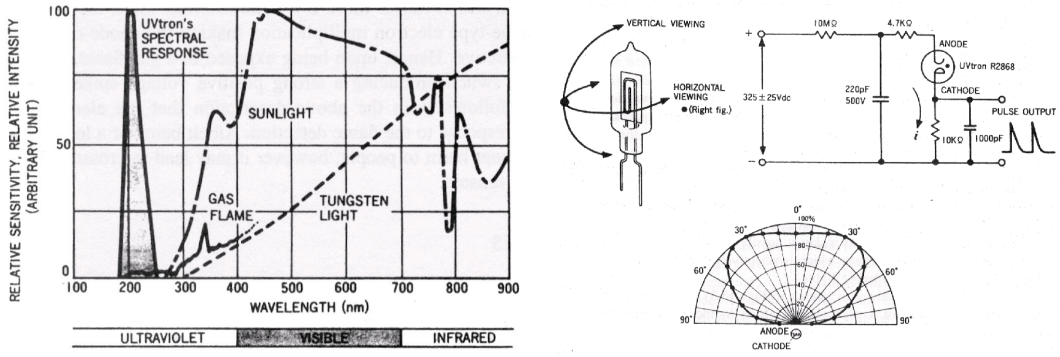
These sensors are used to detect missile plumes or gas flames because both produce spectra that extend into the UV below 200nm. Sunlight after passing through the atmosphere loses a large portion of its UV spectrum below 250nm.

The UVtron detector is a transparent gas filled tube with a high voltage applied across its electrodes. Upon being exposed to a plume, the high energy photons strike the cathode and release free electrons by the external photoeffect. The electrons gain kinetic energy as they are accelerated towards the anode by the potential difference between the two plates.



If an electron strikes a gas molecule with sufficient energy, the molecule will be ionised and release more UV radiation. This results in an avalanche multiplication effect and the tube becomes conductive. A relaxation oscillator circuit is used to produce a pulse train in the presence of UV radiation.

These detectors are more sensitive than smoke detectors for outdoor applications.



(a) **(b)**

Figure 3.8: Details of the UVtron (a) spectral response and (b) operational details

Obviously the cathode material is selected so that it is sensitive to ultraviolet light in the region below 250nm to be away from the residual UV present in sunlight, but not below about 200nm.

The lowest suitable work function for the cathode is easily calculated from the following formula

$$E = hf = \frac{hc}{\lambda} \text{ J} \tag{3.25}$$

A photon of light with a wavelength of 250nm will have an energy of

$E = 6.625\times10^{-34} \times 3\times10^{8} / 250\times10^{-9} = 7.95\times10^{-19}$ J

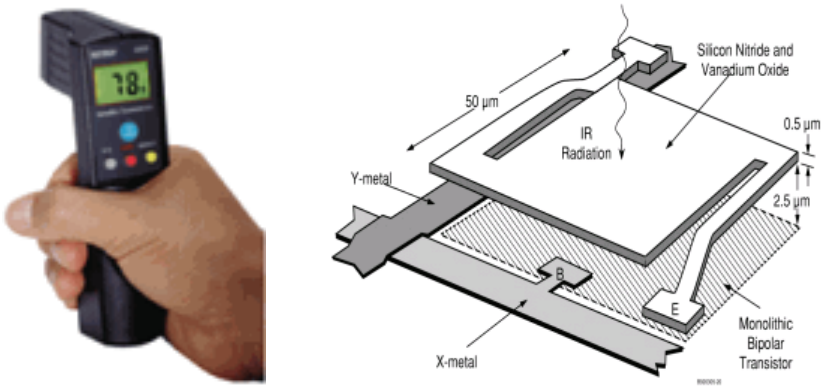
$E = 4.97$eV

It can be seen that none of the materials listed in Table 3.2 would be suitable, and that silver (E = 4.6eV) be the best of a bad lot with a sensitivity that extends up to a wavelength of 270nm.

3.8.2. Radiation Thermometer (Internal Photoeffect: Photoconductive)

Non-contact temperature sensor consists of a sensing element strongly responsive to IR radiation such as a Thermistor (micro bolometer). It has a low thermal conductivity support structure and electrical contacts which is hermetically sealed within a housing containing an inert gas and a transparent window

These thermometers usually assume a fixed emissivity, ε, which can make them inaccurate under certain circumstances



(a) (b)

Figure 3.9: Non contact thermometer (a) and a single microbolometer element used to measure the incident energy in the IR band

3.8.3. Passive Infrared Sensor (Internal Photoeffect: Pyroelectric)

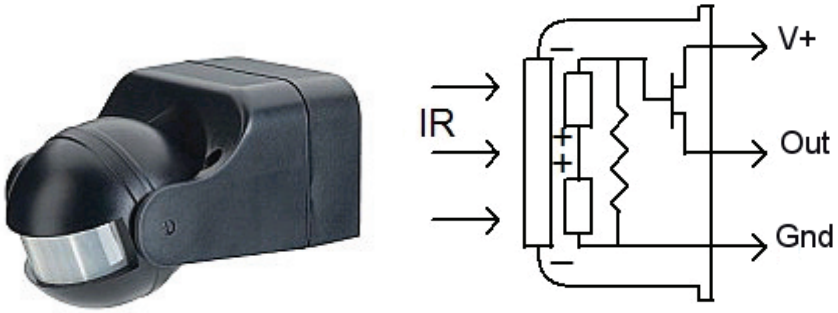


Figure 3.10: PIR sensor and schematic

Passive infrared (PIR) sensors are responsive to far infrared radiation within the spectral range from 4 to 20μm where most of the thermal power radiated by human beings is concentrated. An interesting pyroelectric material that is commonly used is a polymer film PVDF that, though not as sensitive as most solid state crystals, has the advantage of being flexible and low cost.

As an object moves across the sensor field of view, the image moves across the PVDF film moving from one pair of electrodes to the next generating an alternating current of the order of 1pA. A FET follower with an input impedance of about 50GΩ converts this to a 50mV signal that can be amplified and detected.

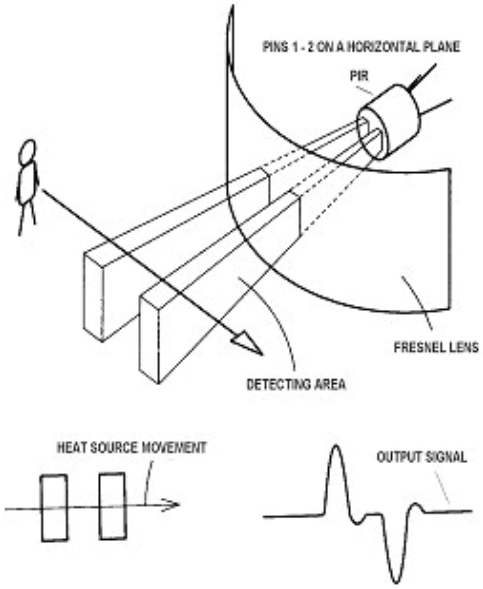


Figure 3.11: Operational principle of PIR sensor

The Fresnel facet lens and the PVDF film are often curved with the same radii to ensure that the surface of the film is always in focus. A Fresnel lens both captures more IR radiation and focuses it to a small point. This focal point moves across the sensor as the IR source moves and exposes one element at a time. A Fresnel lens can extend detection range of the sensor to about 30m.

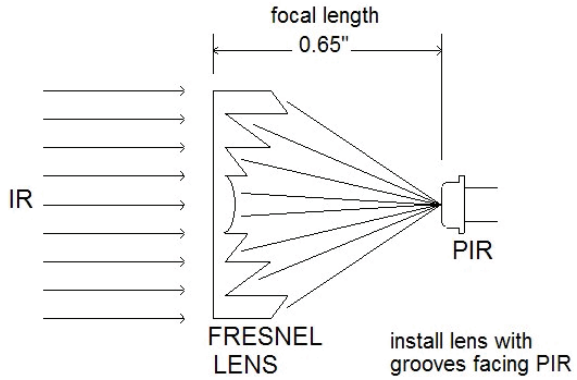
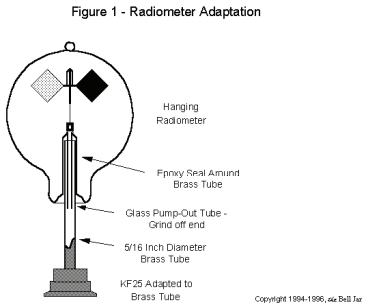


Figure 3.12: PIR sensor Fresnel lens

3.8.4. Crookes Radiometer



A Crookes radiometer consists of a glass bulb from which most of the air has been removed. The rotor mounted on a vertical support within the bulb bears four light horizontal arms to which are attached vertical metal vanes. Each vane has one side polished and the other side blackened.

When radiant energy (at any frequency) strikes the polished surfaces with high reflectivity, most of it is reflected away, but when it strikes the blackened surface with low reflectivity , most of it is absorbed, raising the temperature of those surfaces. The air near a blackened surface is heated and exerts a pressure on the blackened surface, causing the rotor to turn.

3.9. Introduction to Thermal Imaging Systems

Electro-optical thermal imagers, also referred to as forward looking infrared (FLIR) systems, thermal imaging systems (TIS) or infrared search and tracking (IRST) systems are passive devices which operate in the IR region of the frequency spectrum. They use the temperature gradient of the object to produce TV like images at night as well as during the day. They are often called night vision devices, but should not be confused with image intensifiers that use the same name.

Every object not at absolute zero (0K) emits electromagnetic radiation. The wavelength of the emitted radiation is a function of the temperature of the object. The short wavelengths between 0.45 and 1μm (the human eye sees between 0.46 and 0.65μm) used by visible cameras are emitted by very hot sources such as the sun or incandescent light bulbs, whereas the longer IR wavelengths between 3 and 14μm are emitted by objects at around 300K or 27°C.

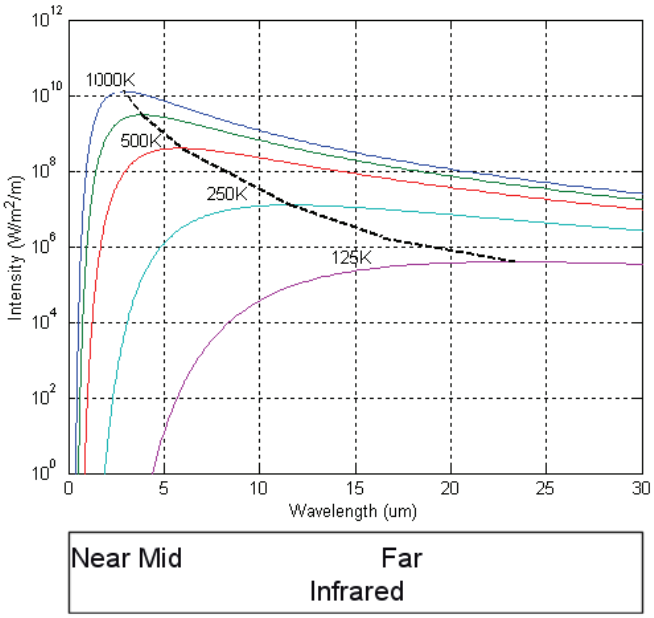


Figure 3.13: Planck's function in the infrared

Radiation is either scattered or absorbed as it propagates through the atmosphere. From the figure, note that there are three good IR transmission windows at 1-3µm, 3-5µm and 8-12µm, these windows dictate the choice of wavelengths used in IR sensor design.

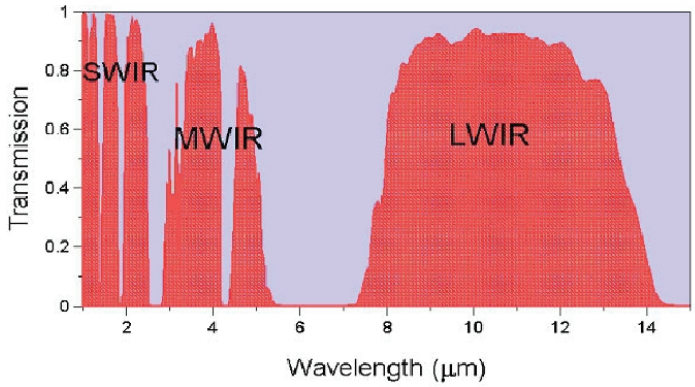


Figure 3.14: Transmission of infrared radiation through the atmosphere

The Infrared wavelengths generally used for missile seekers are influenced by target radiation characteristics, atmospheric attenuation and detector spectral response. Five bands are usually used spanning from 1.5 to 12.5µm.

Modern seekers generally use a combination of two bands to improve target detection probability and to minimise the effects of countermeasures (flares). To improve performance, one of the channels is often in the millimetre-wave band.

3.9.1. Thematic Imaging

Other bands are used for thematic mappers (both visible and IR) to enhance specific characteristics of the biosphere.

Table 3.4: Wavelengths used for specific applications by the Landsat thematic mapper

Band	Range (μm)	Characteristic
1	0.45 – 0.52	Coastal water mapping
2	0.52 – 0.60	Healthy vegetation
3	0.63 – 0.69	Chlorophyll absorption
4	0.76 – 0.9	Water delineation, vegetation vigour
5	1.55 – 1.75	Snow or cloud differentiation
6	10.4 – 12.5	Plant heat stress measurement
7	2.08 – 2.35	Hydrothermal mapping

3.10.Scanning Mechanisms and Arrays

Early thermal imagers called serial scan sensors, used either a single or a small array of detector elements preceded by horizontal and vertical mirror or prism based scanners to produce an image.

This configuration allows the small detector area to be cooled easily and effectively.

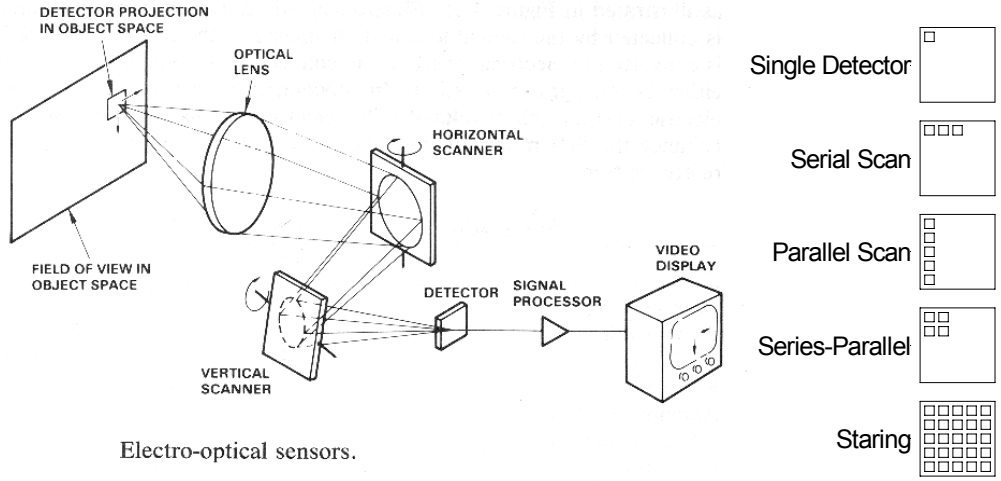


Figure 3.15: Scanning mechanisms for electro-optical sensors

In parallel scan sensors, a vertical column of between 112 and 180 detectors (1st generation sensors) or 240 to 960 detectors (2nd generation) are scanned horizontally using a single mirror. In these scanners, the column is often displaced by one pixel on alternate scans to produce an interlaced display

More recently un-scanned 2D staring arrays similar to a TV's charge-coupled device (CCD) that sample the scene image at the focal plane of the sensor optics have become available. These arrays are constructed of between 480 and 300,000 (often 320×240 pixels) detectors along with amplifiers, multiplexers and other supporting electronics on a single chip called a sensor chip assembly (SCA).

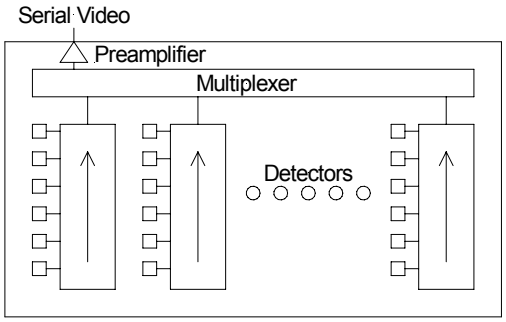


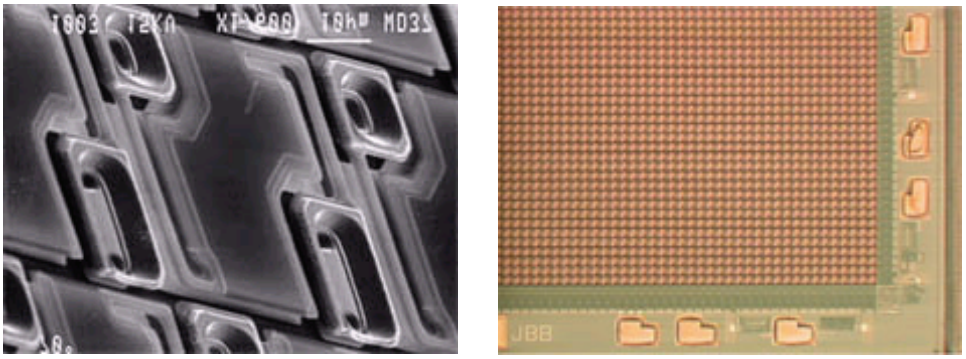
Figure 3.16: Focal plane array sensor chip assembly

Staring arrays (focal plane arrays) can either use high performance cooled sensors which are generally photovoltaic in nature and based on Mercury Cadmium Telluride (HgCdTe) or Platinum Silicide (PtSi), or uncooled microbolometer, Vanadium Oxide (VOx) thermistor based systems that offer reduced sensitivity, but much lower cost.

3.10.1. Microbolometer Arrays

The individual sensor elements in a microbolometer array use the change in electrical resistance of a VOx thermistor deposited onto the tiny "platelets" fabricated by silicon micro-machining in a silicon foundry. Incoming target radiation heats the VOx causing a change in electrical resistance, which is readout by measuring the resulting change in bias current. 80,000 and more sensors can be fabricated together into a two-dimensional array. The structure can be dimensioned to operate at 30 Hertz. That is, the thermal conductance of the isolating legs can be adjusted to match the time-constant for 30-hertz operation.

It consists of a two-layer structure. An interconnecting readout circuitry is applied to the silicon process wafer and then the microbolometer structure is built on top of the readout circuitry. First a pattern of islands are deposited on the readout circuitry. The islands are made of a material that can be selectively etched away later to form a bridge structure. Three layers -- silicon nitride, vanadium oxide, and silicon nitride -- are deposited over the sacrificial islands. The sacrificial islands are then etched away leaving the thermally isolated bridge structure of vanadium oxide. A photo of an early Honeywell microbolometer element is shown in the figure below followed by a photo of one corner of a 320 by 240 microbolometer array



(a) (b)

Figure 3.17: Scanning electron microscope image of a microbolometer element (a) and a photograph of part of an array (b)

Most of today's camera manufacturers use the 320 by 240 microbolometer array. However there is an excellent alternative for many commercial applications – the 160 by 120 array. The smaller array and its resulting camera can be produced at a much lower cost. Far more arrays can be produced on a single wafer and the yield is higher for the smaller array. In addition, one of the most expensive components of an infrared camera is the lens and its cost is proportional to the array size.

The only advantage of the larger array is field of view (FOV). With the same f# and focal length lens and the same detector size, a camera with 320 by 240 or 160 by 120 will have identical spatial resolution. But the target size for a fixed distance between the camera and target will be twice as large in both dimensions for the camera with the larger array. For many commercial applications the cost savings of the smaller array size over shadows the advantage of a larger FOV.

3.11.Key Optical Parameters

The following four parameters are key to describing the optical performance of an imaging system.

3.11.1. The Aperture Diameter (D_o)

The entrance pupil defines the limiting resolution of the optics as described by the Rayleigh criterion for resolution

$$\Phi_{\lim} = \frac{2.44\lambda}{D_o} \text{ rad} \tag{3.26}$$

where $\Phi_{\lim}$ is the smallest angular separation resolvable between two objects (diffraction limited resolution) (rad) and Do is the aperture diameter (m)

The reason that a scale factor of 2.44 is used (as opposed to the 1.22 used in radar) is because these sensors operate as receivers and do not transmit and receive.

3.11.2. F number (*f/#)*

This is the ratio of the limiting aperture and the focal length. It determines the light gathering capability or how much energy is focussed on the detector. The scene energy reaching the focal plane is $\propto [1/f/\#]^2$.

3.11.3. Focal length (f_1)

Distance from the optical centre (or pole) to the principal focus of a lens or curved mirror; $f_1 = (f/\#)D_o$

3.11.4. Transmission

Light is lost in an optical system through absorption and reflection. As light passes from one refractive medium to another, some portion of the energy is refracted at the medium boundary, and some absorbed within the medium.

$$\tau = \frac{(1-\rho)^2 e^{-\alpha x}}{1-\rho^2 e^{-\alpha x}} \tag{3.27}$$

where: $\rho = \left[\frac{(1-n)}{(1+n)}\right]^2$

x is the lens thickness (m), α the absorption coefficient of the optical material, ρ the surface reflectance and n the refractive index of the lens material

To increase the transmission of a lens, multilayer coatings are deposited on it to grade the refractive index change from air to the lens. Uncoated germanium (a common IR optical material) has a transmittance of about 47%. Using multilayer anti reflection coatings, this can be increased to 97%.

Table 3.5: Lens characteristics

Material	Transmission Band (μm)	Refractive Index
Germanium	2-20	4
Sapphire	0.4-5	1.63
Zinc selenide	0.5-20	2.4
Zinc sulphide	0.6-15	2.2

For mirrors, the transmission is the product of the reflectance of all the mirrors in the system.

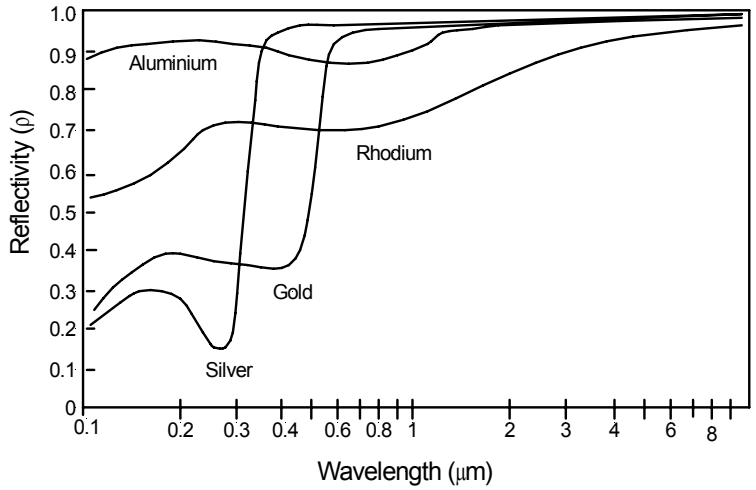


Figure 3.18: Spectral reflectance of some mirror coatings

3.11.5. Detector Area and Field of View

For a square detector, the area A_{det} is related to the instantaneous field of view (IFOV) α_d (sr) and the focal length f_1

$$\begin{aligned} A_{\text{det}} &= \left(\alpha_d^{1/2} f_1\right)\left(\alpha_d^{1/2} f_1\right) = \alpha_d f_1^2 \\ A_{\text{det}} &= \alpha_d \left[(f/\#) D_o\right]^2 \end{aligned} \tag{3.28}$$

3.12.System Performance Measures

3.12.1. Spatial Frequency

Spatial frequency is defined as the number of intensity cycles that exist in an image per milli-radian of subtended angle. It is one of the critical measures that define the performance of an imaging system.

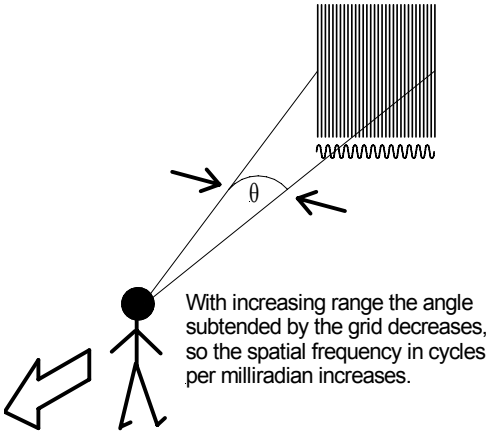


Figure 3.19: Definition of spatial frequency

Human visual perception is has a non-linear sensitivity to both the spatial frequency and contrast as shown in the following figure. This makes quantitative analysis of target detection quite difficult to achieve.

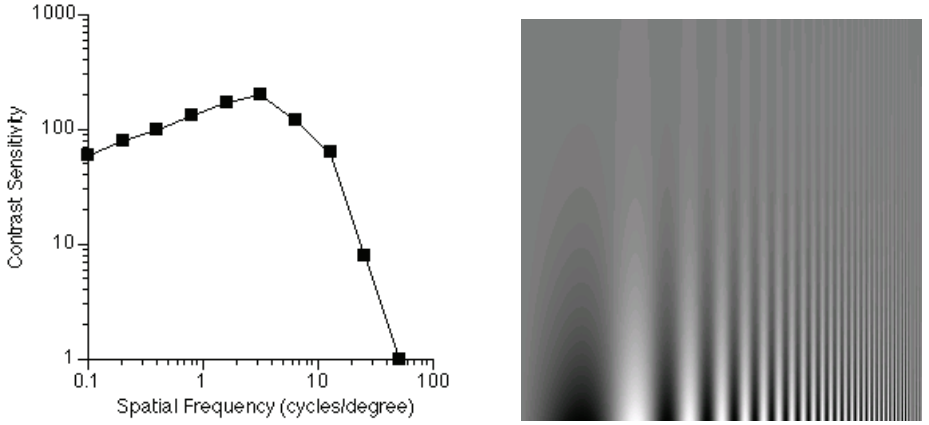


Figure 3.20: Human visual response to spatial frequency and contrast

3.12.2. Noise Equivalent Temperature Difference (NETD)

NETD measures the performance of the detector and the processing electronics. It is defined as the temperature difference that will produce a signal to noise ratio of unity.

3.12.3. The Minimum Resolvable Temperature (MRT)

MRT combines both the sensor and the operator characteristics into one measurement to define the sensors ability to display a structured target with small temperature differences.

The MRT can be calculated analytically, but because it is probabilistic in nature and because the ability of human operators is also considered, it is often determined experimentally by adjusting various spatial frequency bar patterns and noting the ability of a trained operator to resolve them.

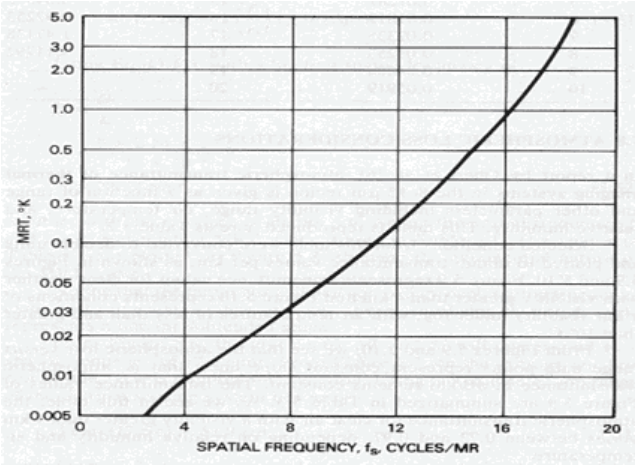


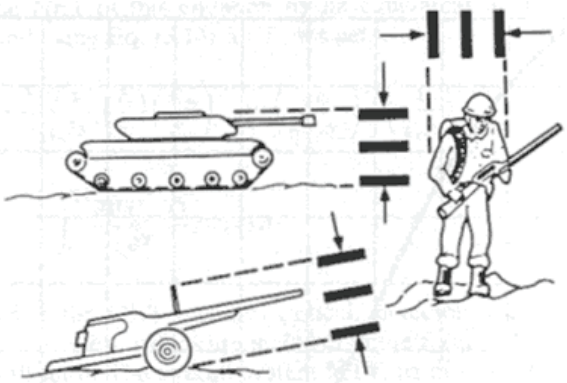
Figure 3.21: Measured MRT versus spatial frequency for a FLIR

3.13.Target Detection and Recognition

The Johnson criterion relates the number of resolution lines across a target critical dimension to the probability that the operator can detect or recognise the target. These criteria are listed in the following table as a function of the operator task and confidence. They are used along with the MRT curve to determine the maximum range at which detection (or recognition) will occur.

Table 3.6: Target detection criteria

Observer Task		Resolution Element (Half Cycles) per Target Dimension	
Night Vision & Electronic Sensors Designation	Naval Air Defence Centre Designation	Observer Confidence	
		0.5	0.95
Detector	Detect (required)	2	4
	Orient	3	
Recognise		7	14
	Recognised (classify)	8	16
	Identify	12	
Identify		13	32



Target	Resolution per minimum dimension in line pairs			
Broadside view	Detection	Orientation	Recognition	Identification
Truck	0.90	1.25	4.5	8.0
M-48 tank	0.75	1.20	3.5	7.0
Stalin tank	0.75	1.20	3.3	6.0
Centurion tank	0.75	1.20	3.5	6.0
Half-track	1.00	1.50	4.0	5.0
Jeep	1.20	1.50	4.5	5.5
Command car	1.20	1.50	4.3	5.5
Soldier (standing)	1.50	1.80	3.8	8.0
105 Howitzer	1.00	1.50	4.3	6.0
Average	1.0 ± 0.25	1.4 ± 0.35	4.0 ± 0.8	6.4 ± 1.5

Figure 3.22: Johnson criteria for detection

3.14.Example of FLIR Detection

To calculate the detection and recognition ranges for a tank target (3×3m) with a temperature difference of 2.5° using the measured MRT curve.

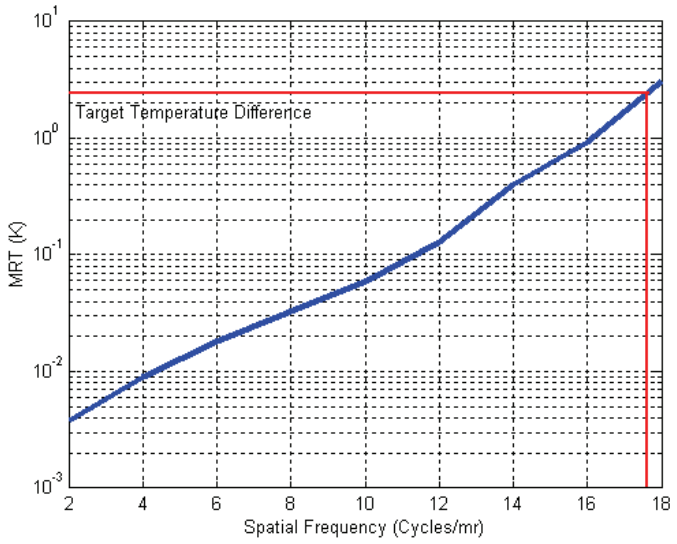


Figure 3.23: Measured MRT curve for a thermal imaging system

From the graph we can see that for an MRT of 2.5° the spatial frequency that can be resolved is 17.6 cycles/mrad. However, from the Johnson criteria, detection requires 1 cycle (2 half cycles), and recognition about 4 cycles.

Noting that the target has dimensions of 3m, each spatial frequency cycle corresponds to a range of 3km (where the target subtends one mrad).

The maximum range for detection and recognition is therefore

R_{det} = (3km)×17.6 = 52.8km

R_{rec} = (3km)×17.6/4 = 13.2km

These results do not include any atmospheric losses, however, even for clear-air with a visibility of greater than 4km, transmittance losses of between 0.72 and 0.97dB/km are experienced depending on relative humidity and air temperature. We use 0.9dB/km to for this example.

In fog with a visibility of 1km transmittance losses of between 3 and 4dB/km occurs. We use 4dB/km for this example

Because the MRT curve is experimental, the easiest way to solve for these additional losses is graphically. There is a linear relationship between detection range and spatial frequency resolution, so we can re-plot the MRT curve with R_{det} replacing Spatial Frequency with a scaling factor equal 3 (for the detection range calculation) and ¾ for the recognition range calculation.

The transmittance loss is then plotted on this graph as a function of range starting with the actual temperature difference of 2.5K.

$Loss = 2.5\times10^{dB/10}$

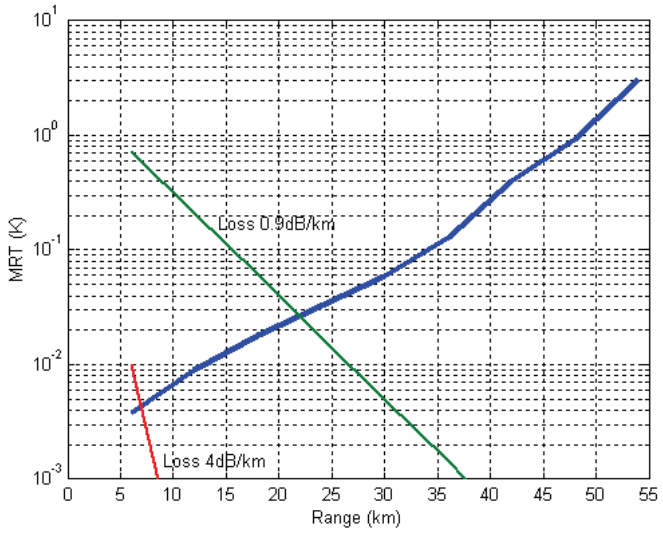


Figure 3.24: MRT curve scaled for a detection range of 52.8km

From the intersection of the loss-lines with the MRT graph it can be seen that the detection range in clear-air reduces to 22km, and in fog down to 7km.

In a similar manner the recognition range can be determined under good weather conditions and in foggy conditions.

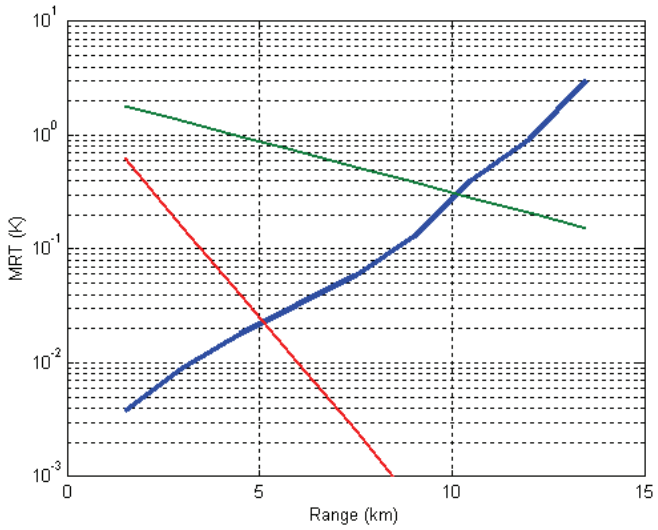


Figure 3.25: MRT curve scaled for a detection range of 13.2km

```
% FLIR DETECTION RANGE CALCULATIONS
xtar = 3;                       % target 3m across
% MRT curve
sfreq =(2:2:18);                % spatial frequency
mrt=([0.00384,0.00909,0.01812,0.03253,0.05919,0.12918,0.40287,0.92253,3.11792]);
semilogy(sfreq,mrt);
grid
xlabel('Spatial Frequency (Cycles/mr)')
ylabel('MRT (K)')
pause
% for a temperature difference of 2.5 kelvin,
% the resolvable spatial frequency is
sfdet = interp1(mrt,sfreq,2.5);
% for detection we need a spatial frequency of only 1 cycle at the detection range
%rdet = sfdet*xtar/1.0;         % detection
% for recognition we need a spatial frequency of 4 cycles at the detection range
rdet = sfdet*xtar/4.0;          % recognition
range=sfreq*rdet/sfdet;
dB1= -0.9*range;
loss1 = 2.5*10.0.^(dB1/10);     %clear air transmittance loss
dB2 = -4*range
loss2 = 2.5*10.0.^(dB2/10);     % fog 1km visibility transmittanceloss
% plot the scaled MRT curve and the transmittance loss
semilogy(range,mrt,range,loss1,range,loss2)
grid
%axis([0,15,0.001,10])          % detection
axis([0,15,0.001,10])           % recognition
xlabel('Range (km)');
ylabel('MRT (K)');
```

3.15.Thermal Images

The following figure shows a number of thermal images made using uncooled microbolometer based IR imaging systems.

Note particularly that the contrast of natural terrain is quite low because the temperatures of the various objects and their emissivities are similar. It is only objects like people and vehicles which are much hotter than the surroundings that stand out.

An interesting and unexplained phenomenon are the black throats of the chimneys at lower left.

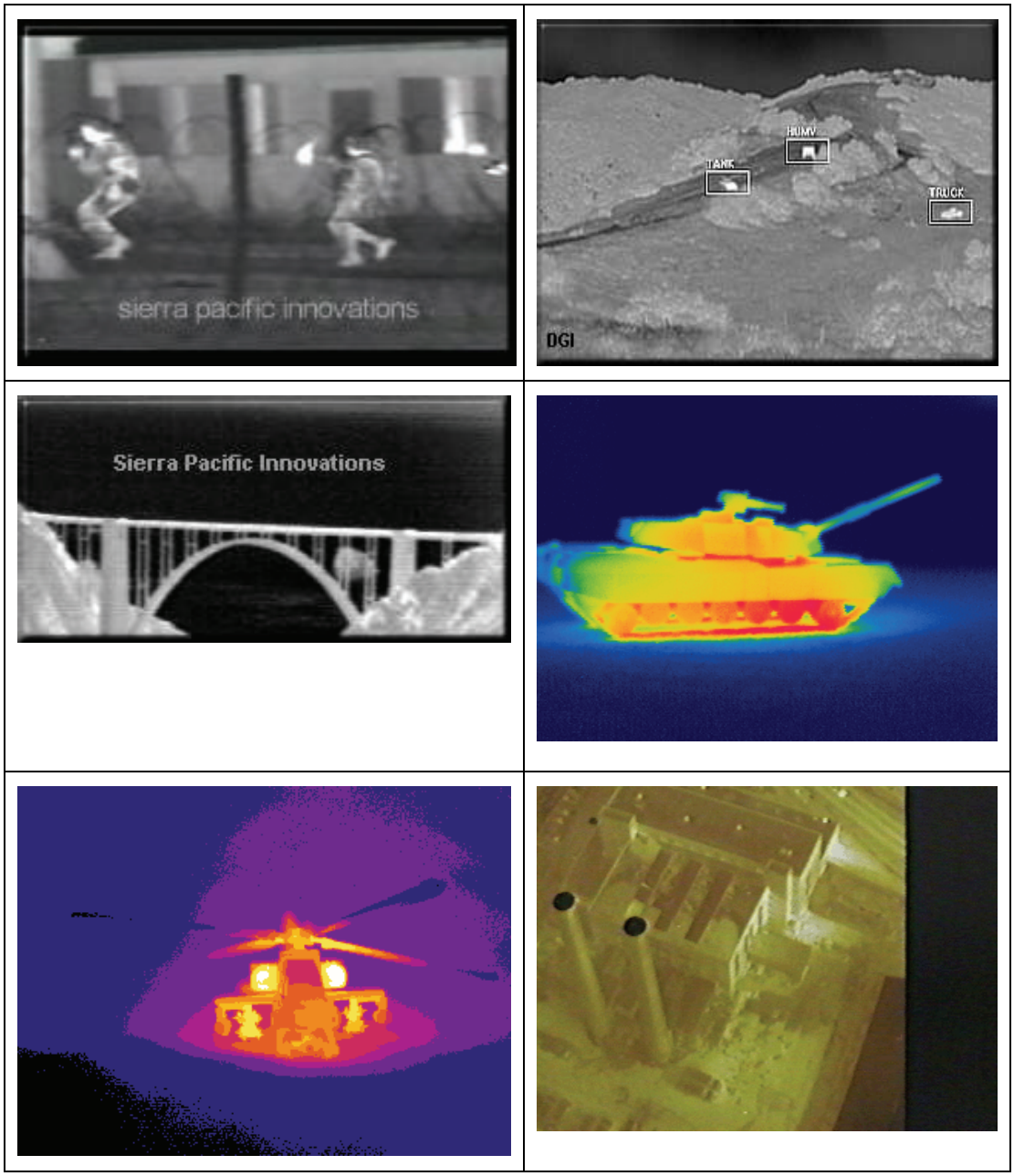


Figure 3.26: Examples of some thermal images from an uncooled FLIR sensor using micro-bolometer technology

3.16.Image Intensifiers

Image intensifiers are passive sensors that amplify ambient ultra violet, visible and near-infrared radiation. The scene is imaged onto a photocathode which has an energy-band structure such that on absorption of light, electrons are emitted from the surface. At best one electron can be emitted per photon.

3.16.1. First Generation Tubes

In first generation devices, these electrons are focussed and accelerated by an electric field before they impact on a luminescent screen. Because the accelerated electrons

posses increased energy, each electron gives rise to many photons (in the visible part of the spectrum) when it is absorbed by the phosphor. This results in an amplification or gain.

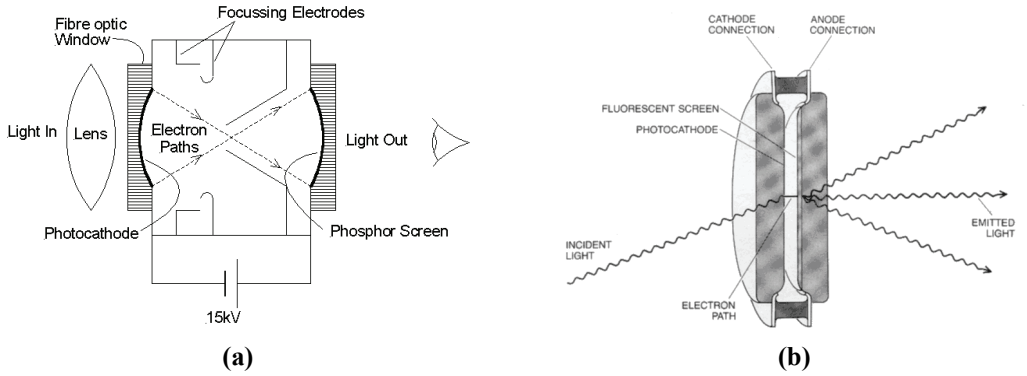


Figure 3.27: First generation (a) inverting tube and (b) non-inverting tube

To maintain crisp images, the distance travelled by each electron is limited, this limits the distance between the electrodes, and hence the allowed acceleration voltage is not very high.

First generation tubes feature especially high image resolution, a wide dynamic range (the ability to reproduce the ratio between the bright and dark parts of the image) and low noise.

Single stage devices offer gains of between 50 and 100 which gives satisfactory performance down to moonlight illumination levels. For starlight levels, gains of >50000 are required. This can be achieved by cascading multiple stages as shown in the figure.

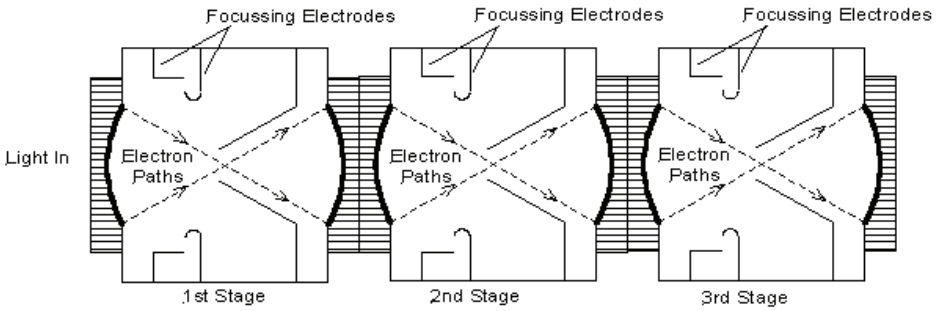


Figure 3.28: Cascading tubes to increase gain

Table 3.7: Natural illumination levels

Condition	Illumination (lux)		Condition	Illumination (lux)
Direct sunlight	10^5		Deep twilight	1
Full daylight	10^4		Full moon	10^{-1}
Overcast day	10^3		Quarter moon	10^{-2}
Very dark day	10^2		Clear starlight	10^{-3}
Twilight	10		Overcast starlight	10^{-4}

3.16.2. Second Generation Tubes

Second generation tubes use a micro channel plate (MCP) between the photocathode and the phosphor. The early MCP's consisted of about a million hollow glass tubes fused together into a disc. These tubes are about 10μm in diameter and 1mm long. The inside walls of the individual tubes are coated with a secondary emitting material and so act as electron multipliers. A single stage tube with a MCP can produce gains of up to 5×10^4.

Second generation devices offer poorer resolution and lower dynamic range than their first generation counterparts.

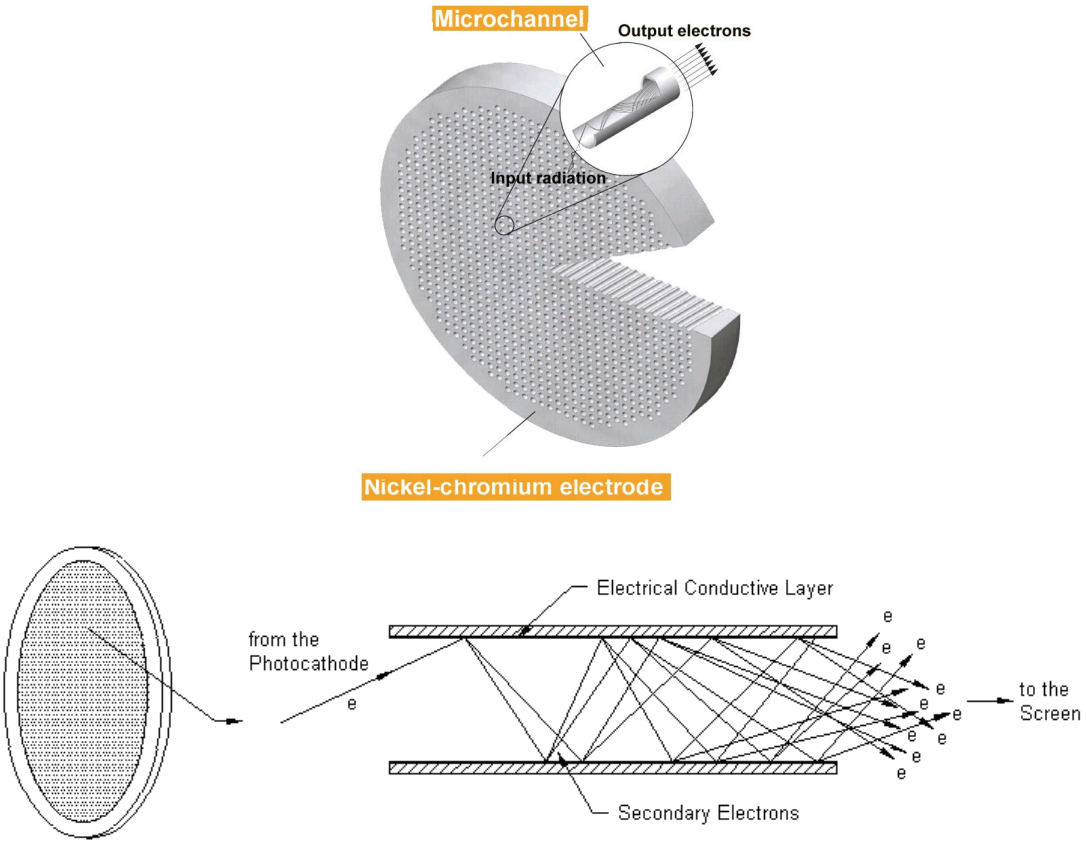


Figure 3.29: Micro channel plate schematic

Modern MCP's contain between 2 and 6 million holes. This number is a major factor in determining resolution.

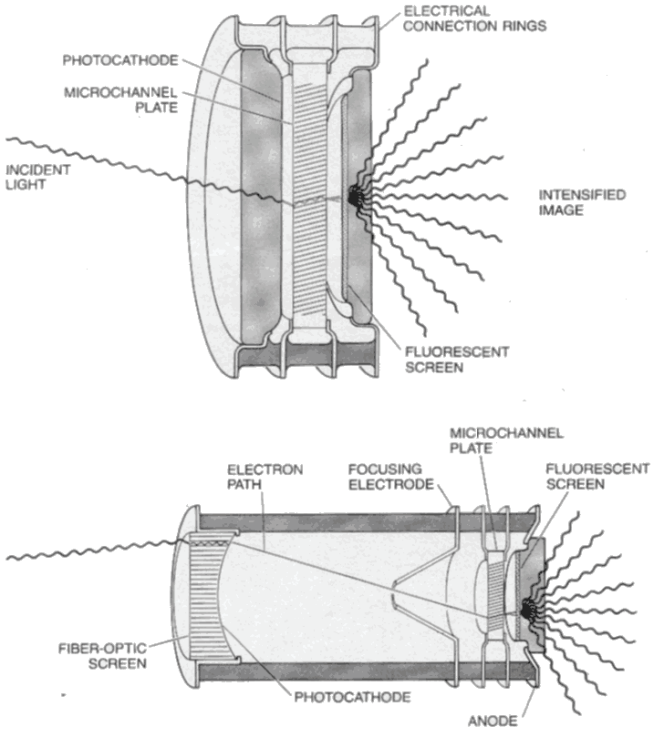


Figure 3.30: 2nd generation tubes using MCPs

In the configurations shown, the potentials applied to the MCP are intermediate to those of the anode and the cathode. The proximally focussed intensifier (upper diagram) is similar to a two electrode intensifier in as far as the gaps between the photocathode and the MCP and the MCP and screen are sufficiently short to minimise electron dispersion. In the focussed intensifier case, the electrons emitted by the photocathode are focussed by and electric field onto the MCP

Limitations of Microchannel Plates

Four major physical constraints limit their performance:

- Average output signal cannot exceed the maximum current that can be sustained within the walls of the microchannel. When the electron flux is too large, electric charge removed from the glass is not replaced immediately and the gain is reduced. Bright parts of an image can saturate and lose contrast.
- If one of the molecules of gas that remain in the tube becomes ionised, it becomes accelerated towards the input of the tube where it could strike the wall and initiate a new cascade of electrons. This cascade can mask the signal. Curved microchannel paths minimise this effect.
- The charge density of electrons in the tube is limited to about 10^7 per mm before mutual electrostatic repulsion returns additional secondary electrons to the surface of the channel before the field can accelerate them.
- The surface area of the channels is less than the surface area of the entire plate. Geometric constraints limit this to a maximum of 91%, but because the walls must be a finite thickness, this is generally worse (typically about 55%). This means that about half of the input flux hits the area between the channels. Channels can be made funnel shaped to minimise this effect.

3.16.3. Third Generation Tubes

Third generation tubes are similar in construction to 2nd generation devices but have a GaAs photocathode which is more sensitive and extends into the near infrared (NIR) band (450 to 950nm) as shown in the figure below.

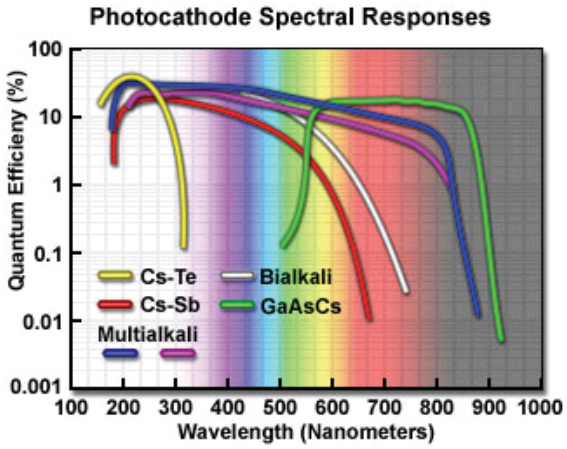


Figure 3.31: Photocathode response comparison

3.16.4. Spectral Characteristics of the Scene

Though the spectral content of sky illumination for sunlight and moonlight peaks in the visible region, the spectral content of clear starlight is weighted towards the NIR. Hence, for very low light applications, photo-cathodes are made sensitive in this region.

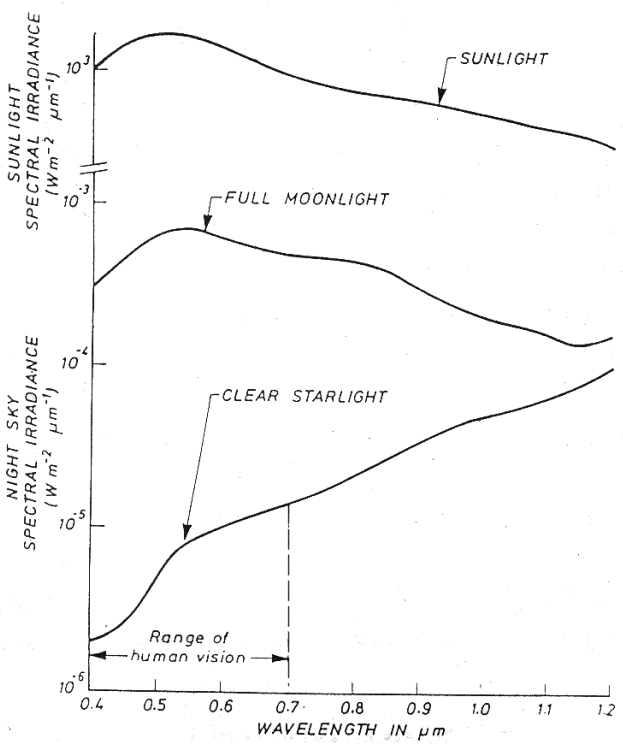


Figure 3.32: Spectral characteristics of outdoor scenery under different lighting conditions

Another advantage of operating at NIR is that vegetation exhibits higher reflectivity above 800nm and hence target background contrasts can be larger.

The following figure shows an image produced by a 3rd generation tube which illustrates the good contrast with foliage that can be achieved (left) and good dynamic range (right)

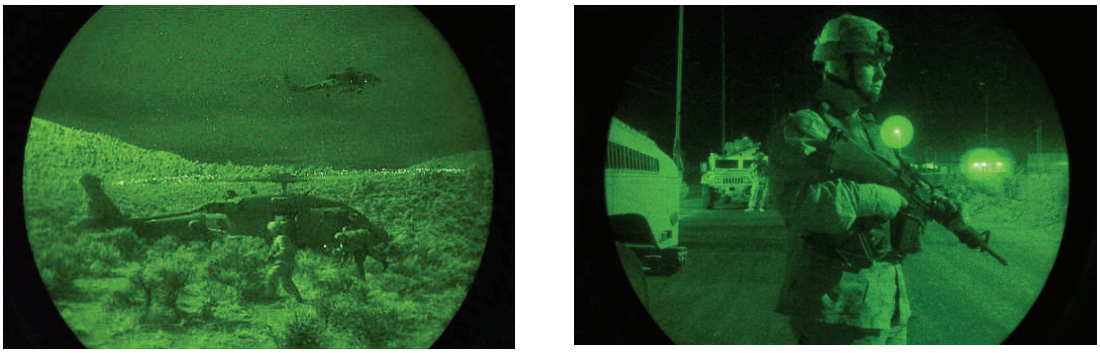


Figure 3.33: Image intensifier images showing the good resolution and high gain available from modern 3rd generation night vision systems

3.16.5. Off the Shelf



Figure 3.34: Night vision for everyone

3.16.6. Time Gating MCPs

Because MCPs offer huge amplifications, it is possible to operate these devices over very short illumination periods.

One useful technique that has been developed to exploit this is to illuminate a scene with a short burst of laser light (typically <10ns) and then time gating the MCP for the same period but delayed by a fixed length of time. This will capture the reflections from a 1.5m slice of the scene at a range determined by the selected delay.

Underwater imaging through turbid conditions minimises the backscatter from suspended particulates except at the range of interest and results in cleaner images.

Because the human body is fairly transparent in the infrared, this technique is being used to image its internal structure, albeit with much higher gating speeds (typically <200ps).

3.17.References

[1] F.J.Crawford, Electro-Optical Sensors Overview, *IEEE AES Systems Magazine,* Oct 1998
[2] D. W Craig, RSAF Platform Engineers Course: Electro-Optics Module., DSTO, August 1994
[3] J. Fraden, *Handbook of Modern Sensors,* Air Press, 1996.
[4] S.A.Hovanessian, *Introduction to Sensor Systems,* Artech House 1988.
[5] SPI Library, http://www.x20.org/library/thermal/infrared.html, 17/02/2000
[6] Boeing Flir, http://www.boeing.com/defence-space/infoelect/flir, 24/07/2000
[7] Laser Beam Measurement http://beammeasurement.mellesgriot.com/tut_photo_det.asp
[8] F.J.Crawford, Electro-Optical Sensors Overview, *IEEE AES Systems Magazine,* Oct 1998
[9] SPI Library, http://www.x20.org/library/thermal/infrared.html, 17/02/2000
[10] http://www.phy.hw.ac.uk/~peckham/envphy_4/basic_principles!/black_body.html, 09/02/2001
[11] http://www.phys.virginia.edu/classes/252/black_body_radiation.html, 09/02/2001-02-24
[12] http://www.flir.spiral8.com/recource_ctr/theory_of_thermography.html, 13/02/2001
[13] A.J.Sparius, Electro-optical Imaging Target Trackers, *Transactions of the S A Institute of Electrical Engineers,* November 1981.
[14] Jacob Fraden, *Handbook of Modern Sensors 2nd Ed.,* AIP Press, 1996
[15] Duncan W Craig, RSAF Platform Engineers Course: Electro-Optics Module., DSTO, Aug 1994
[16] E Uvarov, *A Dictionary of Science*, Penguin, 1964
[17] Encyclopaedia Britannica, Standard CD, 2000
[18] Pyroelectric Infrared Sensors, www.glolab.com/focusdevices/focus.shtml
[19] Microbolometer Technology, www.infraredsolutions.com/html/technology/mic...
[20] J. Fraden, *Handbook of Modern Sensors 2nd Ed.,* AIP Press, 1996
[21] D. W Craig, RSAF Platform Engineers Course: Electro-Optics Module., DSTO, August 1994
[22] Encyclopaedia Britannica, Standard CD, 2000
[23] Sunday Telegraph (Sydney), 04/02/2001
[24] M.Lampton, The Microchannel Image Intensifier, *Scientific American,* November 1981
[25] http://www.x20.org/nightvision/nightvisiontheory.html, 17/02/2000.
[26] D. Appell, Seeing Stars with Digital Eyes, *IEEE Spectrum,* July 2001
[27] Introduction to Image Intensifier Tubes, http://www.proxitronic.de, 20/01/200

Chapter 4.

Millimetre Wave Radiometers

4.1. Thermal Emission

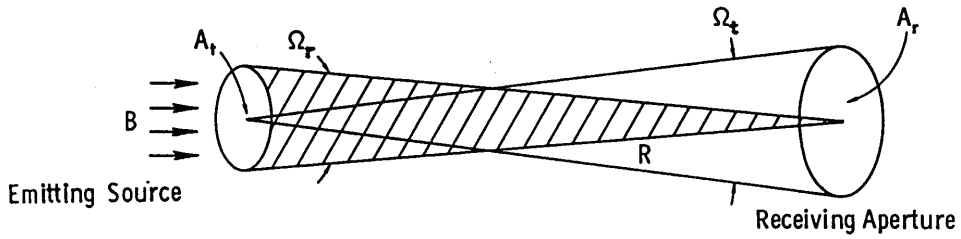
The Rayleigh-Jeans approximation for Planck's law can be written as follows

$$B_f = \frac{2kT}{\lambda^2} \quad \text{W/m}^2\text{/Hz/sr.} \tag{4.1}$$

This approximation is accurate to within 1% for frequencies below 100GHz, and to within 3% for frequencies below 300GHz.

4.2. Antenna Power Temperature Correspondence

Consider a lossless antenna with area A_r and beamwidth Ω_r at a distance R from a source of radiation with an area A_t. Assuming R is sufficiently great that the power density S_t is uniform over the solid angle Ω_t.



The power P, intercepted by the antenna is

$$P = S_t A_r \quad \text{W}, \tag{4.2}$$

where P – Received power (W),
S_t – Target power density (W/m^2),
A_r – Receiver Antenna Aperture (m^2).

The power can also be expressed in terms of the radiation intensity as

$$P = \frac{F_t A_r}{R^2} \quad \text{W}, \tag{4.3}$$

where F_t is the radiation intensity (W/sr).

The brightness can then be defined as

$$B = \frac{F_t}{A_t} \quad \text{W/m}^2\text{/sr}, \tag{4.4}$$

and the solid angle Ω_t subtended by the source of the radiation is given by

$$\Omega_t = \frac{A_t}{R^2} \quad \text{sr}. \tag{4.5}$$

Substituting into the power equation

$$P = BA_r\Omega_t \quad \text{W}. \tag{4.6}$$

For a differential solid angle

$$\partial P = A_r B(\theta,\phi) F_n(\theta,\phi)\partial\Omega, \tag{4.7}$$

where $B(\theta,\phi)$ – Source brightness as a function of solid angle (W/m^2/sr),
$F_n(\theta,\phi)$ – Normalised radiation pattern of antenna as a function of solid angle,

If this is integrated over all 4π steradians and over the frequency band f_1 to f_2,

$$P = \frac{A_r}{2}\int_{f_1}^{f_2}\iint_{4\pi} B(\theta,\phi)F_n(\theta,\phi)\partial\Omega\partial f \quad \text{W}. \tag{4.8}$$

This allows for the calculation of the power incident on the antenna in terms of the brightness of the source of the radiation and the gain pattern of the antenna.

This received power is reduced by one half in this case because the direct polarisation from the source is random and it is being received by a linearly polarised antenna.

Considering that this antenna is placed within a blackbody, and if the detected power is limited to a small bandwidth such that the brightness is constant with frequency, then the Rayleigh-Jeans approximation can be substituted for $B(\theta,\phi)$ to obtain the received power

$$P_{bb} = \frac{kT(f_2 - f_1)A_r}{\lambda^2}\iint_{4\pi} F_n(\theta,\phi)\partial\Omega \quad \text{W}. \tag{4.9}$$

From basic antenna theory it can be shown that the integral above equates to the pattern solid angle Ω_p which is given by

$$\Omega_p = \frac{\lambda^2}{A_r}. \tag{4.10}$$

This is substituted into the power equation to give the fundamental equation of radiometry

$$P_{bb} = kT(f_2 - f_1) \quad \text{W}. \tag{4.11}$$

Certain points are worth noting here:

- The detected power is independent of the antenna gain because the source of radiation is extended and uniform and not a point source
- The equation is independent of the distance from the radiating target
- The temperature of the antenna structure has no effect on the output power
- Temperature and power are interchangeable so all the gain calculations can be applied directly to the measured "temperature"
- The power detected is directly proportional to the bandwidth.

Example

Consider the power received by an antenna operating at 100GHz with a bandwidth of 2GHz observing a blackbody with a temperature 310K

$$P = 1.38\times10^{-23}\times310\times2\times10^{9} = 8.56\times10^{-12} \text{ W}$$

$$= -80.68\text{dBm}$$

4.3. Brightness Temperature

$T_b(\theta,\phi)$ is defined as the brightness temperature of the thermal source B.

All real bodies are to some extent "grey", as they radiate less than a black body. In addition the brightness temperature, $T_b(\theta,\phi)$ for a grey body can also angle dependent because of variations in its emissivity.

T is defined such that the brightness of the grey body is the same as a blackbody at the brightness temperature. It can be obtained from the physical temperature

$$T_b(\theta,\phi) = \varepsilon(\theta,\phi)T, \tag{4.12}$$

where $\varepsilon(\theta,\phi)$ – Emissivity,
T – Physical temperature of the radiating element (K).

In the example above, if the target has an emissivity $\varepsilon = 0.8$, then the brightness temperature $T_b = 0.8\times310 = 248$K, and the received power is reduced accordingly (-81.64dBm).

4.4. Apparent Temperature

In radiometry the apparent antenna temperature T_{AP} replaces the received power W as the measure of signal strength, where T_{AP} is defined as the temperature of a matched resistor with noise power output equal to W; that is $W = k.T_{AP}$ at the antenna port

The apparent antenna temperature T_{AP} is calculated from the brightness temperature including atmospheric and antenna losses.

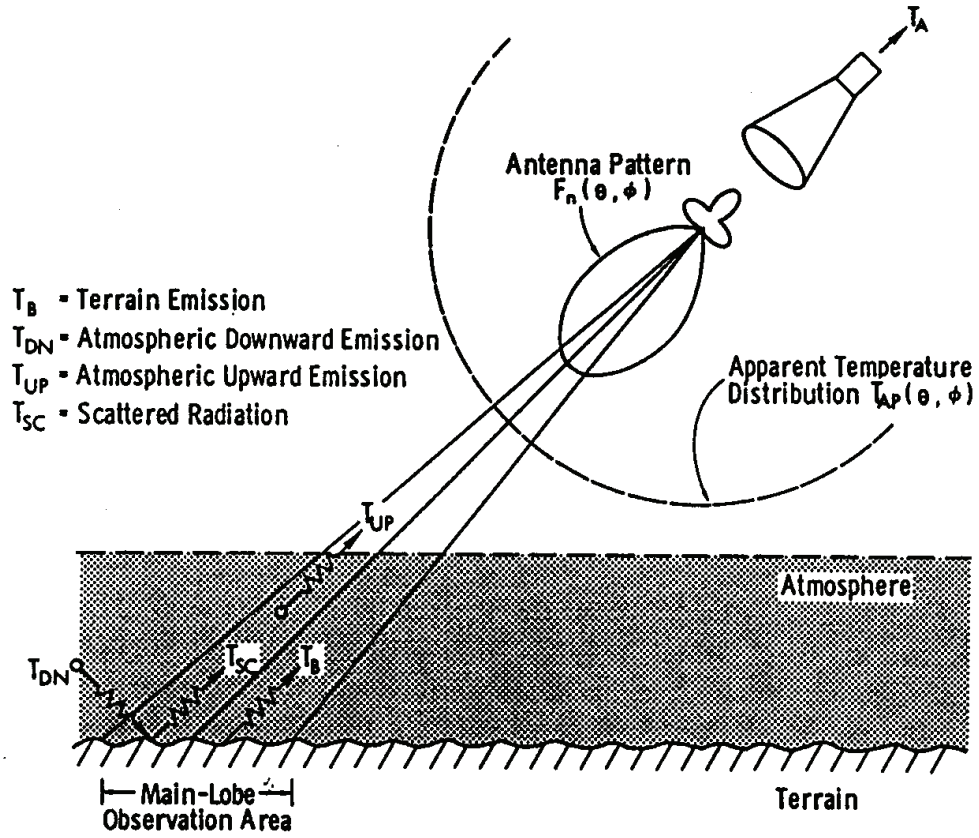


Figure 4.1: Radiometer configuration showing effects

The radiation from the main lobe of the antenna is made up of two components:

- The brightness temperature T_B from terrain emissions
- The scatter temperature T_{SC} which is the radiation reflected from terrain in the main lobe but not generated by it. Radiation from both the atmosphere T_{DN} (the downward or downwelling temperature) and galactic radiation may be reflected. At frequencies greater than 10GHz only the downward radiation from the atmosphere need be considered.

These contributors to the total radiation are then attenuated by the atmosphere before they reach the antenna.

In addition to this there is the upward (or upwelling) radiation from the atmosphere.

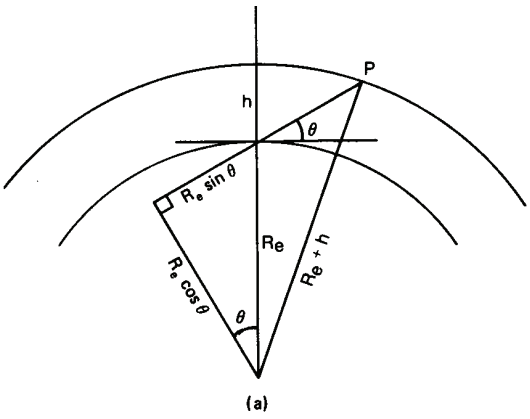
$$T_{AP}(\theta,\phi) = T_{UP}(\theta,\phi) + \frac{1}{L_A}\left[T_B(\theta,\phi) + T_{SC}(\theta,\phi)\right], \tag{4.13}$$

where T_{AP} – Apparent Temperature (K),
T_{UP} – Upwelling temperature from the atmosphere (K),
L_A – Atmospheric loss factor,
T_B – Brightness of the observation area (K),
T_{SC} – Brightness of the radiation scattered from observation area (K).

4.5. Atmospheric Effects

4.5.1. Attenuation

Atmospheric attenuation is a function of the air density, and, for horizontal or oblique paths through the atmosphere, it must be calculated by integration. The graph below shows the attenuation right through the atmosphere



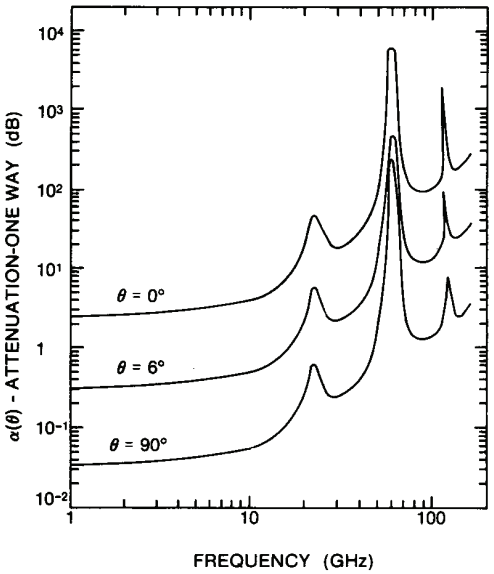


Figure 4.2: Attenuation through the atmosphere

The upward and downward brightness temperatures of the atmosphere vary with frequency, and will obviously be higher where the attenuation is higher as the atmosphere is more opaque.

At 94GHz, the attenuation through the entire atmosphere can be calculated as follows:

$$L_A = 0.17 + 0.06\rho_o, \tag{4.14}$$

where L_A – Atmospheric attenuation (dB),
ρ_o – Water vapour concentration (g/m^3).

For aircraft based radiometers, the attenuation is far more complex, and will be dealt with in some detail later in this chapter.

4.5.2. Downwelling Radiation

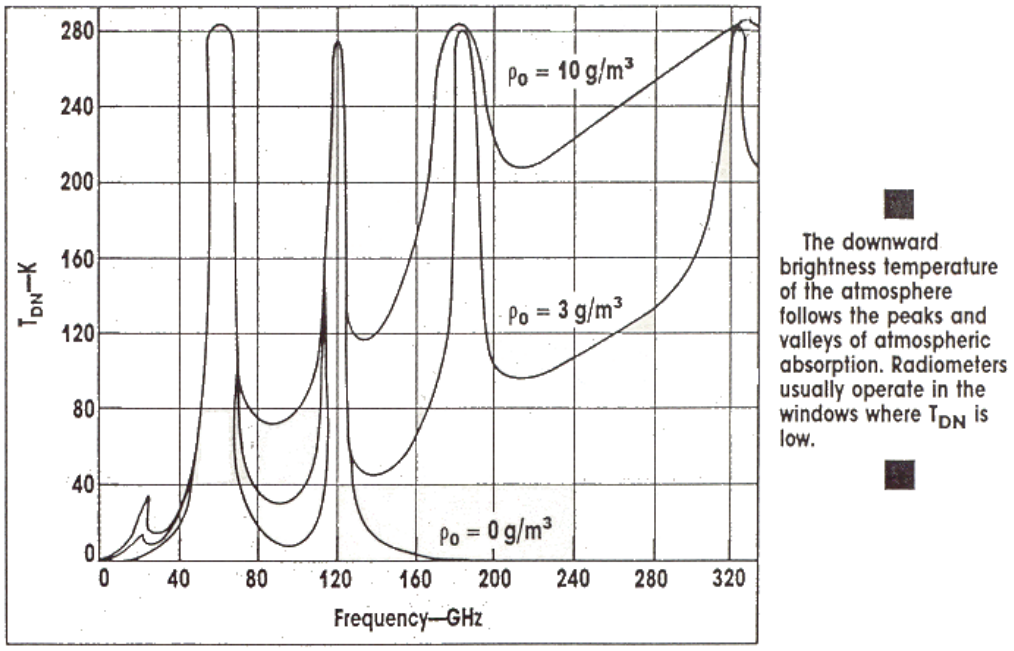


Figure 4.3: Downwelling brightness temperature as a function of frequency with water vapour concentration as a parameter

For operation at 94GHz, typical values of the downwelling temperature as a function of the atmospheric conditions are shown in the following table

Table 4.1: Downwelling temperature under different weather conditions

Conditions	Downwelling Temperature (K)
Clear Sky	10-60
Thick Fog	120
Overcast	150
Fog	180
Thick Clouds	180
Moderate rain	240

4.5.3. Upwelling Radiation

For space borne radiometers the upwelling radiation is that of the entire atmosphere and is equal to the downwelling radiation

For aircraft, only part of the atmosphere contributes to the upwelling radiation, and as the atmosphere can be stratified and complex, it is easiest to treat it as an attenuator and calculate the upwelling radiation in those terms.

To maintain thermal equilibrium, any medium that absorbs radiation (attenuates) must also radiate. As the atmosphere can be modelled as an attenuator, it can be shown that its effective temperature is

$$T_e = (1 - \frac{1}{L_A})T \,, \tag{4.15}$$

where T_e – Effective temperature of attenuator (atmosphere),
L_A – Attenuator loss factor = $10^{\alpha/10}$,
T – Physical temperature of the attenuator (K).

4.6. Terrain Brightness

Various forms of terrain have completely different brightness temperatures

Metallic Objects: These are lossless and opaque and so are perfectly reflecting. As a result their brightness will be the same as the downwelling radiation.

Water: The brightness of water is dependent on polarisation, angle of view, and to a lesser extent, temperature, purity and surface conditions. Because it is also reflective, its brightness is also dependent on the downwelling temperature. At 94GHz the reported brightness for water (vertical polarisation) varies between 150 and 300K.

Soil: As with water, it is dependent on polarisation and angle of view. It is also dependent on moisture content and surface roughness. At 94GHz the reported brightness for soil (vertical polarisation) varies between 160 and 280K.

Vegetation: Brightness of vegetation depends on its type and moisture content. At 94GHz it is reported to vary between 230 and 300K.

Built-Up Areas: This will be complex, however at 94GHz, asphalt is given to be 260 to 300K.

Though there is a significant overlap between the brightness temperatures in these cases, this is due to the fact that the data were taken under a variety of weather conditions. In general, there will be a significant contrast between different materials under the same weather conditions.

4.7. Example

A space based radiometer operating at 94GHz with a bandwidth of 2GHz looks directly downwards to the ground at a temperature of 27°C which has an average emissivity ε (over the footprint) of 0.9. What is the received power?

As discussed earlier, the reflectivity $\rho = 1-\varepsilon = 0.1$

From Figure 4.2, the total attenuation directly downwards through the atmosphere at 94GHz is 1dB. The loss is $L_A = 10^{\mathrm{dB}/10} = 1.26$

Assuming that the air has a water content of 3g/m^3, From Figure.4.3, the downwelling brightness temperature at 94GHz is 30K. Assume that the upwelling and the downwelling temperatures are the same.

$$T_{AP} = T_{UP}(\theta,\phi) + \frac{1}{L_A}\left[\varepsilon T_B + \rho T_{SC}\right], \tag{4.16}$$

$$T_{AP} = 30 + \frac{1}{1.26}\left[300\times 0.9 + 30\times 0.1\right],$$

$$T_{AP} = 30 + 216.7 = 246.7K\ .$$

For a bandwidth of 2GHz

$$P = 30 + 10\log_{10} kT\beta \quad \text{dBm},$$

$$P = -81.7dBm\ .$$

4.7.1. Temperature Contrast

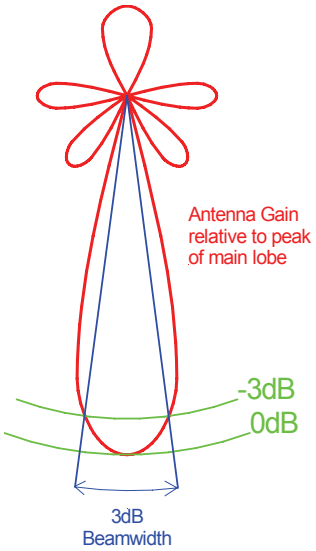
Typical temperature contrasts of metallic objects to other materials is summarised in the table below.

Table 4.2: Temperature contrast of metallic objects and other materials under different weather conditions

Material	Atmospheric Conditions		
	Clear	Fog	Rain
Vegetation	220K	200K	40K
Water	120K	100K	30K
Concrete	190K	170K	40K

4.8. Antenna Considerations

4.8.1. Beamwidth



The 3dB beamwidth of the antenna can be approximated by the following formulae

$$\theta_{3dB} \approx \frac{70\lambda}{D} \text{ degrees,} \qquad (4.17)$$

where D – Diameter of the antenna (m),
λ - Wavelength (m).

4.8.2. Efficiency

In the previous discussion, it has been assumed that the antenna is lossless, however, in reality an antenna absorbs a certain amount of the power incident on it, and hence it also radiates.

$$T_{AO} = \eta_l T_A + (1-\eta_l) T_p , \qquad (4.18)$$

where: T_{AO} – Equivalent apparent temperature at the antenna output port (K),
η_l – Radiation Efficiency of the Antenna (Typ 0.6),
T_A – Scene Temperature measured by the antenna (K),
T_P – Physical Temperature of the antenna (K).

Note that η_l is equivalent to the surface reflectivity ρ of the antenna.

4.8.3. Fill Ratio

The size of the antenna footprint does not affect the terrain's brightness temperature. However, the footprint area is important when an object with a different emissivity than that of the terrain is present in the footprint. If such an object is completely enclosed by the antenna, then the observed brightness temperature can be calculated to be

$$T_B = T_{BG}(1-F) + T_{BT} F , \qquad (4.19)$$

$$F = \frac{A_T}{A} , \qquad (4.20)$$

where F – Fill in ratio,
T_{BG} – Ground Brightness Temperature (K),
T_{BT} – Target Brightness Temperature (K),
A_T – Target Area (m^2),
A – Antenna Footprint (m^2).

4.9. Receiver Considerations

4.9.1. Mixer Implementations for Microwave Receivers

At millimetre wavelengths (>50GHz) low noise amplifiers still expensive, so many radiometers use mixers fed directly from the antenna port.

In the block diagram shown below, the mixer generates the two frequencies f_{RF}-f_{LO} and f_{RF}+f_{LO}. The f_{RF}-f_{LO} term will become the IF signal.

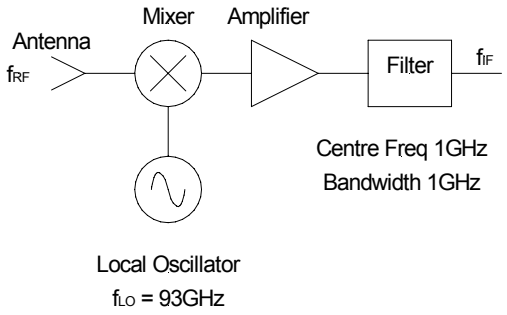


Figure 4.4: The down conversion process

Two different frequencies satisfy the requirement for $f_{IF} = f_{RF}$-f_{LO}. If $f_{RF} = f_{LO}$+f_{IF} then the output of the mixer will be f_{LO}+f_{IF}-$f_{LO} = f_{IF}$ If $f_{RF} = f_{LO}$-f_{IF} then the output of the mixer will be f_{LO}-f_{IF}-$f_{LO} = -f_{IF}$. This latter response is called the image response of the mixer and is indistinguishable from the direct response.

In the example shown, f_{RF} = 93+1 = 94GHz for the direct response and f_{RF} = 93-1 = 92GHz for the image response.

If the radiometer receiver is implemented as shown in the diagram, then it will receive signals over the band from 91.5 to 92.5GHz and from 93.5 to 94.5GHz both of which will be down converted to the IF band from 0.5 to 1.5GHz.

4.9.2. Mixer Specifications

The mixer conversion loss is defined as follows:

$$L_c = 10\log\frac{available.RF.input.power}{IF.output.power}\ \text{dB} \qquad (4.21)$$

Practical mixers usually have a conversion loss between 4 and 8dB. It generally increases with increasing frequency, it is also a function of LO drive (or pump) power. Typical microwave and millimetre wave mixers require LO powers of between 10 and 13dBm but many can be biased externally using a small DC current in which case the required LO drive is reduced to between 0 and +3dBm.

Mixer noise characteristics are important, so when specifying or reading mixer specifications a distinction must be made as to whether the input is a single sideband or a double sideband signal. It was shown above that the mixer produces an IF output from two input frequencies, and will therefore collect noise power from both frequencies. When used with a DSB input, the mixer will have desired signals at both

RF frequencies, while an SSB input provides the desired signal at only one of those frequencies. The DSB noise figure will be 3dB lower than the SSB noise figure.

Millimetre Wave Mixers

The following table lists the basic specifications for mixers made by TRG.

Table 4.3: Mixer specifications

Model Number	**960K**	**960A**	***960B***	**960U**	**960V**	**960E**	**960W**	**960F**
Frequency Range GHz	18-26.5	26.5-40	33-50	40-60	50-75	60-90	75-110	90-140
Waveguide	WR-42	WR-28	WR-22	WR-19	WR-15	WR-12	WR-10	WR-8
DSB Noise Figure dB[1]	3.5	4.0	4.0	4.5	4.5	5.0	5.0	5.5
Conversion Loss dB[2]	5.0	5.5	5.5	6.0	6.0	6.5	6.5	7.0

1. DSB noise figure assumes a +7dBm LO, IF frequency of 10-1000MHz and a 1.5dB IF amplifier noise figure.

2. Conversion loss SSB (dB) assumes a +7dBm LO. Starved or high LO drive versions are available e.g. 0dBm < LO < +16dBm

Both Farran and Millitech also offer balanced mixers with similar characteristics. At W-band (75-110GHz) these are as follows:

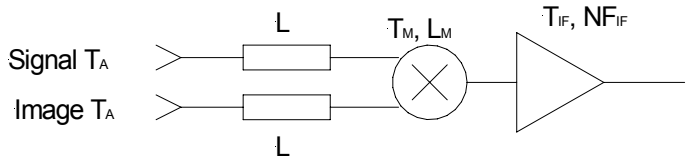
Table 4.4: Farran and Millitech mixer specifications

	Farran	**Millitech**
DSB Noise Figure dB	7.5	7.0
Conversion Loss dB	7.2	8.0

The specifications assume an IF amp with a 1.5dB noise figure and 13dB LO drive

Noise Figure

A down converter block can be represented by two separate inputs to the mixer as shown in the figure below.



It can also be shown that the Noise Figure NF for a cascaded receiver chain made up of a number of stages each with gain and individual noise figures

$$NF = NF_1 + \frac{NF_2 - 1}{G_1} + \frac{NF_3 - 1}{G_1 G_2}. \qquad (4.22)$$

For the double sided mixer implementation shown in the diagram, substitute $L = NF_1$, $L_M/2 = NF_2$, $1/L = G_1$, $2/L_M = G_2$ and $NF_{IF} = NF_3$ to obtain the total noise figure for the double sided implementation

$$NF_{DSB} = L + L\left(\frac{L_M}{2} - 1\right) + (NF_{IF} - 1)\frac{L.L_M}{2} = \frac{L.L_M NF_{IF}}{2}. \tag{4.23}$$

Similarly for single side-band (SSB) operation

$$NF_{SSB} = L.L_M NF_{IF}. \tag{4.24}$$

These two equations can be written in terms of temperature where T_o is the ambient temperature (290K),

$$T_{SYS} = T_O(NF_{DSB} - 1) \quad \text{for DSB operation,} \tag{4.25a}$$

$$T_{SYS} = T_O(NF_{SSB} - 1) \quad \text{for SSB operation.} \tag{4.25b}$$

4.10. The System Noise Temperature

Even without any external input, the radiometer will produce an output because the receiver is not at *absolute zero*. This output is also defined in terms of an equivalent noise temperature T_{sys} of a matched resistor at the antenna port. The available noise power P_N from such a resistor is expressed as

$$P_N = kT_{sys}\beta.G, \tag{4.26}$$

where k – Boltzmann's Constant 1.38x10^{-23} J/K,
β – System Bandwidth (Hz),
G – System Power Gain.

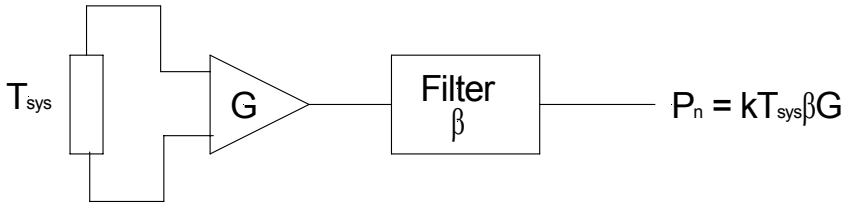


Figure 4.5: System noise temperature equivalent circuit

The receiver introduces additional noise into the system that is incorporated into the equation for T_{sys}

$$T_{sys} = T_o(NF - 1), \tag{4.27}$$

where NF is the Noise Figure for the receiver and is defined as the ratio of the input SNR to the output SNR with the input terminated at T_o = 290k.

4.11.Radiometer Temperature Sensitivity

The ability of a radiometer to detect changes in the input temperature ΔT is determined from the analysis of the detector output when the input is band limited white noise.

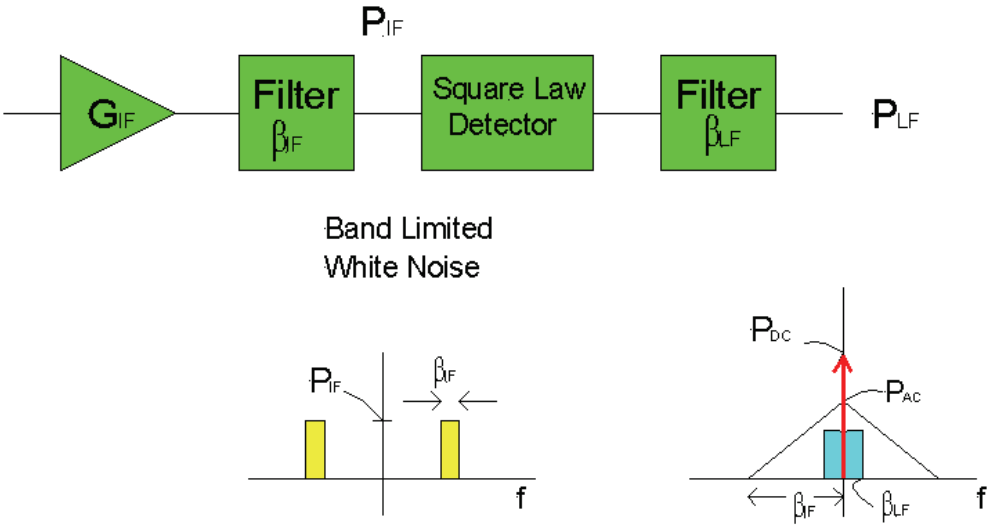


Figure 4.6: The basic radiometer circuit

4.11.1. The IF Signal

Assuming a rectangular filter, the double-sided spectrum at IF has a bandwidth β_{IF} and a height $\frac{1}{2}\, kT_{sys}G_{IF}$ as shown in the figure above.

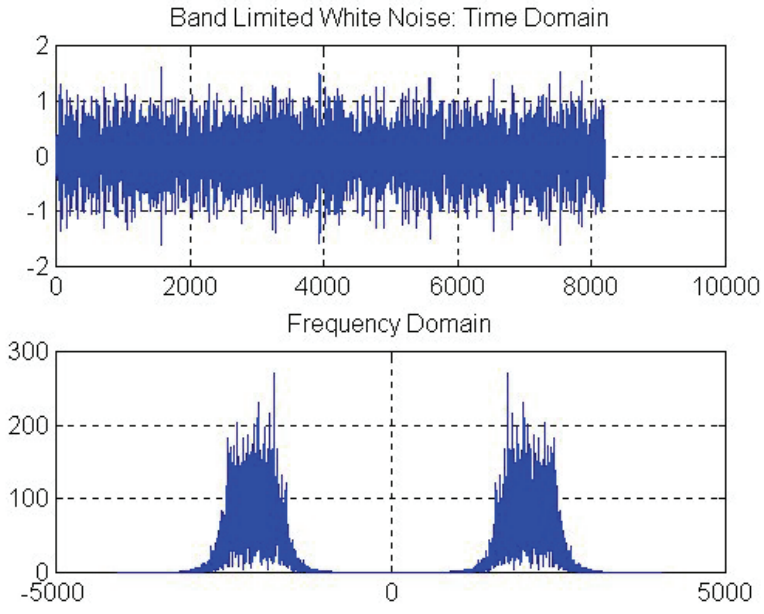


Figure 4.7: Radiometer signals in the time and frequency domain

4.11.2. The Detected Signal

A square law detector produces an output signal proportional to the square of the input envelope. It can be shown that the post detection probability density function includes a DC component P_{DC} and a double-sided triangular noise component P_{AC}.

The magnitude of the DC power component is given by: $P_{DC} = \left(\frac{kTG}{2}\right)^2 \beta_{IF}{}^2$

And the AC power density has height: $P_{AC} = \left(\frac{kTG}{2}\right)^2 \beta_{IF}$ and width β_{IF}.

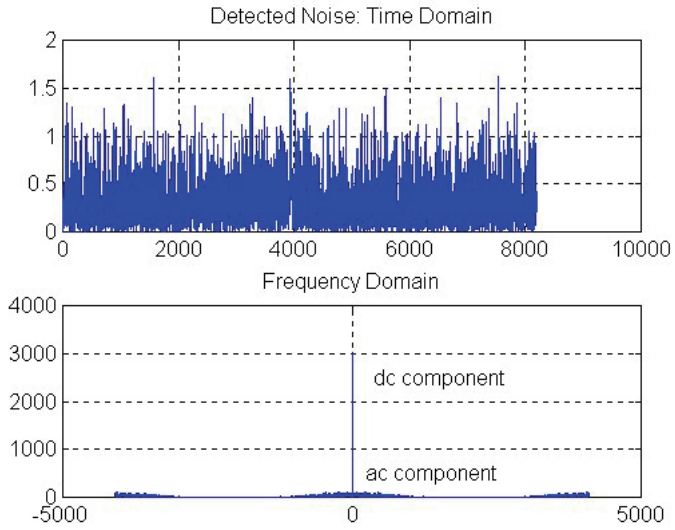


Figure 4.8: Radiometer signals after detection

4.11.3. Lowpass Filtered Signal

The signal is then passed through a low pass filter with a bandwidth β_{LF} which does not alter P_{DC} but reduces the AC component to an almost rectangular density function (because $\beta_{LF} << \beta_{IF}$) of height $\left(\frac{kTG}{2}\right)^2 \beta_{IF}$ and width β_{LF}.

The AC power output for a double sided spectrum after the low frequency filter is the area,

$$P_{AC} = 2\left(\frac{kTG}{2}\right)^2 \beta_{IF} \beta_{LF} . \qquad (4.28)$$

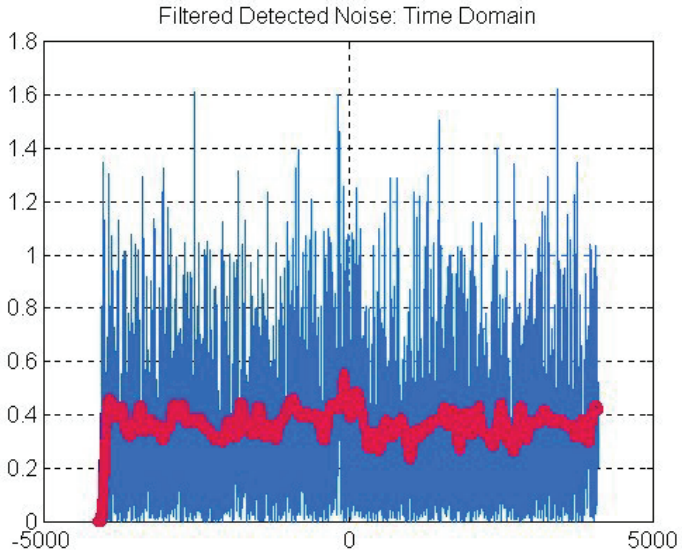


Figure 4.9: Radiometer signal extracting the DC component

The ratio of AC power component to the DC power component is

$$\frac{P_{AC}}{P_{DC}} = \frac{2\left(\frac{kTG}{2}\right)^2 \beta_{IF}\beta_{LF}}{\left(\frac{kTG}{2}\right)^2 \beta_{IF}{}^2} = \frac{2\beta_{LF}}{\beta_{IF}}. \tag{4.29}$$

In terms of voltages this can be rewritten as

$$\frac{V_{AC}}{V_{DC}} = \sqrt{\frac{2\beta_{LF}}{\beta_{IF}}}. \tag{4.30}$$

Since the temperature change ΔT can be measured by V_{AC} while the sum of the antenna and system temperatures T_A+T_{sys} determines V_{DC} then the two ratios will be the same and we can write

$$\frac{\Delta T}{T_A + T_{sys}} = \sqrt{\frac{2\beta_{LF}}{\beta_{IF}}}, \tag{4.31}$$

and

$$\Delta T = (T_A + T_{sys})\sqrt{\frac{2\beta_{LF}}{\beta_{IF}}}. \tag{4.32}$$

If the lowpass filter is implemented as an ideal integrator with a time constant, τ, then $\beta_{LF} = 1/2\tau$ and the temperature change ΔT can be rewritten as

$$\Delta T = \frac{T_A + T_{sys}}{\sqrt{\beta_{IF}\tau}}. \tag{4.33}$$

Note that β_{IF} is not the 3dB bandwidth but the reception bandwidth. For a 2 pole RC filter $\beta_{IF} = 1.96\beta_{3dB}$.

4.11.4. Detection Probability

These formulae determine the minimum detectable signal, where the signal level is equal to the noise level. However, in reality the SNR required will be much higher, typically 13dB, and so the acceptable temperature difference will have to be scaled appropriately. This is discussed in Chapter 10.

4.12. Radiometer Implementation

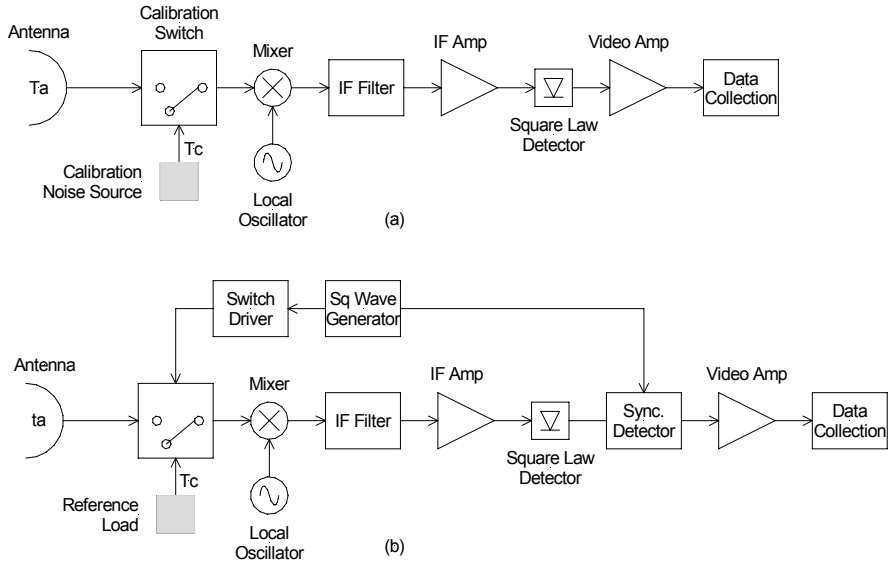


Figure 4.10: Block diagram of radiometer types (a) total power (b) Dicke

4.12.1. Total Power Radiometer

A square law detector cannot distinguish between an increase in the signal power (an increase in T_A) from an increase in the pre-detection gain G. If the gain varies by ΔG around the average gain G, then the minimum detectable temperature change ΔT_{min} is determined as

$$\Delta T_{\min} = (T_A + T_{sys})\left[\frac{1}{\beta_{IF}\tau} + \left(\frac{\Delta G}{G}\right)^2\right]^{1/2}. \tag{4.34}$$

4.12.2. Dicke Radiometer

In the block diagram for the Dicke Radiometer shown above it can be seen that the receiver input is switched at a constant rate between the antenna port and a reference load maintained at a constant temperature.

The output of the square law detector is then synchronously detected as shown in the figure below, such that the final output is proportional to the difference between the temperature of the antenna and the reference load.

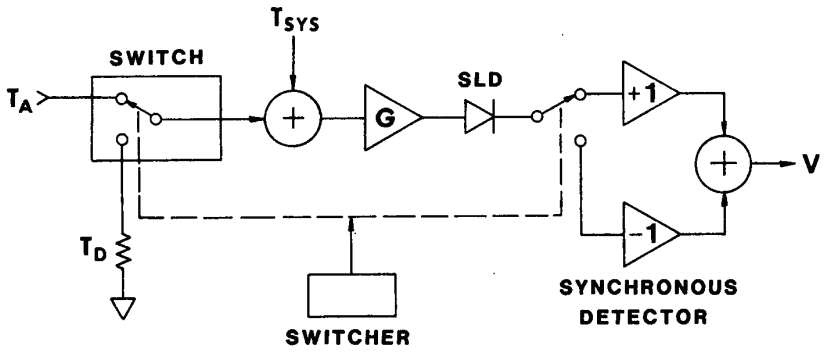


Figure 4.11: Dicke radiometer simplified schematic diagram

The derivation of this formula is beyond the scope of this course, but it should be noted that it is based on the premise that only one half of the switching time, τ, is used to view the antenna, while the other half is used to view the Dicke reference.

$$\Delta T_{\min} = (T_A + 2T_{sys} + T_D)\left[\frac{1}{\beta_{IF}\tau} + \left(\frac{\Delta G}{G}.\frac{T_A - T_D}{T_A + 2T_{sys} + T_D}\right)^2\right]^{1/2}. \qquad (4.35)$$

4.12.3. Comparison Between Radiometer Types

If the performance of these two radiometer implementations is compared for the following realistic scenario: T_A-T_D = 10K, T_D = 300K, T_{sys} = 1000K, β_{IF} = 1GHz and τ = 0.1s then the following results are obtained

Total Power $\Delta T_{min} = 0.185\text{K for } \frac{\Delta G}{G} = 0.01\%,$

$\Delta T_{min} = 1.87\text{K for } \frac{\Delta G}{G} = 0.1\%.$

Dicke $\Delta T_{min} = 0.241\text{K for } 0.01\% < \frac{\Delta G}{G} < 0.1\%.$

Hence if $\frac{\Delta G}{G} > 0.0172\%$ then the Dicke configuration is superior.

4.13.IF and Video Gains

The required IF gain is determined from the signal level required by the square law detector and the video gain is then determined from the output of the square law detector and the display or signal logger requirements.

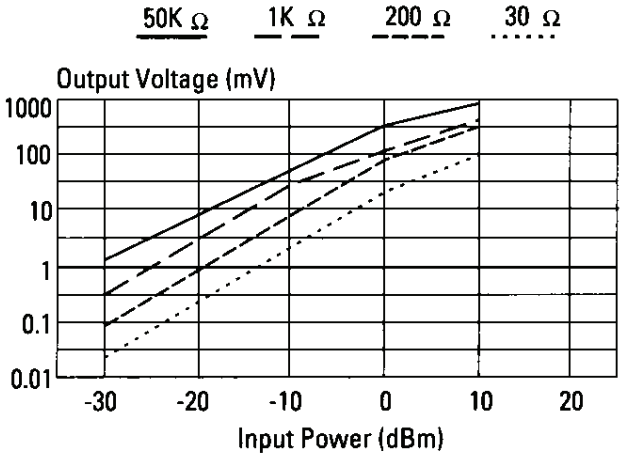


Figure 4.12: Typical microwave square law detector characteristics

The power at the antenna of a typical uncooled radiometer is -75dBm, so to make it compatible with the square law detector, a gain of 65dB is required.

In that case, the detector output would be 10mV, so for an operating voltage of 1V, additional video voltage amplification of ×100 (40dB) is required.

4.14.Target Seeker Design Example

Millimetre wave anti tank missiles, mortar shells and other sub-munitions often resort to radiometric tracking or detection for the final phases of the engagement.

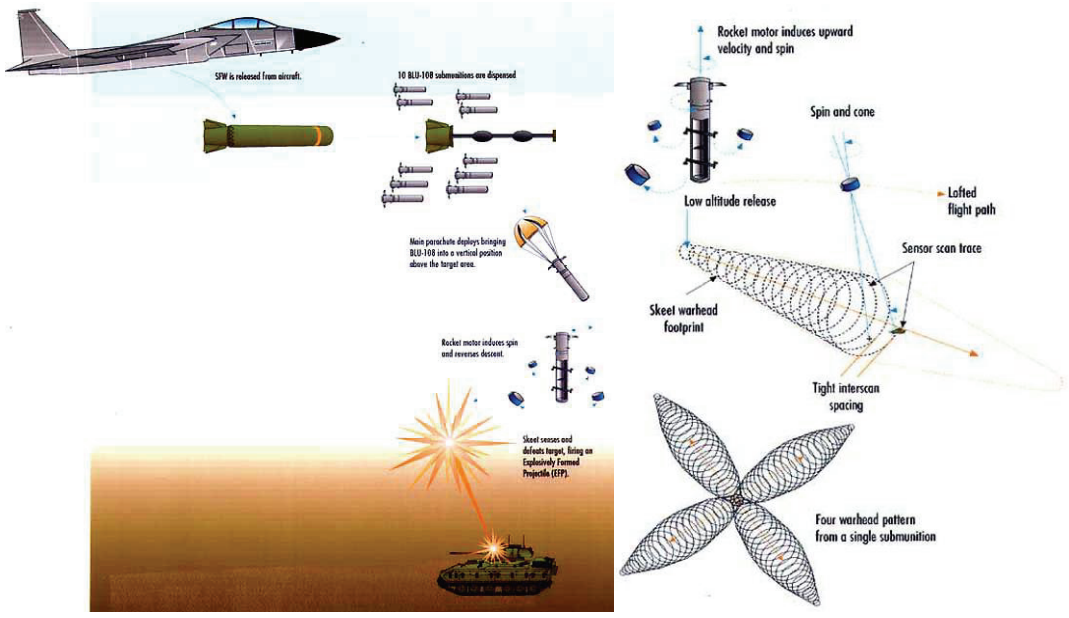


Figure 4.13: Textron submunition releases anti tank skeet

In this example assume that a skeet is launched from a height of 25m with an upward velocity of 50m/s, a horizontal velocity of 10m/s, that the cone angle remains constant at 10° and that it spins at a constant rate of 2rps.

It is fitted with a radiometric seeker with a 50mm aperture, that operates at 95GHz with a receiver bandwidth of 2GHz.

Start by writing a MATLAB procedure that generates the position of the skeet over the 10s after release and the beam pattern that is generated on the ground

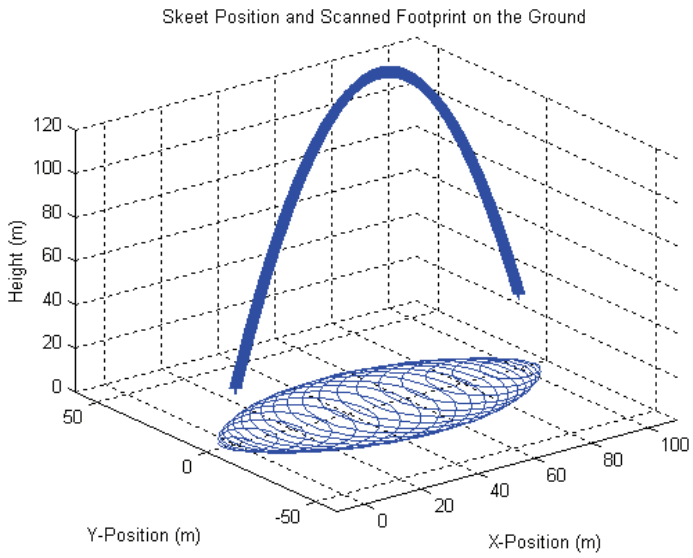


Figure 4.14: Skeet position and search footprint on the ground

Determine the radiometric temperature difference between the target and the surrounding ground as follows:

The average sky temperature at 95GHz is assumed to be 60K over a 140° angular sweep, 150K over a 20° sweep and 300K up to 10° above the surface around the target as shown in the following figure.

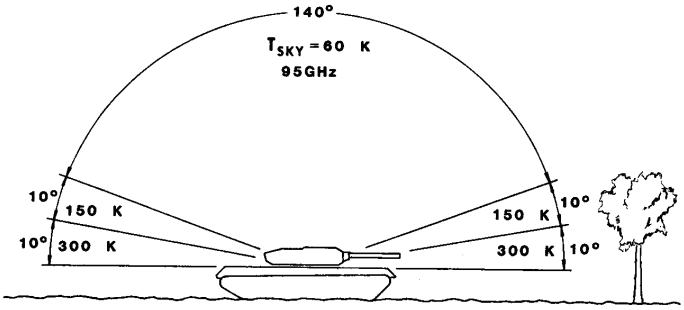


Figure 4.15: Passive target detection temperature scenario

Assuming that the tank has been driving and its temperature is 35°C (308K) and its emissivity $\varepsilon = 0.1$. The average target temperature over the full hemisphere will be the sum of reflected temperatures, scaled by their various areas and the tank reflectivity, and the emitted temperature scaled by the emissivity.

The area of each of the 10° sections can be found by integration

$$A_1 = \int_0^{\alpha} 2\pi r^2 \cos\theta d\theta \,, \tag{4.36}$$

$A_1 = A_2 \approx 1.09$ steradians.

Therefore the area of the remaining 140° section is

$A_3 = 2\pi - (A_1 + A_2) = 4.10$ steradians.

The reflected (scattered) temperature is

$$T_{SC} = \rho\left[\frac{A_1}{2\pi}T_1 + \frac{A_2}{2\pi}T_2 + \frac{A_3}{2\pi}T_3\right] = 0.9\left[\frac{1.09}{2\pi}\ 300 + \frac{1.09}{2\pi}150 + \frac{4.10}{2\pi}60\right] = 105.5K$$

The radiated (brightness) temperature of the target is

$T_{BT} = \varepsilon T_T = 0.1 \times 308 = 30.8K$.

The apparent temperature is the sum of the reflected and radiated brightness temperatures modified by the loss through a portion of the atmosphere, plus the upwelling temperature. For a path length of 100m and a clear air attenuation of 0.2dB/km, the loss is very small and $L_A \approx 1$. The upwelling temperature, which is related to the attenuation is also very small ($T_{up} \approx$ 1K), so can also be ignored

$$T_{AP} = T_{UP}(\theta,\phi) + \frac{1}{L_A}\left[T_{BT} + T_{SC}\right] \approx 30.8 + 105.5 = 136.3K \,.$$

Assuming that the surrounding terrain is at a temperature of 20°C (293K) and the emissivity of the ground 0.92 (typical for grass and soil), then the apparent temperature of the ground calculated in the same way as it is for the target

$$T_{APG} = \varepsilon_G T_G + \rho T_{GSC} = 0.92 \times 293 + 0.08 \times 117 = 279K \,.$$

Note that the scattered contribution is very small as the reflectivity of the ground is low.

Note also that the tank appears to be much colder than the surrounding terrain because it scatters the cold sky temperature.

The actual brightness temperature seen by the radiometer is determined by the temperature difference, and the percentage beamfill.

An antenna looking straight down will illuminate a circular footprint on the ground. The diameter of the footprint will be a function of the antenna beamwidth (generally to the half-power or 3dB contour), and the distance to the ground.

A reasonable approximation for the beamwidth

$$\theta_{3dB} = \frac{1.22\lambda}{D} \text{ radians,}$$

where λ - Wavelength (m), and D – Antenna diameter (m).

The footprint area on the ground is

$$A_B = \frac{\pi}{4} d^2 = \frac{\pi}{4} (R\theta_{3dB})^2,$$

where R – Range to the ground (m).

For D = 50mm and λ = 3.16mm $A_B = 4.67 \times 10^{-3} R^2$.

The cross sectional area of a tank as seen from above A_T = 20m^2 so the scene temperature measured by the antenna will just be the sum of the background and tank brightness temperatures scaled by their relative areas

$$T_A = T_{APG} \frac{A_B - A_T}{A_B} + T_{APT} \frac{A_T}{A_B}.$$

A MATLAB procedure is then used to plot this scene temperature as the range to the ground varies.

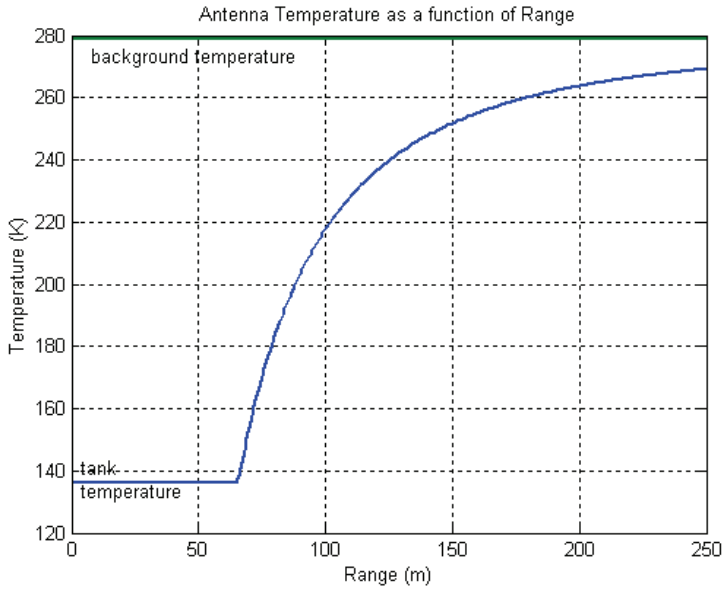


Figure 4.16: Temperature variation due to beamfill effects

As the antenna scans across the terrain, it will measure the background temperature of 279K, however, when it encounters the tank, the measured temperature will dip to the apparent temperature shown in the figure.

The actual temperature difference will depend on the range to the target.

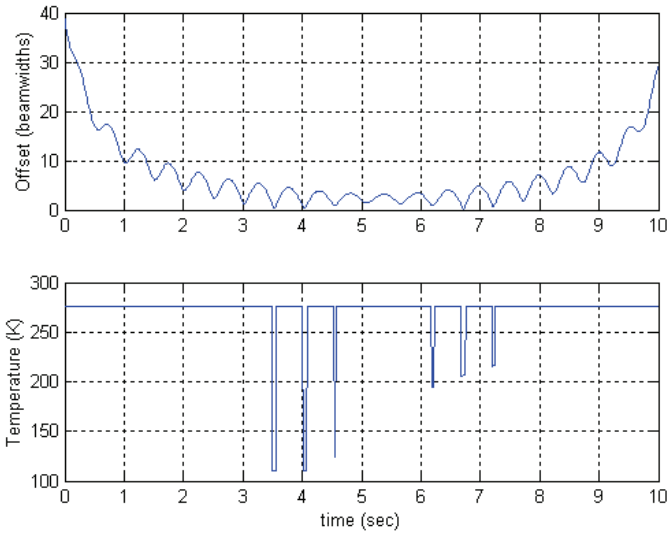


Figure 4.17: Antenna temperature as a function of time showing instances when the beam scans through the cold target

The figure shows that the antenna will sweep across the target a number of times during its flight (assuming it does not detonate its warhead), and that each time this occurs, the measured antenna temperature will dip

Radiometer Implementation

To keep the cost of the skeet as low as possible, a total power radiometer is used with an uncooled front end.

The allowable integration time is made equal to the dwell time of the antenna on a target.

The circumference of the circle scanned by the beam with a cone angle of 10° (0.17rad) is

$$circ = 2\pi(R\theta_{cone}).$$

The size of the antenna footprint on the ground is

$$D_{foot} = R\frac{1.22\lambda}{D}.$$

As the skeet scans at 2rps, the dwell time in seconds

$$\tau = \frac{D_{foot}}{circ}\frac{1}{2} = \frac{1.22\lambda}{4\pi D\theta_{cone}} = \frac{1.22\times 3.21}{4\pi\times 50\times 0.17} = 37ms.$$

Receiver Noise Temperature

It is assumed that the system operates using a single sideband and that the mixer is placed with the feed horn at the focal point of the antenna so that the waveguide losses are minimised.

L = 0.2dB = 1.05 (feed loss from the antenna to the mixer)
L_m = 6dB = 3.98 (mixer conversion loss)
NF_{IF} = 1.5dB = 1.41 (low noise amplifier noise figure)

$$NF_{SSB} = L.L_M NF_{IF} = 1.05 \times 3.98 \times 1.41 = 5.88$$

$$T_{SYS} = T_O(NF_{SSB} - 1) = 290(5.88 - 1) = 1415K$$

Minimum Detectable Temperature Difference

The formula to determine the minimum temperature difference is given by:

$$\Delta T_{\min} = (T_A + T_{sys})\left[\frac{1}{\beta_{IF}\tau} + \left(\frac{\Delta G}{G}\right)^2\right]^{1/2}.$$

Assume that the system gain is completely stable so $\Delta G = 0$, the equation reduces to

$$\Delta T_{\min} = \frac{T_A + T_{sys}}{\sqrt{\beta_{IF}\tau}} = \frac{275 + 1415}{\sqrt{2 \times 10^9 \times 37 \times 10^{-3}}} = 0.2K,$$

where T_A – Background Temperature (279K),
T_{sys} – Receiver System Temperature (1415K),
β_{IF} – Receiver Bandwidth (2GHz),
τ - Integration time (37ms).

Hence, a 0.2K temperature drop should just be detectable. However, for a good probability of detection, a signal to noise ratio of at least 13dB is required, and so a temperature drop of at least $0.2 \times 10^{13/10}$ = 4K is required.

Going back to the formula for the beamfill effects, a temperature difference of 4K will occur at a range of 420m, which exceeds the height reached by the skeet, so the tank will always be detectable.

The received power from the scene temperature will be

$$P = 30 + 10\log_{10} kT\beta = 30 + 10\log_{10}(1.38 \times 10^{-23} \times 279 \times 2 \times 10^9) = -81dBm$$

The actual output power will be higher than that as it includes the noise generated within the receiver as well

$$P = 30 + 10\log_{10} k(T_A + T_{sys})\beta = -73dBm.$$

From the graph for the square law detector, we need a signal level of -10dBm, and so an IF gain G_{IF} = -10 + 73 = 63dB.

If we want to amplify the signal out of the detector so that the DC level is 1V, then a voltage gain of 100 is required.

4.15.Airborne Push-Broom Scanner

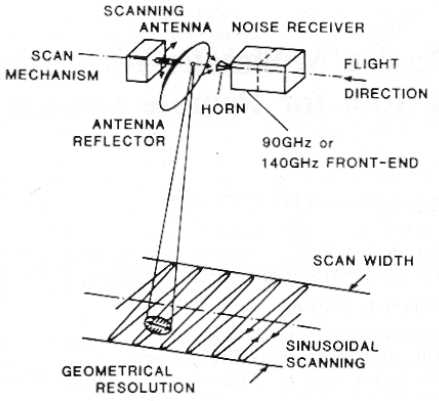


Figure 4.18: Configuration of a typical airborne push broom scanner

4.15.1. Image Processing

Histogram Modification

It is usual to enhance the temperature range of interest by expanding a small percentage of the output voltage to correspond to the full range of colours or shades.

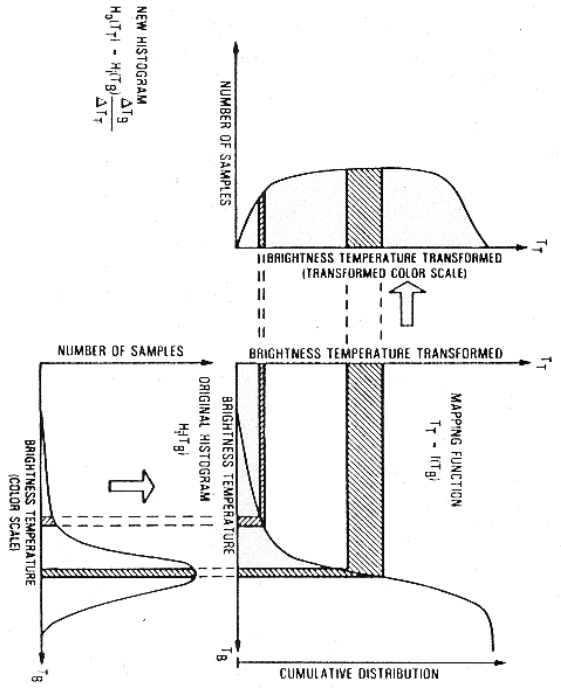


Figure 4.19: Transforming the brightness temperature histogram to enhance the region of interest

An example of this processing technique is shown in the transform from (b) to (c) in the figure below. This technique suppresses small targets, as their weight is proportional to the number of pixels at that temperature.

In millimetre wave images, boundaries appear as temperature discontinuities. To aid human operator interpretation of such images, edge enhancement techniques are used as shown in (d) below.



Figure 4.20: Visible (a) and radiometric (b) images of an airport. The contrast enhanced radiometric image is shown (c) and the edge enhanced image (d)

4.16.Radiometric Applications

4.16.1. Airborne Scanned Millimetre Wave Radiometer

German Aerospace Research Establishment (DLR)

- Function Experimental
- Height >80m
- Aircraft speed 50m/s
- Scan width +/-14.5°
- Ground Resolution (nadir) <1.5x1.5m
- Frequency 140GHz
- Sensitivity <0.4K
- Type Total power with cooled front-end

Note from the image that flat metal roofs of hangars appear white, concrete areas are yellow and aircraft are violet. Grass covered areas have the highest brightness temperatures and appear green.

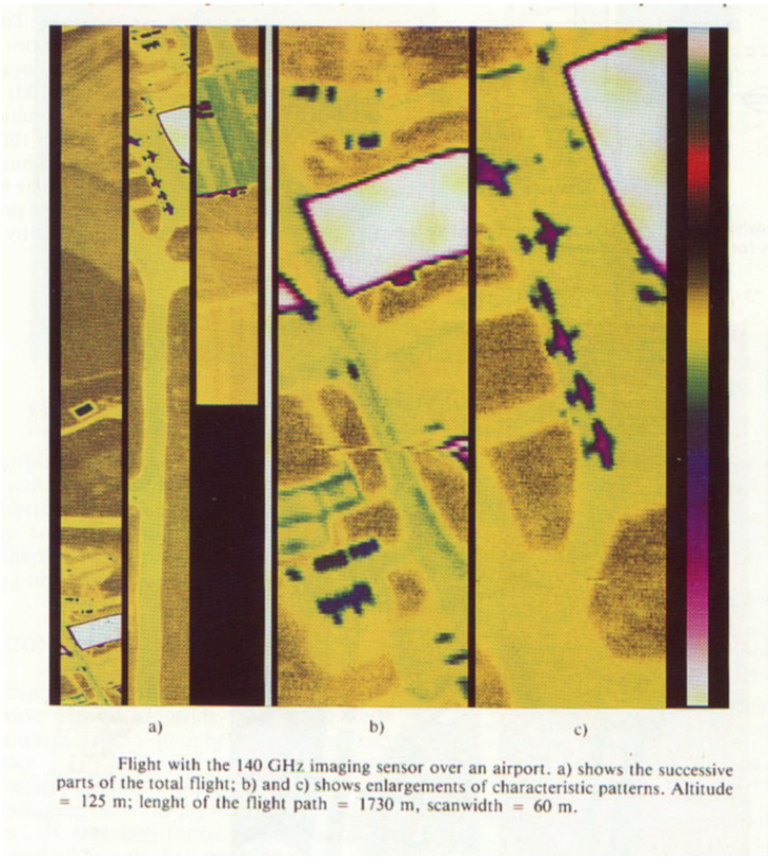


Figure 4.21: Radiometric images of an airport showing progressively enlarged sections of the image

Forward-looking radiometers can be used to produce ground images through mist and fog or rain. Such systems would be useful for both low visibility landing and airport ground movement control.

4.16.2. Scanning Multichannel Microwave Radiometer (SMMR)

- Manufacturer Jet Propulsion Laboratory
- Satellite NIMBUS 7
- Operational October 1978 to August 1987
- Function Sea surface temp, wind stress & sea ice cover
- Slant Range 1380km
- Height 960km
- Beam Nadir 42°
- Beam Incidence 50.3°
- Satellite speed 6.5km/s
- Scan width+/-25° (780km)
- Ground resn 17×25km at 0.81cm to 100××150km at 4.54cm
- Band 0.81, 1.36, 1.66, 2.8, 4.54cm
- Sensitivity Not available
- Type 6×Dicke Radiometer
- Calibration Ambient RF termination and a deep space horn
- Polarisation Alternating for 4 low frequency channels
  Dual for 0.81 and 1.36cm channels

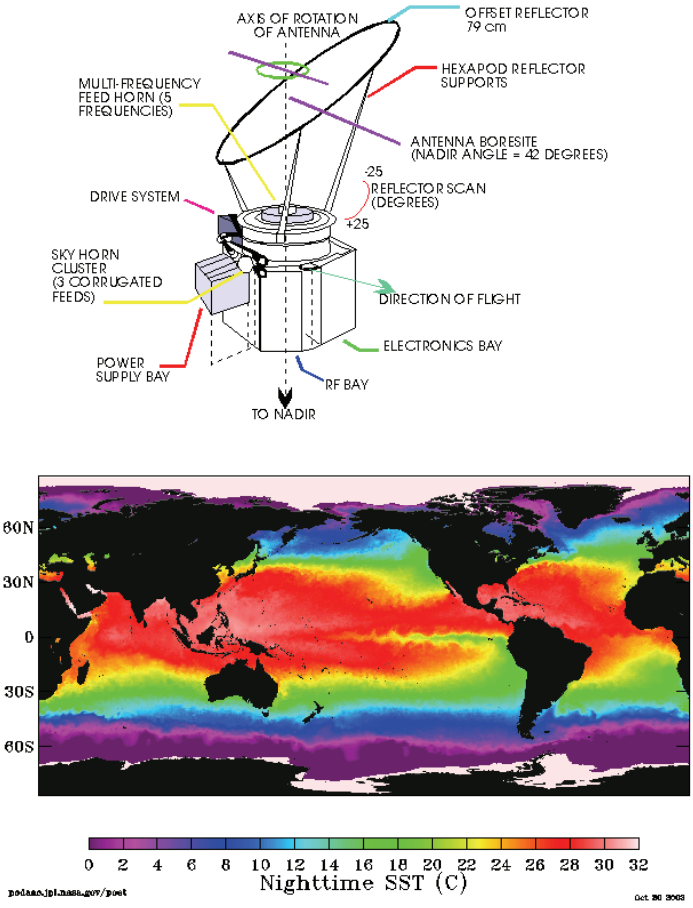


Figure 4.22: Space borne radiometer and sea temperature measured using the instrument

4.16.3. Ground Based Millimetre Wave Radiometers

Low Visibility Imaging

The passive millimetre wave images illustrate quite clearly, how the low attenuation at 94GHz even through mist and fog can be exploited to produce images in bad weather. The two images produced by DERA show a visible image above a passive radiometric image of the same scene.

The top image shows a view of the Severn valley taken from the Malvern Hills. The visible image is hazy with visibility limited to a few kilometres, while the lower radiometric image shows fields and hedgerows at a much greater range.

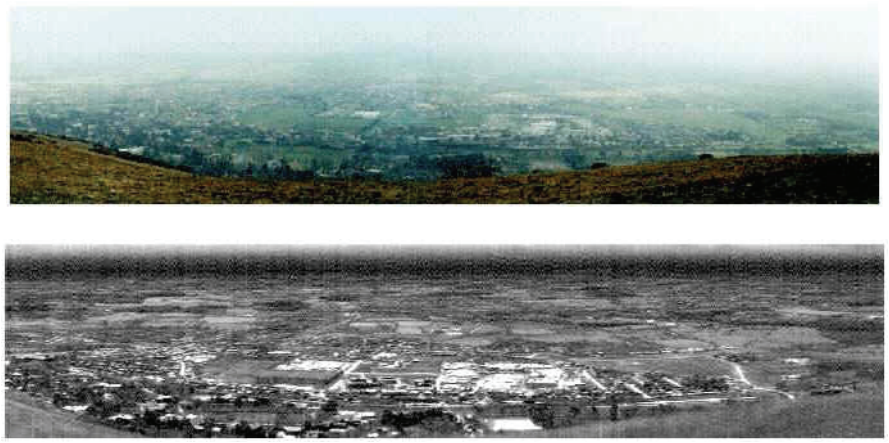


Figure 4.23: Visible and radiometric images of the Severn valley made at 94GHz

This image pair shows a view of the Malvern Hills made at a range of 2 to 2.5km made with a 35GHz radiometer. The mist in the visible image completely masks the hills. However, at 35GHz they are clearly visible.



Figure 4.24: Visible and radiometric images of the Malvern hills made at 35GHz

High speed image enhancement and super resolution techniques have been developed in the UK, the USA and Russia which are capable of producing "photographic" quality images such as the one shown below



Figure 4.25: High resolution radiometric image processed using super resolution techniques

Concealed Weapon Detection

Since the World Trade Centre attack, many institutions including Millitech, Farran, QinetiQ, ThruVision and DERA have been developing high-speed scanned radiometers that can be installed in the entrances to airports, stations, banks, sports arenas and other areas where security is important.

Radiometric images such as those shown below can see weapons concealed beneath clothing. In the DERA image, the man is carrying a replica Beretta 92F pistol in his right pocket.

One of the main advantages of this technology is that it is able to produce images of non-metallic low-density materials, and because it is totally passive (unlike X-ray techniques), it is not harmful.

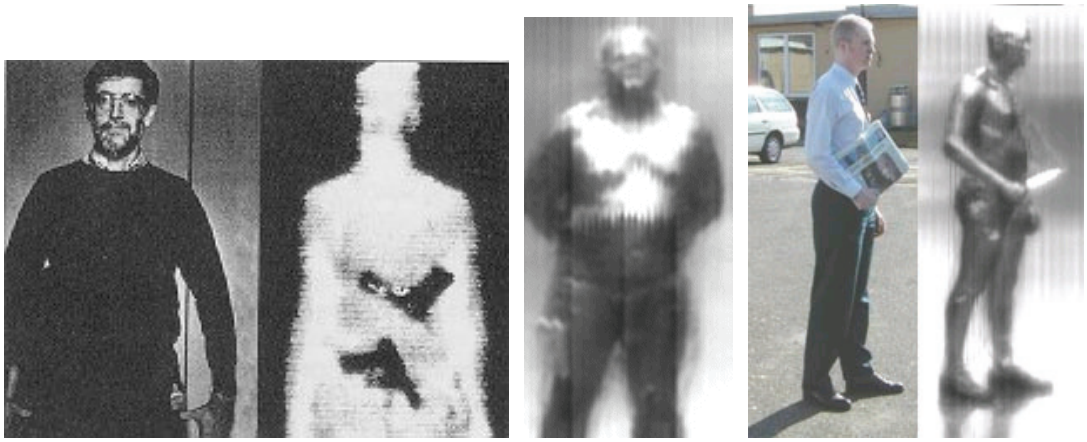


Figure 4.26: Advances in radiometric images of human beings from the first Millivision results on the left to recent images made by DERA

Since the Terrorist attack on the Twin Towers and the subsequent clampdown on airline security, a great deal of interest has been shown in the development of this technology. Of particular interest is the speed at which scans can be made as that determines the number of people per hour than can be processed.

Surveillance and Law Enforcement

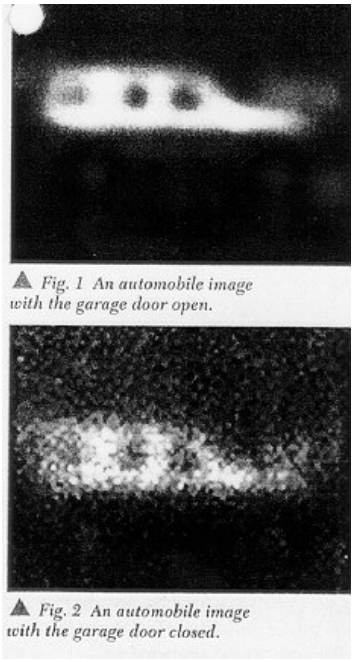


Figure 4.27: Radiometric images of SUV showing the penetration capabilities of mm wave radiation

Because millimetre wave radiation can penetrate thin layers of wood, plasterboard (dry wall) and plastics. Objects hidden behind these can often be viewed.

To illustrate this capability, the images here show a Ute with the garage door open, and with the garage door closed.

In this case, the door was constructed of two layers of plywood with a foam core.

In conjunction with active Doppler technology, such imaging capabilities are extremely useful to the military in urban warfare situations.

Medical Imaging

Because millimetre wave radiation can penetrate the top millimetre or so, medical radiometers are useful in identifying skin cancers and the like. The following prototype radiometer has been built by St Andrews University for research

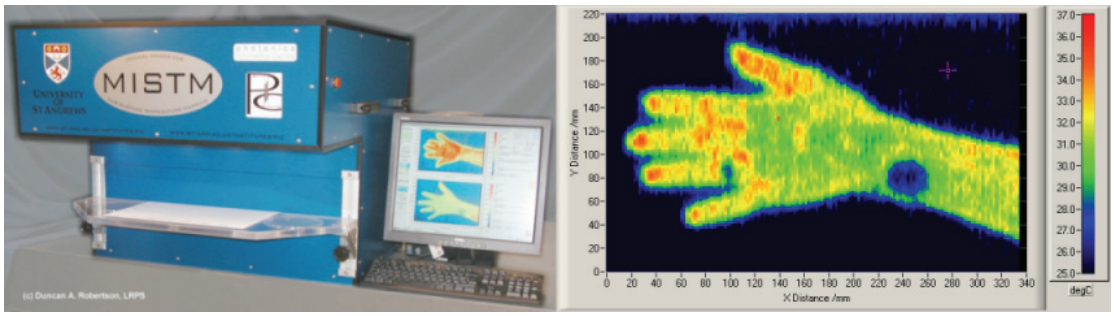


Figure 4.28: Medical imager for sub-surface temperature monitoring

Another advantage of this form of imaging, is because the signal also penetrates clothing, information may be obtained while the patient is still dressed.

Radiometric Cameras

To date, most of the radiometric imaging techniques rely on 2D scanning of large antennas operating either with a single receiver element, or at best a small array. However, the development of full electronic scanning is a priority with most of the producers of passive imaging systems.

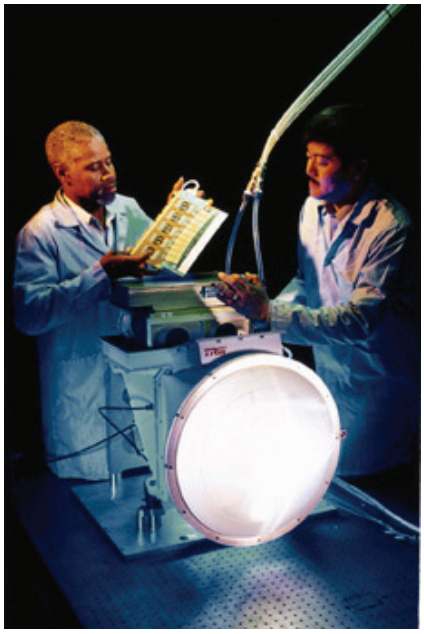


Figure 4.29: Langley Research millimetre wave radiometric camera

4.16.4. Radio Astronomy

The radiometer is a basic tool of radio astronomy and radiometers have been used to detect many species of molecules in interstellar clouds. The absorption and emission of molecular lines is primarily governed by their rotational motions, and the resonance lines are more abundant and intense in the millimetre wave region than the centimetre region.

Minimum detectable antenna temperatures of the order of tenths of a degree or less and the inherently weak signals from resonances that might be light-years away coupled with earth atmospheric noise require systems with extremely low noise temperature and high sensitivity.

Single Dish Telescopes

In the 1980's only single dish antennas were available for use at millimetre wavelengths. They were the 13.7m dish at the University of Massachusetts useable to 300GHz, an 11m dish at Kitt Peak useable to 140GHz and a 20m dish at Onsala in Sweden useable to 150GHz. Most of the antennas built for radio astronomy work are not suitable for millimetre wave work because the surface of the dish is not sufficiently smooth at such high frequencies. The James Clerk Maxwell Telescope (JCMT) at Mauna Kea operates in three bands, 210-280GHz, 300-380GHz and 460-520GHz and will shortly be operational at a fourth band 800-900GHz.

In Australia an agreement between the University of NSW and the CSIRO has seen the Mopra radio telescope upgraded to have a solid surface over its full 22m diameter. It is now the largest millimetre wave telescope in the southern hemisphere. During the upgrade, the rms error of the surface was reduced to less than 0.2mm, and if another "holographic" run is undertaken, this could be reduced to 0.15mm (20× less than the typical observing wavelength of 3mm).

The primary task of such telescopes is to survey large regions of the sky looking for objects suitable for scrutiny by the large millimetre wave arrays.

Telescope Arrays



In the last few years, arrays of dishes such as the Berkeley Illinois Maryland Association (BIMA) array at Hat Creek, shown below and the Atacama Large Millimetre Array (ALMA) in Chile have been under development to improve the angular resolution

Figure 4.30: BIMA telescope array

The BIMA array is a 10 antenna aperture synthesis which operates at wavelengths of 3mm (70-116GHz) and 1mm (210-270GHz). Each of the telescopes is 6.1m in diameter with a measured surface accuracy better than 30μm rms. The half power beamwidth is 100" (0.5mrad). The IF bandwidth is 830MHz wide and the noise temperature is 40K at 3mm and 80K at 1mm.

The antennas may be located at various stations along a T shaped track to obtain separations between 7m and 2km. Normally the antennas are deployed in one of 4 stations providing angular resolutions of 0.4", 2", 6" or 14" at 100GHz.

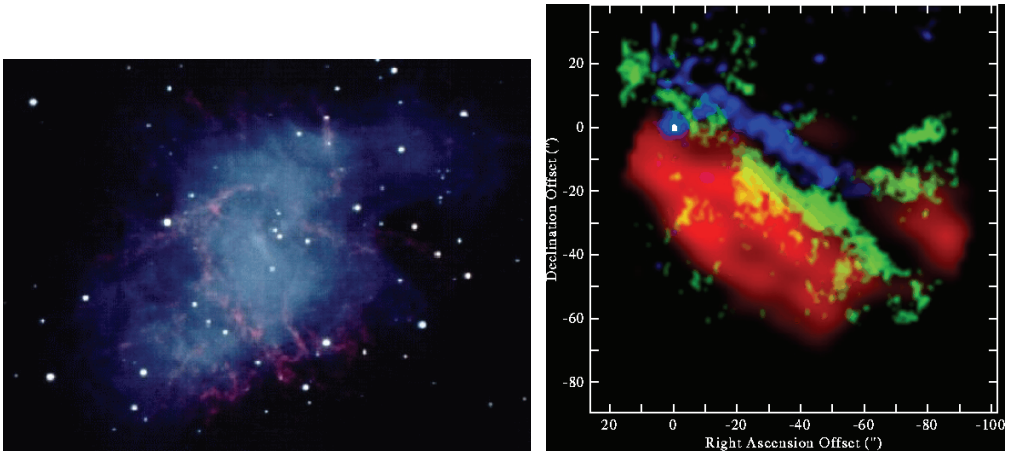


Figure 4.31: Optical and millimetre wave images of regions in Orion

Identification of Molecules in Interstellar Clouds

For linear molecules, the spectra are characterised by a series of almost harmonically related frequencies given by

$$f_s = \frac{nh}{4\pi^2 I}, \tag{4.37}$$

where n is an integer, h is Planck's constant and I is the molecular moment of inertia.

This means that lighter molecules only have spectra in the millimetre or sub-millimetre wave band.

Table 4.5 Some molecules detected in interstellar clouds

Molecule	Frequency (GHz)	Molecule	Frequency (GHz)
SiO	130.246	OCS	108.463
CN	113.492	HNCO	87.925
$C_{12}O_{16}$	115.271	CH_3OH	85.521
$C_{13}O_{16}$	110.201	CH_3CN	110.331-110.383
$C_{12}O_{18}$	109.782	CH_3C_2H	85.457
CS	146.969	X-ogen	89.190
$HC_{12}N_{14}$	88.630-88.634	HNC	90.665

The BIMA array has been able to map the abundance of the different molecules in specific regions around stars and in the remnants of supernovas.

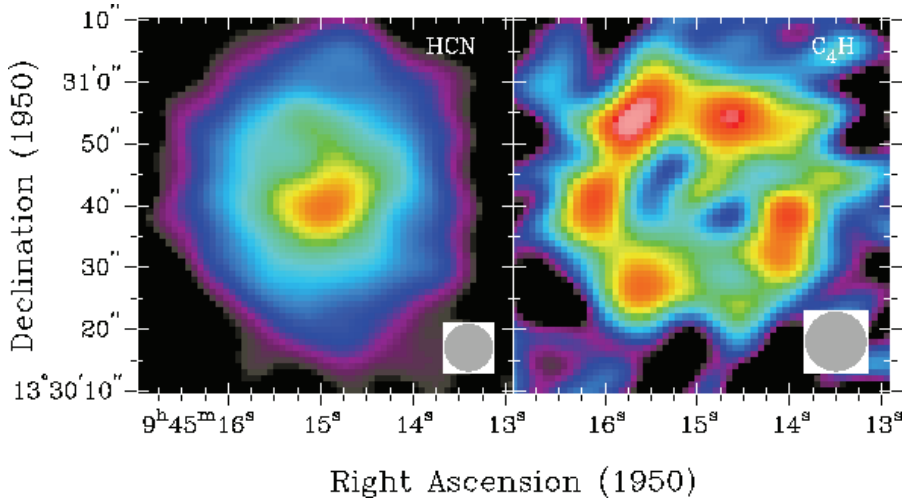


Figure 4.32: Organic molecule distribution in the area around the star IRC+10216 mapped by the BIMA array

Because of the abundance of quite complex organic molecules in space, there is speculation that life evolved there and not on earth as was once thought.

Other Astronomical Applications

Millimetre wave radio astronomy has been used to measure the brightness temperature of the sun, moon and the other planets.

Planets appear brighter at millimetre wavelengths than at lower wavelengths and hence provide information about the surfaces and atmospheres of the planets.

The part of the sun viewed by a radiometer depends on the wavelength since the absorption of electromagnetic energy by the solar constituents is frequency dependant. As the depth of penetration also depends on the frequency, millimetre wave observations provide information not easily obtained at other frequencies.

4.17.References

[1] P.Bhartia, I.Bahl, *Millimeter Wave Engineering and Applications,* John Wiley & Sons, 1984

[2] H.Suss, K.Gruner and W.Wilson, Passive Millimeter-Wave Imaging: A tool for remote Sensing, *Alta Friquenza* No. 5-6, 1989.

[3] http://www.dera.gov.uk, 17/02/2001.

[4] http://bima.astro.umd.edu/bima/home.html, 25/02/2001.

[5] http://eleceng.uks.ac.uk/research/comms/mmw_astronomy/mmw_astronomy.html, 25/02/2000.

[6] http://www.phys.unsw.edu.au/SCHOOL_INFORMATION/MEDIA_ROOM/mopranews .html, 25/02/2000

[7] http://www.astro.uiuc.edu/projects/lai/laipage.html

[8] F.Ulaby, R.Moore, A.Fung, *Microwave Remote Sensing: Active and Passive",* vol 1, Artech House, 1987

[9] M. Inggs et. al., *Radiometry Report,* AMS Document 88-S153/KED, November 1988.

[10] Langley Research Center, www.sti.nasa.gov/tto/spinoff1998/ard7.htm

[11] D Robertson, *Compact mm-wave Medical Imager,* SPIE Defence and Security Symposium, 2004

[12] Smith R.M and Sundstrom B.M, "Technical Feature - The Passive MM-Wave Scenario", Microwave Journal, March 1996, pp 22-34

Chapter 5.

Active Ranging Sensors

5.1. Overview

The sensors operational principles are the same for electromagnetic (radar, laser etc.) and active acoustic sensing.

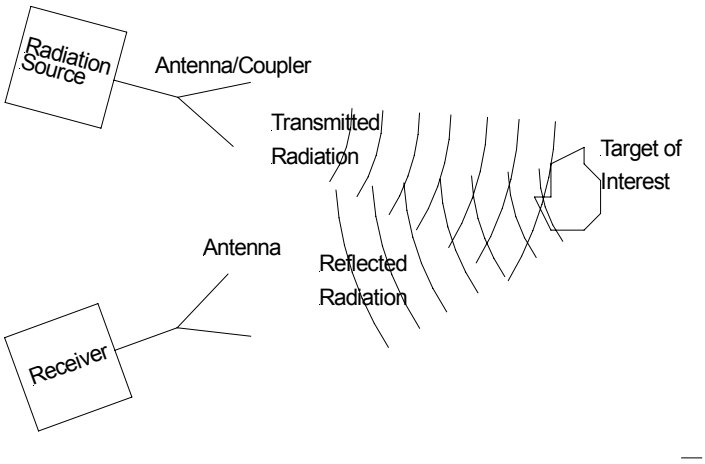


Figure 5.1: Operational principles of a generic time-of-flight sensor

Source of radiation is fed to a transmit antenna

- Tuned to the characteristics of the target (sometimes)
- Matched to the impedance of the medium to maximise coupling and efficiency
- Radiated by a directional antenna to increase the energy on target (sometimes)

Impacts on the target of interest

- Change in impedance results in re-radiation or scattering
- Re-radiation isotropic, random or directional

A small % of the power enters the receiver antenna

- Converted to an electrical signal
- Amplified and detected

5.2. Principle of Operation

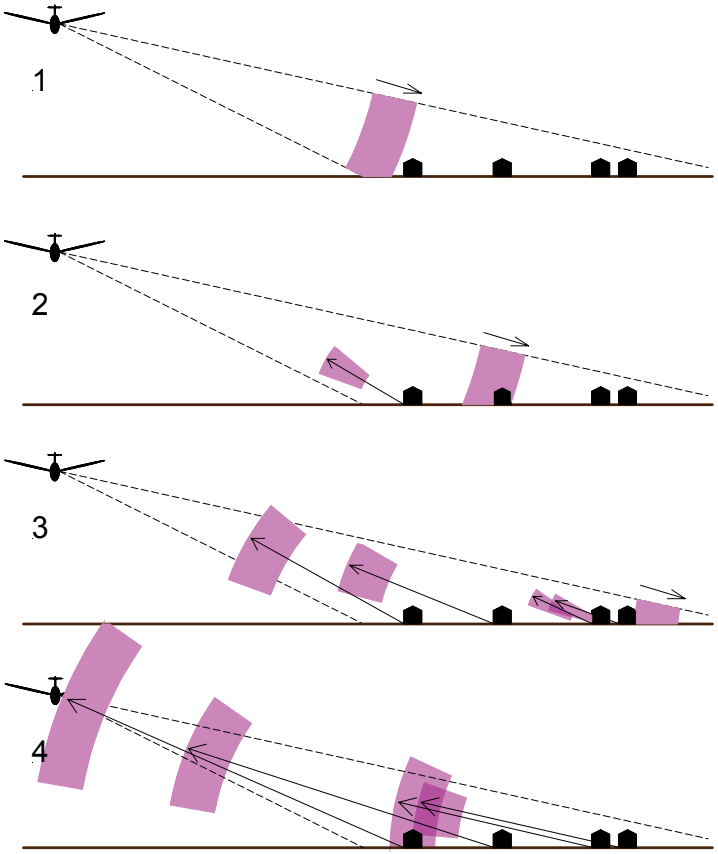


Figure 5.2: Operational principles of a pulsed time-of-flight radar illustrate that the echoes from the nearer houses return to the radar first and that the echoes from the last two houses overlap and appear to the radar as a single return

Time-of-flight is the principle mode of operation for most radar, laser and active acoustic devices. This technique uses the time between the transmission of a pulse and the reception of an echo is measured to provide range.

Because the round-trip time is measured there is a factor of 2 in the formula as shown below,

$$R = \frac{v\Delta T}{2}, \tag{5.1}$$

where R – range (m),
v – wave propagation velocity (m/s),
ΔT – round trip time (s).

5.2.1. Requirements

To operate efficiently a narrow beam must be formed to concentrate the transmitted energy, the transducer must be matched to the characteristics of the medium and the receiver must match the transmitter characteristics.

Some form of modulation must "mark" the carrier signal so that the round trip time can be measured. The figure below shows an amplitude modulated carrier showing the "bang" pulse and then a pair of echoes after a short delay.

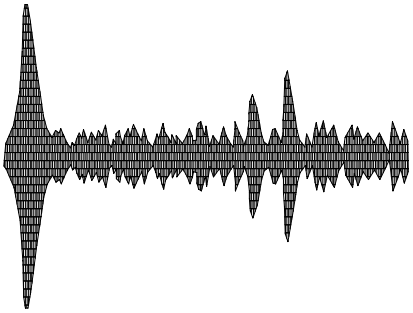


Figure 5.3: Actual received signal showing carrier modulation

It is not easy to process the modulated carrier, so in general it is detected (demodulated) before the timing information can be extracted. The figure below shows the envelope detected output.

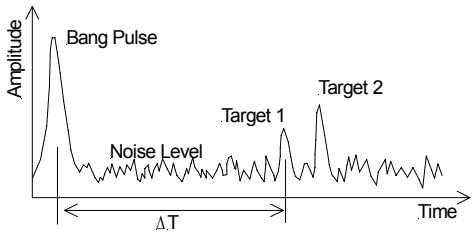


Figure 5.4: Received signal after envelope detection

5.2.2. Speed of Propagation for EM Radiation

Propagation speed is a function of the refractive index of the material

$$v_{matl} = \frac{c}{N} = \frac{c}{\sqrt{\varepsilon_r}} \quad \text{m/s,} \tag{5.2}$$

where c – Speed of Light in a Vacuum (3×10^8 m/s),
N – refractive Index of the material,
ε_r – Relative dielectric constant of the material.

5.2.3. Speed of Propagation of Sound Waves

For acoustic propagation the velocity is a function of the bulk modulus and the density for sound waves as calculated below

$$v_{matl} = \sqrt{\frac{B}{\rho}} = \sqrt{\frac{1}{K\rho}} \quad \text{m/s,} \tag{5.3}$$

where B – Bulk Modulus of the material,
K – Compressibility of the material,
ρ – Density of the material (kg/m^3).

The bulk modulus of the material is defined as ratio of the applied pressure to the strain $B = \dfrac{P}{(V_o - V_n)/V_o}$ Where P is the pressure and V is the volume.

The bulk, B, modulus of air is equal to 1.4 times the pressure. At sea level $P_{air} = 1.01\times10^5$ N/m^2 and $\rho_{air} = 1.20$kg/m^3.

Table 5.1: Speed of sound in different materials

Material	Speed (m/s)	Material	Speed (m/s)
Air (STP)	343.2	Sea water	1450 - 1750
Air 0°C, dry	331.29	Methanol 30°C	1121.2
Nitrogen 0°C	334	Mercury	1451
Oxygen 0°C	316	Aluminium	5000
Carbon dioxide	259	Copper	3750
Helium 0°C	965	Lead	1210
Water 0°C	1402.3	Steel	5250

Note that the temperature, the salinity and the density determine the variation in the speed of sound in the sea. This is addressed in Chapter 8. For time of flight sonar depth sounders, this variation in speed must be considered if accurate measurements are required.

5.2.4. The Antenna

The antenna acts as a transducer between transmission line propagation and free-space propagation.

During transmission it concentrates the radiated energy into a shaped beam that points in the desired direction and only illuminates the selected target and on reception it collects the energy reflected by the target and delivers it to the receiver.

The effectiveness of these two roles is described by the transmitting gain and the effective receiving aperture of the antenna.

In Sonar applications, the gain is known as the *directivity index*

Either one or two antennas can be used.

The gain and beamwidth are determined by the size of the antenna in terms of the wavelength radiated,

$$G = \frac{4\pi A}{\lambda^2}, \tag{5.4}$$

where: G – Antenna gain,
A – Frontal area of the antenna (m^2),
λ – Wavelength (m).

The antenna beamwidth is inversely proportional to the gain, so it can be estimated using the following simple formula.

$$G \approx \frac{4\pi}{\theta_{3dB}\phi_{3dB}}, \tag{5.5}$$

where θ_{3dB} and ϕ_{3dB} are the beamwidths in orthogonal planes (rad).

$$\theta_{3dB} \approx \frac{70\lambda}{d} \text{ deg}, \tag{5.6}$$

where λ is the wavelength (m) and d the antenna diameter (m).

The power pattern in the far field is proportional to the square of the field magnitude. $|F(u)|^2$ for both sonar and radar antennas

For a circular aperture the power pattern is

$$P = \left| \frac{J_1[(\pi D/\lambda)\sin\theta]}{(\pi D/\lambda)\sin\theta} \right|^2, \tag{5.7}$$

where J_1 is the Bessel function of the first kind.

The half power beamwidth is $\theta_{3dB} = \sin^{-1}(1.029\lambda/D)$ which corresponds (in the following figure) to $u = 1.616$.

For a rectangular aperture the transmitted (or received) power pattern as a function of the angle off boresight is

$$P = \left| \frac{\sin[(\pi l/\lambda)\sin\theta]}{(\pi l/\lambda)\sin\theta} \right|^2. \tag{5.8}$$

The half power beamwidth is $\theta_{3dB} = \sin^{-1}(0.887\lambda/l)$ which corresponds to $u = 1.393$

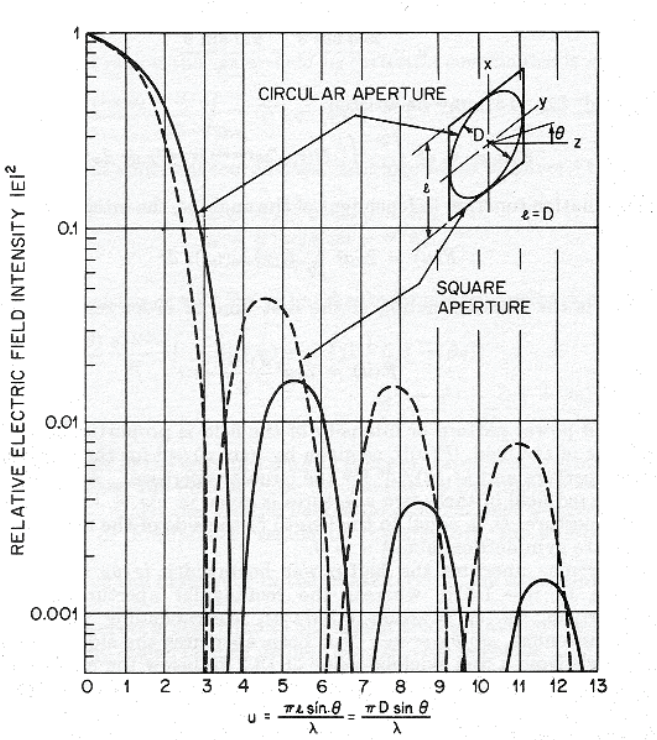


Figure 5.5: Comparison between far field patterns of circular and rectangular antenna apertures

For the Sonar application, the electric field intensity is replaced by the sound pressure. In the following figure, the relationship between the aperture and the beamwidth is compared for three different transducer configurations.

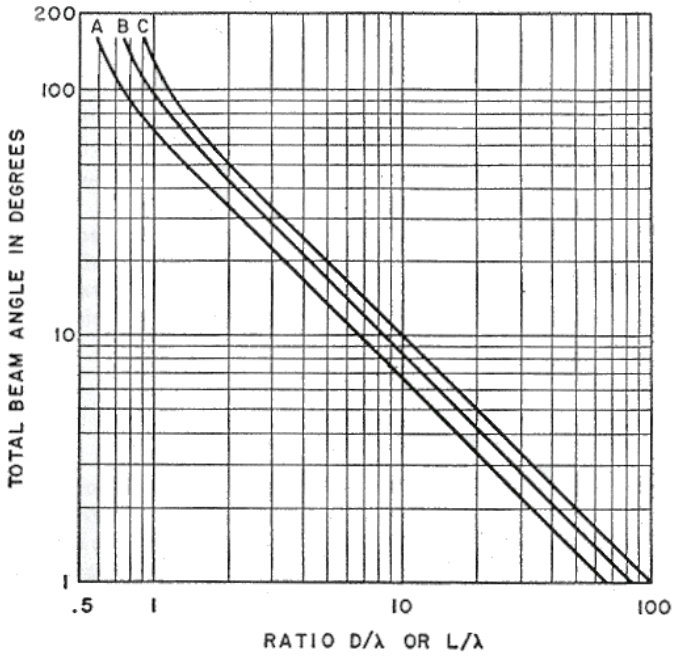


Figure 5.6: Total beam angle for (A) ring diameter, *D*, (B) line length, *L* and (C) piston diameter, *D* relative to the wavelength λ

Circular and Rectangular Antennas

Tracking and ranging antennas are generally symmetrical in the two axes so that they can generate a pencil beam pattern.

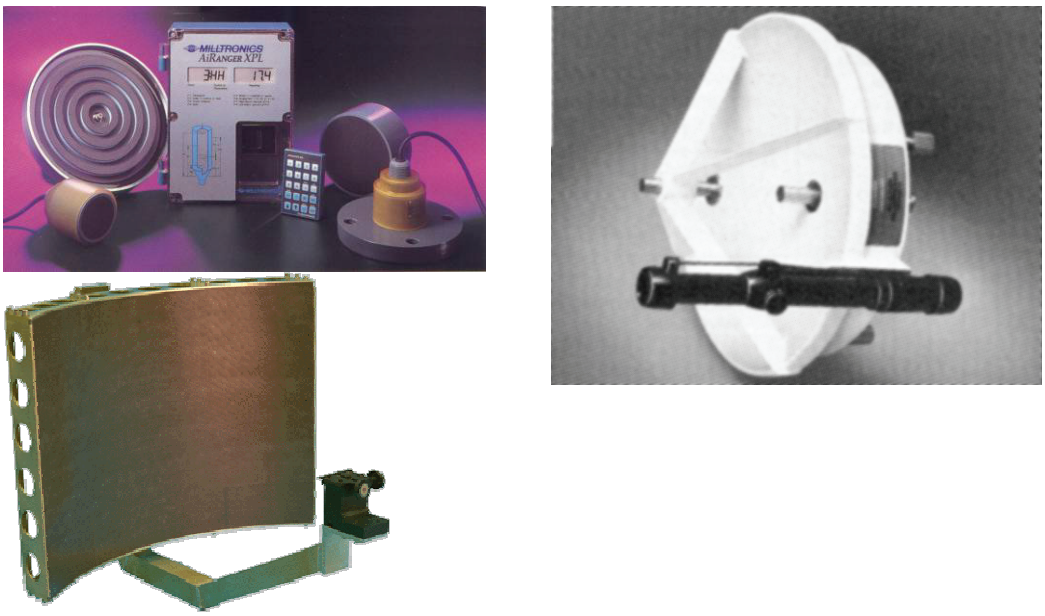


Figure 5.7: Circular and rectangular antennas

Cylindrical Antennas

Antennas and transducers for imaging or searching are generally asymmetrical so that they generate a broad beam in elevation and a narrow beam in azimuth.

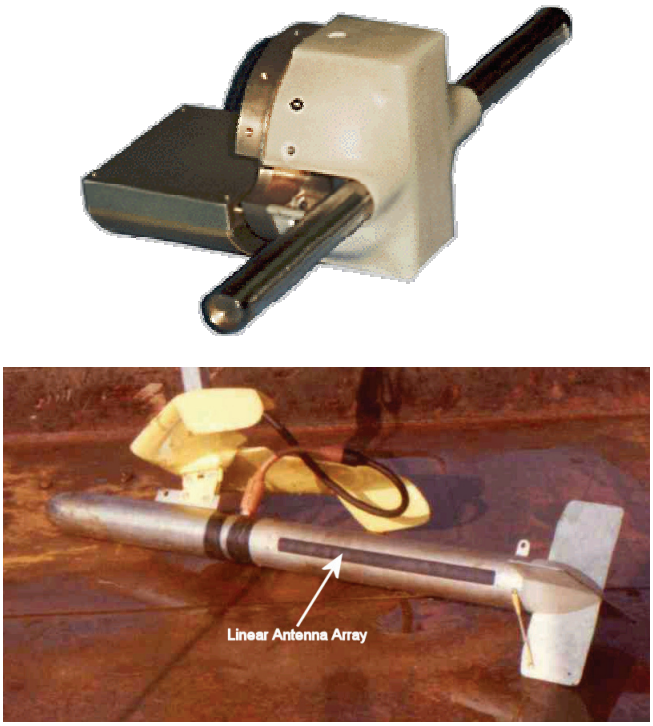


Figure 5.8: Linear antenna arrays

5.2.5. The Transmitter

To mark the carrier so that the time-of-flight can be measured, she signal is "marked" in some way. This is known as modulation, and can be achieved in various ways

- Amplitude modulation (AM)
- Frequency modulation (FM)
- Phase modulation (PM)
- Polarisation modulation

For most simple TOF sensors, the carrier is on-off amplitude modulated as shown in the figure.

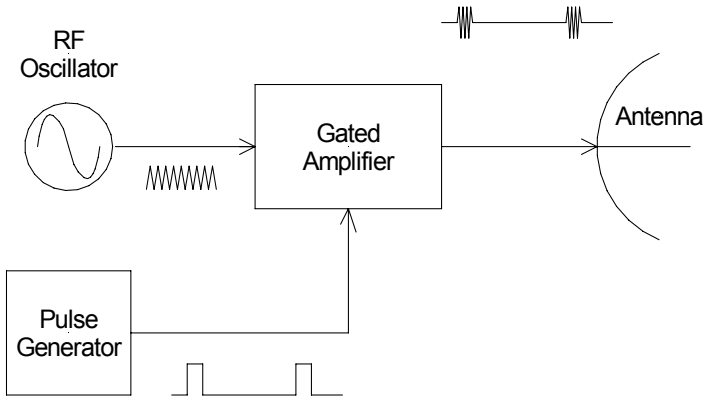


Figure 5.9: On-off amplitude modulation

To maximise the number of pulses that strike the target in a given time, the pulse repetition frequency (PRF) is selected to be as high as possible while still catering for the maximum unambiguous range

$$R_{\max} = \frac{vT_{\max}}{2}, \tag{5.9}$$

where: R_{max} – Maximum unambiguous range (m),
v – Velocity (m/s),
T_{max} – Pulse repetition interval (1/PRF).

The pulse width τ is selected to maximise the amount of energy transmitted (average power) while still giving the required range resolution.

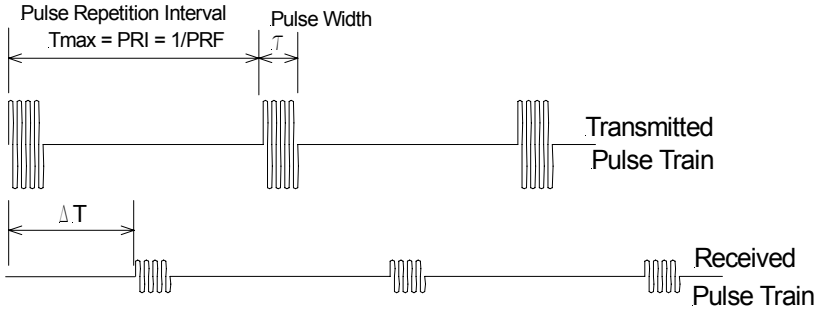


Figure 5.10: Transmitted and received pulse trains

It is possible to use "pulse compression" techniques to increase the average power while still maintaining the required range resolution. This is covered in Chapter 11.

Radar Transmitters

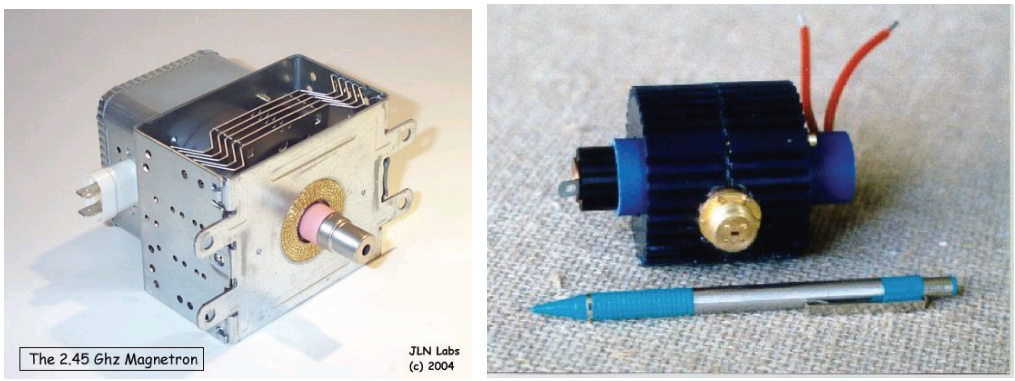
For long range applications, the peak transmit power can be extremely high, many MW for surveillance radars. Microwave tubes such as Magnetrons, Klystrons and Travelling Wave Tubes (TWTs) are generally used to produce the high powers required for radar operation.

The first Magnetron ever made is shown in the figure below. It was made by Randall and Boot of the Physics Department of Birmingham University in February 1940 and had important implications in the successful use of airborne radar during WW II.



Figure 5.11: The first magnetron made by Randall and Boot (without magnets)

Magnetrons are still found in most low-cost radar systems because they are cheap and extremely efficient. The magnetron shown below is typical of the genre.

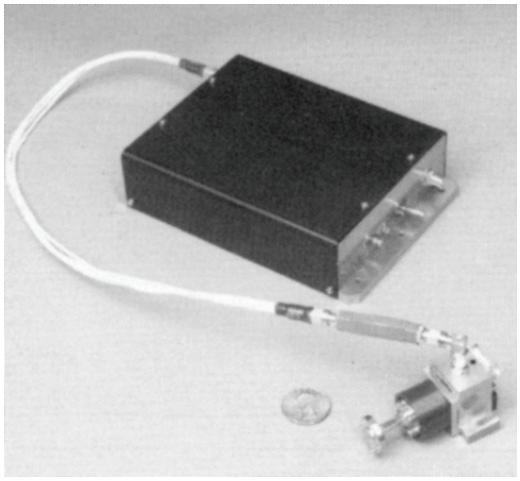


(a) (b)

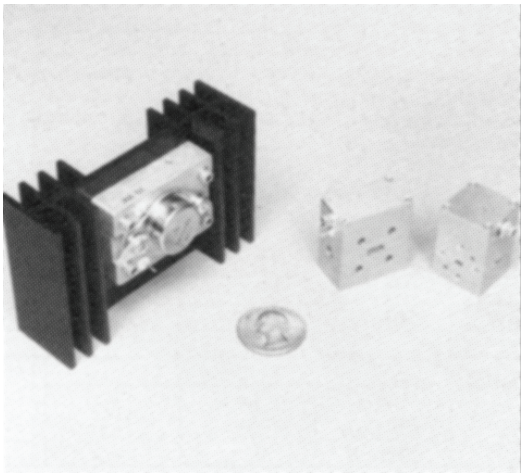
Figure 5.12: Magnetron from (a) microwave oven and (b) 94GHz radar transmitter

For short range applications at high frequency (>10GHz), solid state oscillators using Gunn or IMPATT diodes are often used. These are two-terminal devices which are placed within a resonant cavity where they can be biased to exhibit the appropriate behaviour to sustain high-frequency oscillations.

For operation at lower frequency, voltage controlled oscillators (VCOs) using FETs or high electron mobility transistors (HEMTs) are used. These use the more conventional amplifier and feedback configuration shown in Chapter 2 to produce oscillations.



mmWave IMPATT Oscillator



mmWave Gunn Oscillators

Figure 5.13: Millimetre wave transmitter oscillators

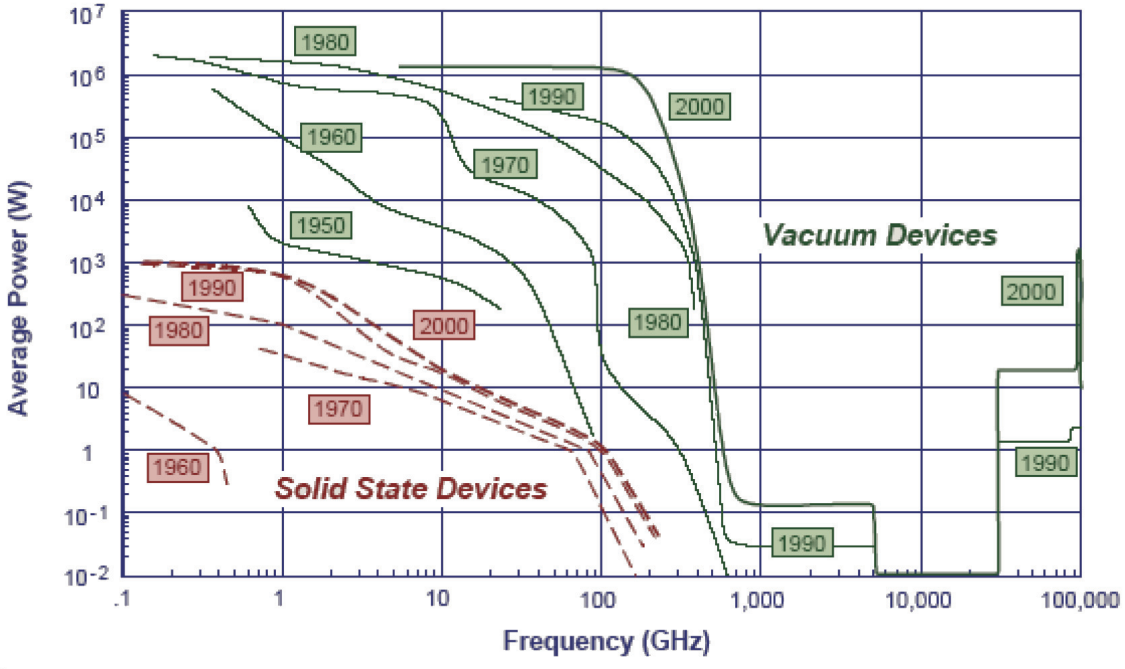


Figure 5.14: State of the art power outputs from various solid state and vacuum tube transmitters

Ultrasonic Transmitters

For short range applications, most ultrasonic transmitters are made from a piezoelectric material. Such materials exhibit a change in size proportional to the applied electric potential.

Natural piezoelectric crystals include Quartz, Rochelle Salt (sodium potassium tartarate) and Tourmaline.

Ferroelectric crystals and ceramics of the barium titanate type and ferromagnetic ceramics of the ferrite type are also used as transducers. These materials exhibit changes in length proportional to the square (or higher power) of the polarisation, and exhibit changes in length which are a function of an applied electric or magnetic field. They are equivalent to piezoelectric and piezomagnetic materials.

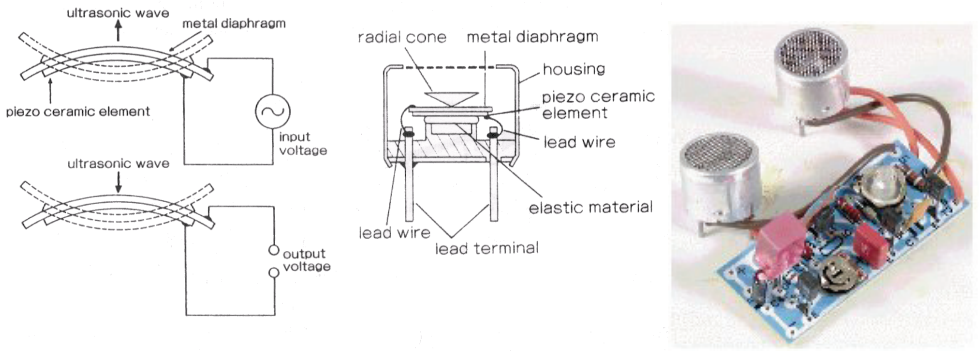


Figure 5.15: Low power ultrasound sonar

Laser Transmitters

Laser transmitters rely on the quantum-mechanical operation of carefully doped PN junctions (diodes) to produce coherent radiation in the IR or visible region when biased by a short current pulse. As shown in the figure below, the cleaved ends of the diode function as mirrors to form a cavity within which the radiation is constrained as it is reflected back and forth.

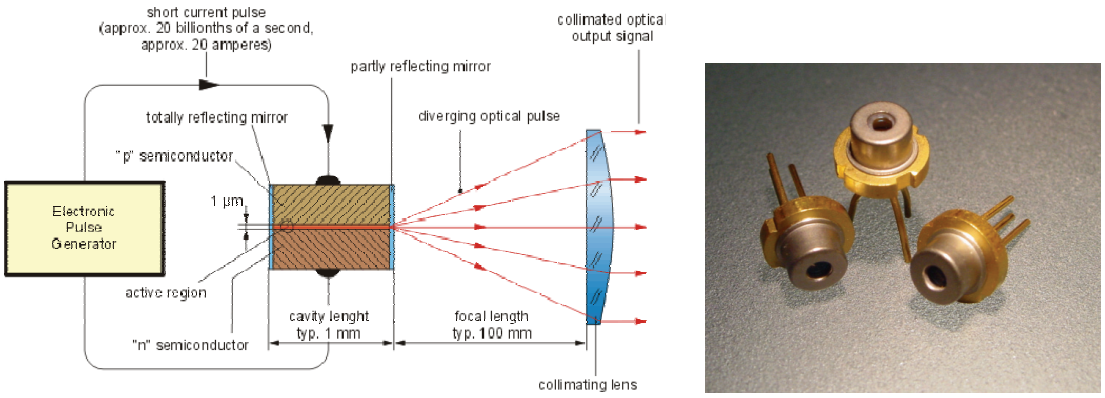


Figure 5.16: Solid state laser transmitter schematic and examples

5.2.6. The Receiver (Matched Filtering and Demodulation)

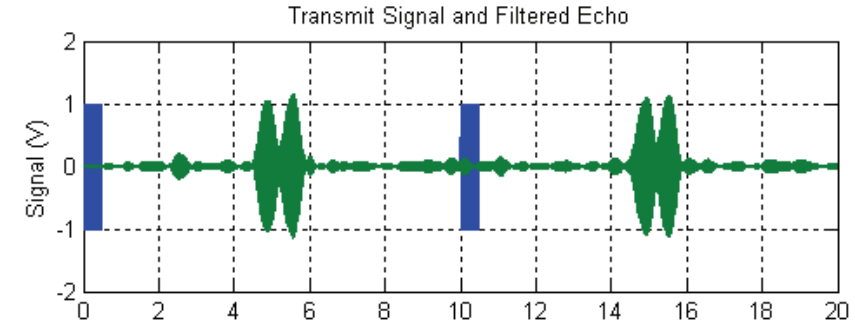
The receiver frequency response is matched to the transmitter modulation pulse characteristics to maximise the peak signal to noise ratio (SNR)

If the bandwidth of the receiver is wide compared to the bandwidth of the transmitted pulse, then noise introduced by the extra bandwidth reduces the SNR

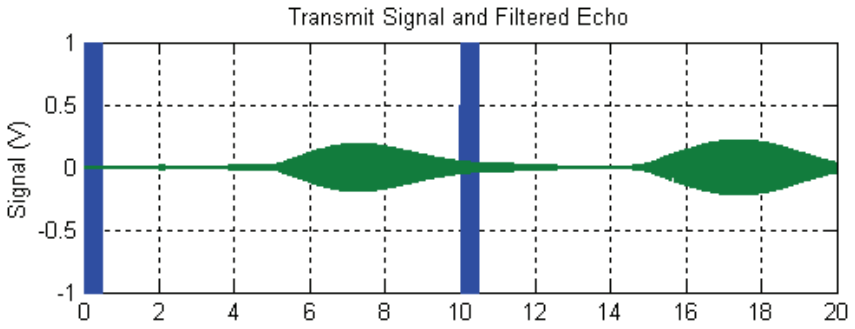
If the receiver bandwidth is too narrow, it will "ring" when it receives a pulse. As a rule of thumb for pulsed measurement systems, the receiver bandwidth β is equal to the reciprocal of the pulse width.

$$\beta = \frac{1}{\tau} \tag{5.10}$$

The following figures show simulation results for an ultrasonic system in which the filter bandwidth is adjusted to illustrate this effect.



(a)



(b)

Figure 5.17: Effect of changing the bandwidth of a matched filter showing (a) the matched bandwidth and (b) the ringing because the bandwidth is too narrow

For other waveforms (pulse compression etc), the matched filter form must be calculated for the specific waveform and the type of noise. This is addressed in Chapter 11.

After filtering, the signal is demodulated (or detected) to extract the timing information from the carrier wave. For basic time-of-flight sensors, this is generally encoded in the amplitude of the carrier (AM).

In most laser sensors the reflected light is detected directly using a fast light-sensitive diode (PIN or avalanche photo diode).

5.3. Pulsed Range Measurement

The simplest method to determine the range to a target is to time the interval from the transmission of a pulse until the echo is received. As shown in the figure below this can be achieved by counting cycles from a high speed clock during this period. If the clock frequency is selected correctly for the propagation velocity, the counter output can be used to display the range directly.

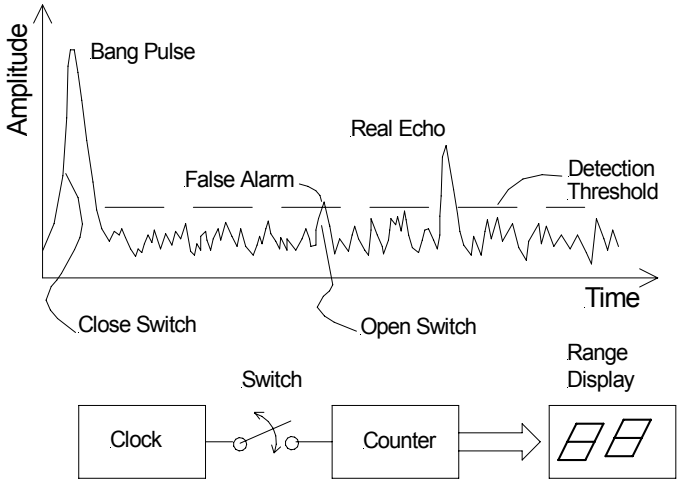


Figure 5.18: Measuring range using a time of flight sensor

In this simple example, the echo signal envelope voltage is compared to a threshold that is adjusted so that the probability of a false alarm (noise alone exceeding the threshold) is low, while there is still a good probability that the signal + noise will exceed the threshold.

The threshold level can be adjusted automatically, or the receiver gain can be adjusted automatically to compensate for the R^{-4} signal-level characteristic. This is called sensitivity time control (STC). It should not be confused with automatic gain control (AGC) which performs a different function.

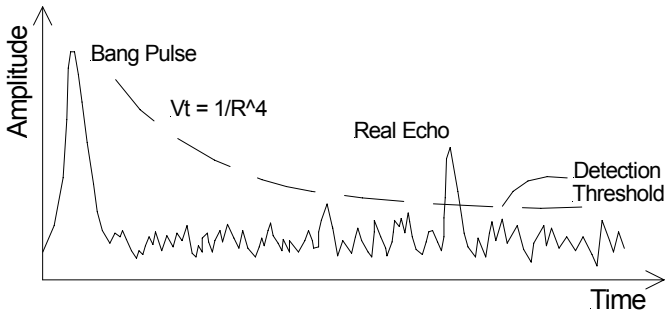


Figure 5.19: Adjustable detection threshold to accommodate propagation losses for targets with a constant cross section

Because the velocity of electromagnetic radiation is so high, this method cannot easily be used to measure the target range to an accuracy of better than about 1m. However it is still possible to measure the range to a fraction of a metre using the split-gate technique described below.

5.3.1. How the Split Gate Estimator Operates

The split-gate estimator generates a linear correction voltage that is proportional to the time of arrival of the echo relative to a clock cycle. Depending on the signal to noise ratio, this correction can improve the accuracy of the range measurement from a few metres down to a few centimetres.

To illustrate the process, in this example it is assumed that the effective pulse width is equal to the +/-1σ limits of the pulse shown in the figure below. For σ = 1, this equates to an effective pulse width of 2 range units.

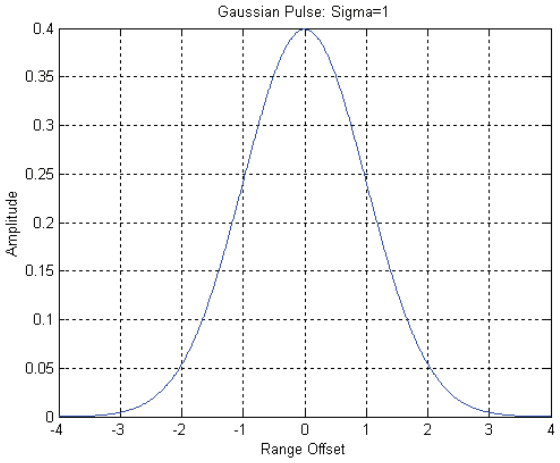


Figure 5.20: Received pulse shape assumed to be Gaussian

Gating process

- A clock with a period of 2 range units runs continuously. This is easy to achieve as a typical LIDAR generates pulses 20ns long. As these are equal to the clock period, its frequency will be 50MHz.
- A comparator with a set threshold starts a counter on the leading edge of the bang pulse. The counter counts on the rising edge of the clock pulse
- A similar comparator stops the counter on the leading edge of the target echo pulse
- This comparator also enables two identical fast sample and hold (S&H) circuits
- These S&H circuits are triggered on the next rising edge of the clock. The first samples the direct echo pulse, and the second samples the same pulse that has passed through a delay (section of coax cable) equal to one half a range unit
- The outputs of these two S&H circuits are combined in the standard normalised split-gate configuration to determine the offset of the pulses from the rising edge of the clock signal.

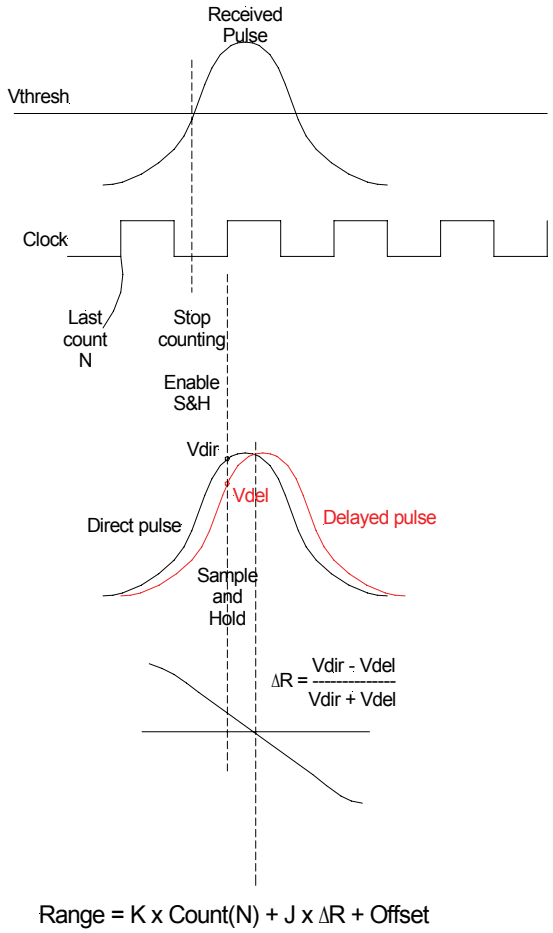


Figure 5.21: Split gate range estimator schematic

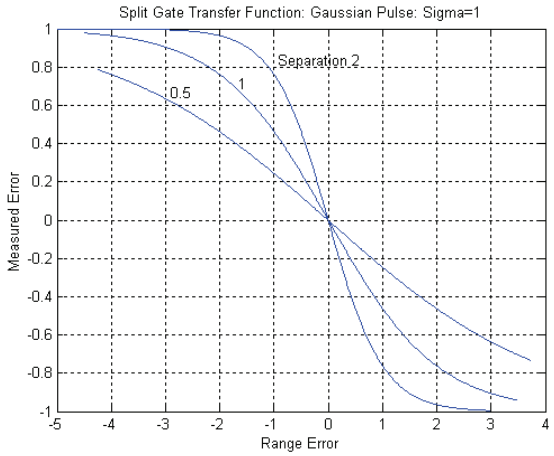


Figure 5.22: Transfer function of different magnitudes of gate separation

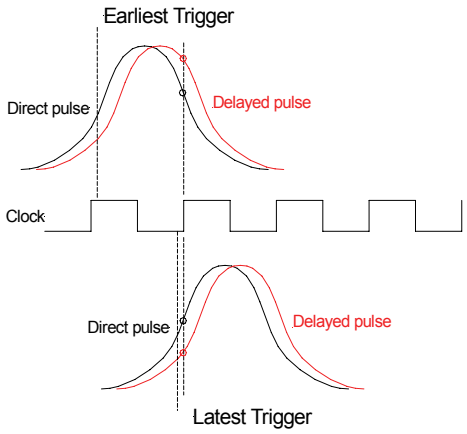


Figure 5.23: Extent of error measurement

The error function must be linear over a single clock cycle (equal to two range units) as shown in Figure 5.23. In Figure 5.24, the linear portion of the transfer function is shown to extend further than that required.

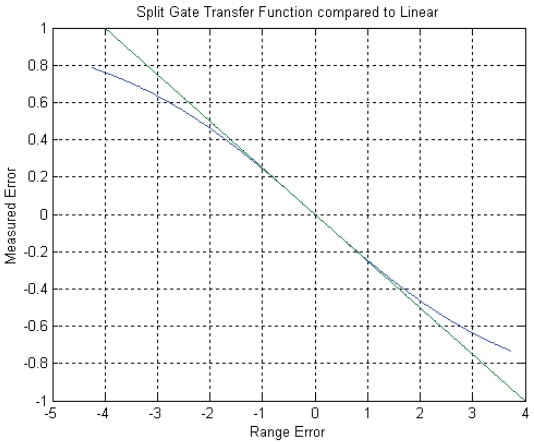


Figure 5.24: Linearity of transfer function: gate separation 0.5R

The following MATLAB script determines the range error transfer function of a split-gate estimator.

```
% split gate technique to measure target range of a laser range finder
% Laser01.m
%
% generate the Gaussian pulse amplitude
r=(-4:0.01:4);
v=exp(-r.*r/2)/(sqrt(2*pi));
plot(r,v);
grid
title('Gaussian Pulse: Sigma=1')
xlabel('Range Offset')
ylabel('Amplitude')
pause

% separate the pulses
dr = 0.5;
vdel = exp(-(r-dr).*(r-dr)/2)/(sqrt(2*pi));
plot(r,v,r,vdel)
grid
title('Delayed and Direct Pulses')
xlabel('Range Offset')
ylabel('Amplitude')
```

```
pause
% Calculate the error transfer function
err = (v-vdel)./(v+vdel);

plot(r-dr/2,err);
grid
xlabel('Range Error')
ylabel('Measured Error')
title('Split Gate Transfer Function: Gaussian Pulse: Sigma=1')
pause

% Measure the error linearity compared to the maximum slope at the origin
slope = -diff(err)/0.01;
smax=max(slope);
y=-smax*r;
plot(r-dr/2,err,r,y)
grid
xlabel('Range Error')
ylabel('Measured Error')
title('Split Gate Transfer Function compared to Linear')
```

5.3.2. Pulse Integration

Pulse integration may be used to decrease the false-alarm rate and to improve the measurement accuracy by increasing the effective signal to noise ratio of the echo. This can be achieved by averaging a number of pulses before thresholding, or by averaging the measured range and discarding outliers.

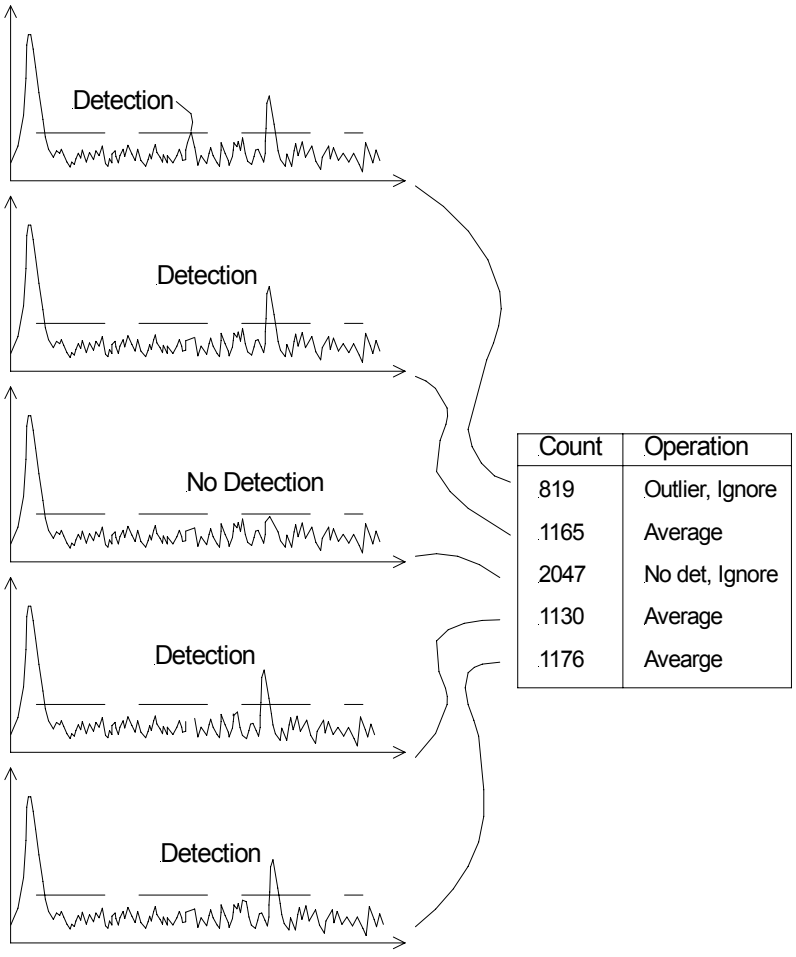


Figure 5.25: Pseudo pulse integration

Integration of pulses requires a storage device that allows the individual echo returns to be averaged. In the old days this was achieved by using a long persistence phosphor on the cathode ray tube radar display combined with the integrating properties of the eyes and brain of the operator. However, modern ranging devices often use digital memory and high-speed signal processing techniques to perform the integration function as shown in the figure below.

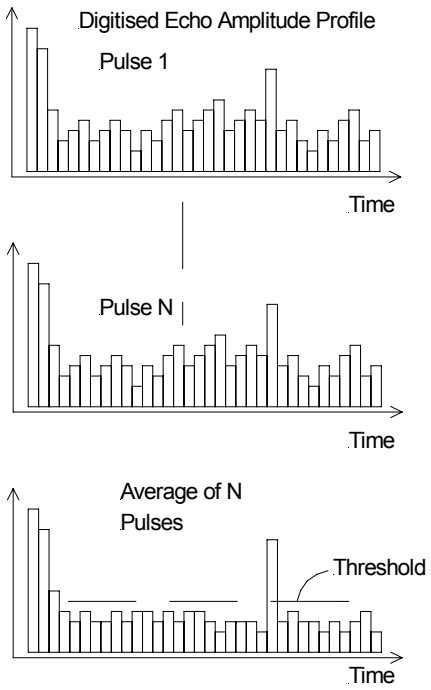


Figure 5.26: True pulse integration

Digital processing of radar signals is limited by the sample rate of the required Analog to Digital (ADC) converters with sufficient dynamic range.

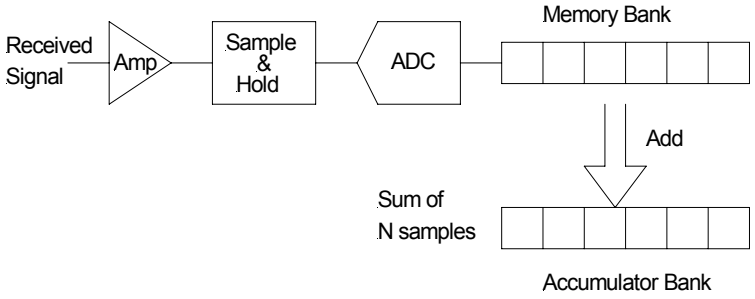


Figure 5.27: Digital pulse integration method

If the pulse width δR = 3m then the aperture time for the Sample and Hold (S&H) should be

$$\tau = \frac{2\delta R}{c} = 20\text{ns}, \tag{5.11}$$

and the ADC must be clocked at 50MHz to capture every sample. It is therefore not practical to use this integration technique for very short-range high-resolution techniques as the S&H aperture would have to be too short.

5.3.3. Time Transformation

Cost sensitive applications such as industrial level sensors and commercial laser range finders use an analogue technique known as “time transformation” to stretch the return time as this leads to a reduction of their clock frequency and ADC requirements. This technique relies on repeated sampling of many pulse cycles.

In this technique, two timing ramps are generated, a fast ramp with a period equal to the pulse repetition interval (PRI) and synchronised to it, and a slow ramp that may span thousands of pulses. A comparator triggers a fast S&H every time the fast-ramp voltage exceeds that of the slow ramp, and the sampled echo signal is held constant for just over one complete PRI cycle. From the figure below, it can be seen that this output tracks the echo sequence but transformed in time from the original PRI by the ratio of the slow-ramp period to that of the fast ramp.

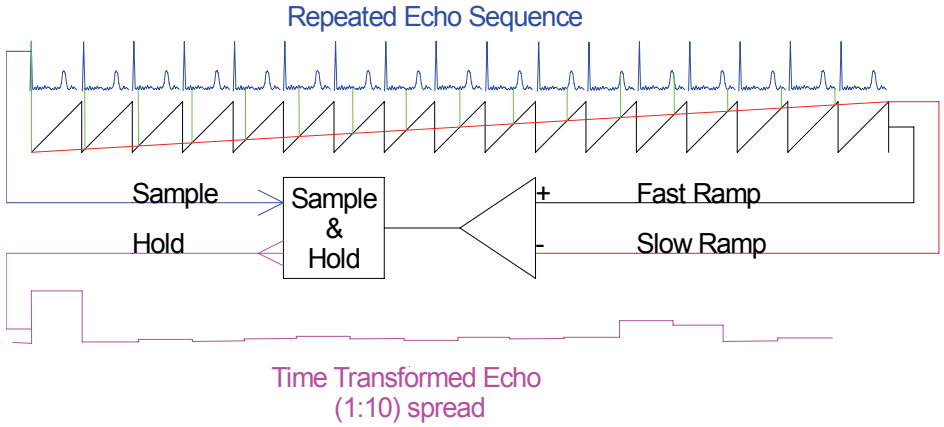


Figure 5.28: Integration by time transformation

Typical transformation ratios of (1:100000) are common in short range level measurement applications as these will transform a time-of-flight sensor that uses EM radiation into the equivalent of an acoustic system

It should be noted that this technique is only required for radar and laser devices because of the high speed of EM propagation. This is not required for sonar devices in air (or even in water) as the speed is sufficiently low.

Finally, time transformation has the added advantage of integrating repetitive signals automatically because each point on the repeated echo is sampled many times and integrated through a low-pass filter

5.4. Other Methods to Measure Range

In theory it is possible to obtain extremely accurate ($\delta R < 1mm$) measurements using standard TOF methods with the split gate process. However, in practise, the pulse widths need to be extremely short and hence the bandwidth must be very high, and this makes achieving the required signal to noise ratio extremely difficult. An alternative is to use the relative phase of the received signal to determine the range.

The original phase reference technique was developed by Dr. T.L.Wadley of the South African National Institute of Telecommunications in 1957 and implemented as a microwave Tellurometer MRA-1.

Phase reference measurements are slow and are only suitable for point targets, so if a faster update rate is required and multiple targets are present in the beam, then more conventional wideband modulation is used.

5.4.1. Ranging using an Unmodulated Carrier

The most basic ranging method involves measuring the phase shift of the unmodulated carrier at at least two different frequencies to determine the range as shown in the following figures.

The equation that defines the range to the target in therms of the wavelength of the signal is

$$2D = n\lambda + \Delta\lambda , \tag{5.12}$$

where D – Distance to the target (m),
λ – Carrier wavelength (m),
n – Number of whole wavelengths,
Δ - Fractions of wavelength.

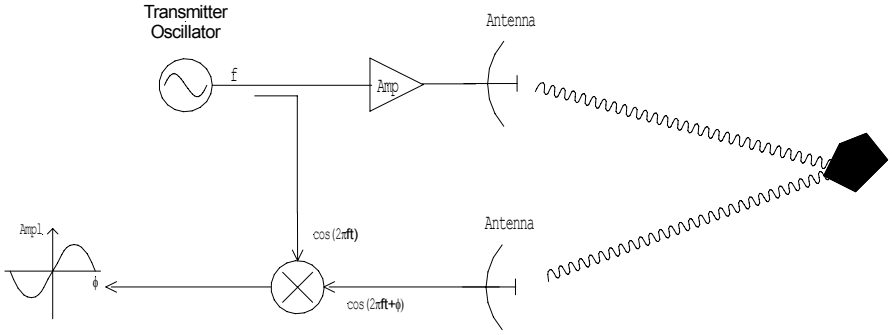


Figure 5.29: Ranging using an unmodulated CW signal

Because it is difficult to measure phase shifts exactly, an alternative technique is to adjust the transmitted frequency so there is a whole number of cycles in 2D.

By shifting the frequency further until this occurs again, two equations can be written and solved for the distance to the target, *D,*

$$\begin{aligned} 2D &= n_1\lambda_1 \\ 2D &= (n_1 + N)\lambda_2 \end{aligned} \tag{5.13}$$

where n_1 – Unknown number of cycles in the round trip distance,
λ_1 – Wavelength at carrier frequency f_1,
N – Number of full cycles of phase shift,
λ_2 – Wavelength at carrier frequency f_2.

Solving for D in terms of the frequency

$$D = \frac{Nc}{2(f_2 - f_1)}, \tag{5.14}$$

where c – Speed of light (m/s),
f_2 – Frequency 2 (Hz),
f_1 – Frequency 1 (Hz).

5.4.2. Ranging using a Modulated Carrier

Most systems use a master station and a remote transponder which amplifies and retransmits the signal to maximise the operating range, however, this is not required if a good reflecting surface is available. As before the round trip distance between the master and the transponder can be given in terms of wavelengths

$$2D = n\lambda_m + \Delta\lambda_m \tag{5.15}$$

where λ_m – Modulation signal wavelength (m)
n – Number of whole wavelengths
Δ - Fractions of wavelength

With this equation, both n and D are unknown, only the phase difference $\Delta\lambda_m$ can be measured using the principles shown in the following figure.

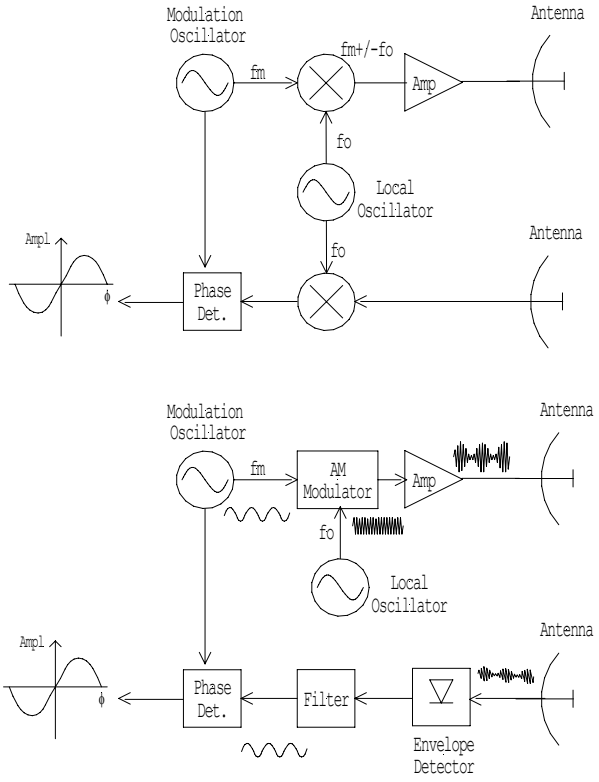


Figure 5.30: FM and AM modulated carrier based CW range measurement techniques

A wavelength of 10m is a compromise between the ability to resolve phase differences (larger λ) and the ultimate resolution (smaller λ)

EM radiation with λ = 10m equates to f = 30MHz for which it is impractical to make directional antennas for portable operation. So, in general, a higher frequency carrier is used that is modulated with a 30MHz signal.

Resolving the integer ambiguity, n, can be achieved in a number of ways

Decade modulation involves stepping λ multiples of 10, eg, 10km->1km->100m->10m->1m etc. The range must be known to within a few km. This technique is used by the CA1000 and MA100 Tellurometers.

This process was manual and time consuming as the results had to be tabulated and the range calculated for each measurement. It also requires the modulator frequency to be very accurate and stable.

Modern IR systems like the MA-200 use a series of similar, stable frequencies. The phase difference is measured for each frequency and the data used to solve a series of simultaneous equations.

Phase Detection

A new range of wideband phase detector ICs such as the AD8302 perform accurate phase comparisons over a wide range of frequencies from close to DC up to 2.7GHz and over a wide range of input powers (-60dBm to 0dBm)

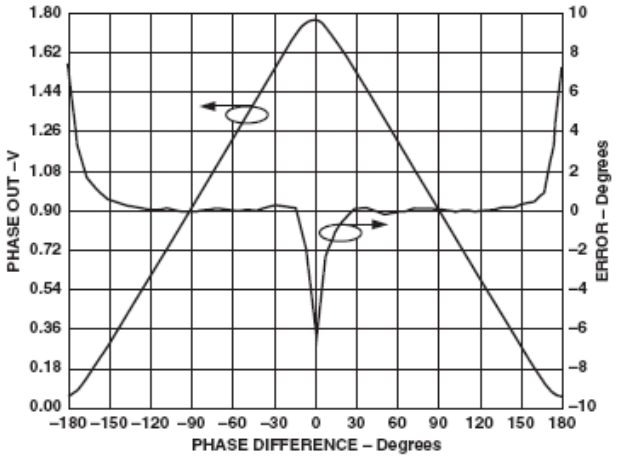


Figure 5.31: AD8302 phase detector performance at 100MHz and -30dBm

The output is calibrated to 10mV/deg with a typical nonlinearity of <1deg.

5.4.3. Tellurometer Example

For example, if the one modulation was set to a 10m wavelength to resolve the distance to the nearest mm, the other would be set to 9.990m.

This would allow the phase to be resolved without ambiguity out to a range of 5km ($2D < 10$km)

$$\begin{aligned} 2D &= n_1\lambda_1 + \Delta\lambda_1 \\ 2D &= n_2\lambda_2 + \Delta\lambda_2 \end{aligned} \tag{5.16}$$

If λ_1 = 10m and λ_2 = 9.99m and the phase difference measurements are respectively $0.0363\lambda_1$ and $0.5989\lambda_2$, what is the measured range.

Because $\Delta\lambda_2 > \Delta\lambda_1$ $n_1 = n_2 = n$, and the equations can be written as follows

$2D = 10n + 0.0363\times10$ [1]

$2D = 9.99n + 0.5989\times9.99$ [2]

[2]-0.999[1]

$2D(1\text{-}0.999)=9.99n\text{-}9.99n+9.99(0.5989\text{-}0.0363)$

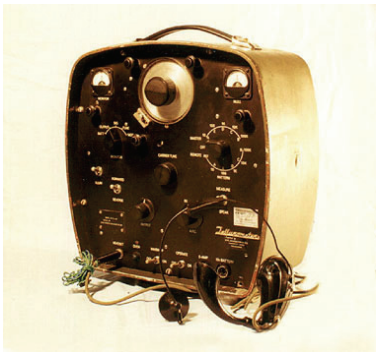
$2D = 5.62036/0.001$

$2D = 5620.363$m

If $\Delta\lambda_2 < \Delta\lambda_1$ then the phase has wrapped by a complete cycle and $n_2 = n_1+1$ and the equations solved accordingly.

5.4.4. Tellurometer Systems

MRA-101 (1962)



- Long range operation up to 50km in master/slave mode
- Narrow beamwidth <6° from 330mm diameter antenna
- Measuring time about 1minute
- Accuracy +/-(1.5cm + 3ppm)
- Frequency range 10.05 to 10.45GHz produced using a Klystron oscillator
- Modulation frequency 7.5MHz

Figure 5.32: MRA-101 Tellurometer

CA-1000 (1970)



- Range 50m to 30km in master/ slave mode
- Slant polarised horn antenna size selected for 10 or 30km operation
- Accuracy +/-(1.5cm + 3ppm)
- Frequency range 10.1 to 10.45GHz produced using a Gunn oscillator
- Will operate as a point to point communication device

Figure 5.33: CA-1000 Tellurometer

MA-100



- Short range (<3km) high accuracy measurements
- Transmitter 0.92µm GaAs diode
- Photodiode receiver
- Amplitude Modulation (AM) frequency 75MHz
- Accuracy at 3km +/-1cm
- Phase measurement performed digitally using a pulse counting system

Figure 5.34: MA-100 Tellurometer

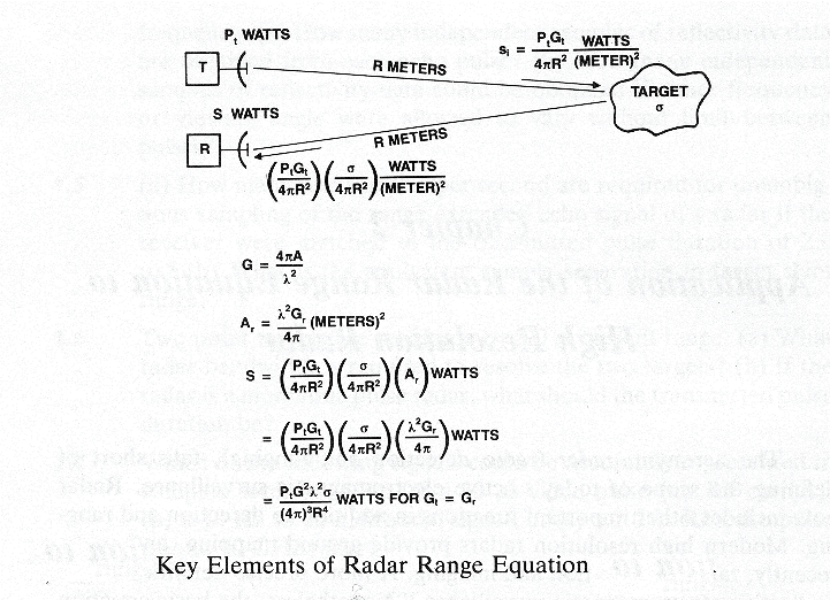
MA-200

- Range 10cm to 1km using a single prism
- IR laser (780µm) based transmitter with a beamwidth of 1.7mrad
- Accuracy +/-0.5mm +0.5mm per km
- Measurement time 45s (1st reading) 25s (subsequent readings)

5.5. The Radar Range Equation

The radar range equation determines the maximum range at which a target can be detected. It takes on many forms depending on what type of system is being described and it can be used to describe Lasers, Radars or Acoustic Systems.

It is generally optimistic because of losses not accounted for and higher than expected noise levels reduce the performance of actual systems.



Key Elements of Radar Range Equation

If the power transmitted by the sensor is P_t through an isotropic radiator (one which radiates uniformly in all directions) then the power density at range R from the transmitter will be equal to the transmitted power divided by the surface area $4\pi R^2$ of an imaginary sphere of radius R.

$$\text{Power density from an isotropic antenna} = \frac{P_t}{4\pi R^2} \quad \text{Watt/area} \tag{5.17}$$

The gain G_t of an antenna is the measure of the increased power in the direction of the target compared to the isotropic case.

The power density S_I at the target is

$$S_I = \frac{P_t G_t}{4\pi R^2} \quad \text{Watt/area.} \tag{5.18}$$

The target intercepts a portion of this power and re-radiates it in various directions.

The measure of the proportion of the incident power re-radiated in the direction of the receiver is denoted as the target backscatter coefficient ρ for a laser radar with a spot size (footprint) smaller than the target, or as the radar cross-section σ for either a laser or microwave radar if the beam footprint is larger than the target.

The radar cross section, RCS, has units of area and is characteristic of a particular target and its orientation.

The power density back at the receiver antenna will be

$$S_R = \frac{P_t G_t}{4\pi R^2} \frac{\sigma}{4\pi R^2} \quad \text{Watt/area.} \tag{5.19}$$

For an effective area of the receiving antenna A_e defined as the ratio of the received power at the antenna terminals to the power density of the incident wave. Then the power received by the sensor is

$$S = \frac{P_t G_t \sigma}{(4\pi)^2 R^4} A_e \quad \text{W.} \tag{5.20}$$

The gain G_r of a lossless antenna is related to its effective aperture by the expression

$$G_r = \frac{4\pi A_e}{\lambda^2}, \tag{5.21}$$

where λ is the wavelength (m).

The effective aperture of antennas that are large in terms of wavelength (d>20λ) can approach the physical antenna size. Substituting for A_e

$$S = \frac{P_t G_t \sigma}{(4\pi)^2 R^4} \frac{G_r \lambda^2}{4\pi} = \frac{P_t G_t G_r \lambda^2 \sigma}{(4\pi)^3 R^4} \quad \text{W.} \tag{5.22}$$

For a monostatic sensor (same receive and transmit antennas) $G_t = G_r = G$

$$S = \frac{P_t G^2 \lambda^2 \sigma}{(4\pi)^3 R^4} \quad \text{W.} \tag{5.23}$$

To account for losses, these are lumped together to reduce the amount of power received and are denoted L where L<1

$$S = \frac{P_t G^2 \lambda^2 \sigma L}{(4\pi)^3 R^4} \quad \text{W.} \tag{5.24}$$

The maximum detection range R_{max} is the range at which the received power just equals the minimum detectable signal S_{min}. Equating R_{max} for $S = S_{min}$.

$$R_{\max} = \left[\frac{P_t G^2 \lambda^2 \sigma L}{(4\pi)^3 S_{\min}} \right]^{1/4} \quad \text{metres} \tag{5.25}$$

5.5.1. Example of Radar Detection Calculation

Determine the maximum detection range of a maritime surveillance radar with the following characteristics

Find the maximum detection range of a maritime surveillance radar with the following specifications

- Gain 34dB
- Transmit peak power 2kW ($10\log_{10}2000 = 33$dBW)
- Frequency 10GHz
- Losses 5dB
- Target RCS 5sqm ($10\log_{10}5 = 7$dBsqm)
- Minimum detectable signal 1.6×10^{-13} W (-128dBW)

Start by converting everything to dB as shown

Apply the radar range equation that has been written for dB values

$$S_{dB} = P_{tdB} + 2G_{dB} + 10\log_{10}\left(\frac{\lambda^2}{(4\pi)^3}\right) + \sigma_{dB} - L_{dB} - 40\log_{10} R \qquad (5.26)$$

Plot the received signal S_{dB} on a semilogx plot along with the minimum detectable power as shown.

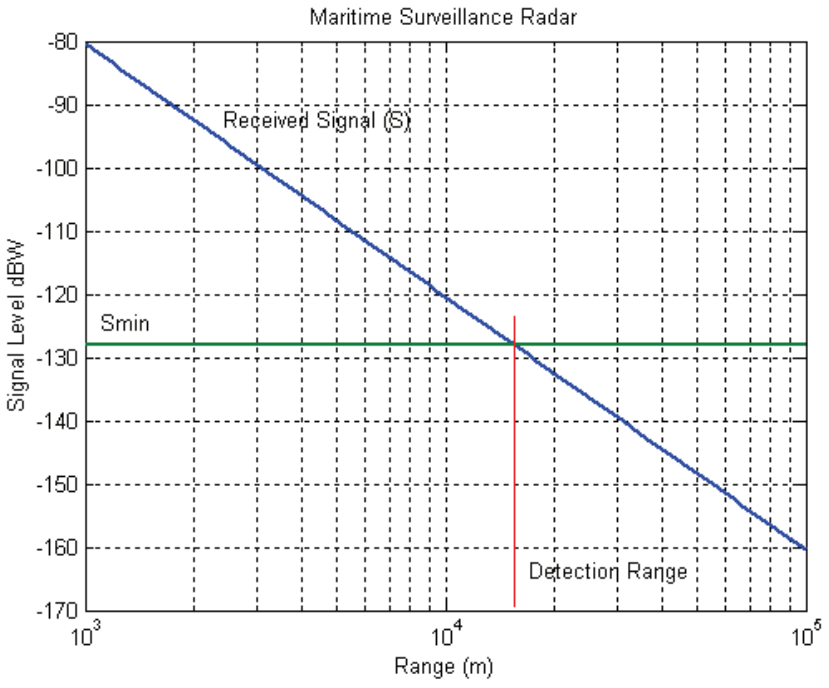


Figure 5.35: Graphical solution to the radar range equation

5.5.2. Receiver Noise

Noise is the unwanted electromagnetic energy that interferes with the ability of the receiver to detect the wanted signal. It may enter the receiver through the antenna along with the desired signal or it may be generated within the receiver.

As discussed in an earlier lecture, noise is generated by the thermal motion of the conduction electrons in the ohmic portions of the receiver input stages. This is known as *Thermal* or *Johnson* Noise.

Noise power P_N is expressed in terms of the temperature T_o of a matched resistor at the input of the receiver

$$P_N = kT_o\beta \quad \text{W}, \tag{5.27}$$

where: k – Boltzmann's Constant (1.38×10^{-23} J/K),
T_o – System Temperature (usually 290K),
β – Receiver Noise Bandwidth (Hz).

The noise power in practical receivers is always greater than that which can be accounted for by thermal noise alone.

The total noise at the output of the receiver N can be considered to be equal to the noise power output from an ideal receiver multiplied by a factor called the Noise Figure, NF.

$$N = P_N NF = kT_o\beta.NF \quad \text{W}. \tag{5.28}$$

5.5.3. Detection Range in terms of Output SNR

Generally the target detection, P_d, and false alarm, P_{fa}, probabilities are given in terms of the output SNR.

The minimum received signal can be written in terms of the output SNR required to achieve a specific P_d and P_{fa} and the pulse width τ

$$\left(\frac{S}{N}\right)_{out} = \left(\frac{S_{\min}}{N}\right)_{in} \beta\tau = \frac{S_{\min}\beta\tau}{kT_o\beta.NF}, \tag{5.29}$$

$$S_{\min} = \left(\frac{S}{N}\right)_{out} \frac{kT_o.NF}{\tau}. \tag{5.30}$$

Substituting S_{min} back into the radar range equation to obtain the detection range

$$R_{\max} = \left[\frac{P_t G^2 \lambda^2 \sigma L \tau}{(4\pi)^3 kT_o (S/N)}\right]^{1/4} \text{ metres.} \tag{5.31}$$

5.5.4. Pulse integration and the probability of detection

If all of the received power from K radar returns is integrated perfectly, then the effective *SNR* is equal to K times the single pulse *SNR* for coherent (pre-detection) integration.

Most radars perform their integration on the video signal after envelope detection as it is much easier to achieve. This is known as non-coherent integration. This is not as effective as coherent integration and the improvement in effective *SNR* is only about $K^{0.8}$ for white thermal noise.

For correlated noise the effectiveness of integration is reduced and is a function of the correlation time and the number of pulses integrated (the integration time).

The following probability density function (PDF) shows the effect of integration on the noise and the target + noise distributions for a non-fluctuating target echo.

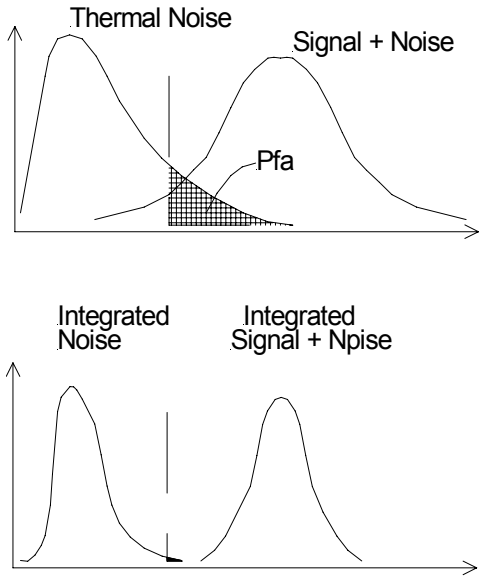


Figure 5.36: Effect of integration on noise and signal PDF's

Note that the means of the PDFs remain unchanged, but their separation (in terms of variance) increases.

For a given threshold, the probability of detection P_d increases and the probability of false alarm P_{fa} decreases.
These probabilities are generally evaluated numerically and displayed graphically as shown in the figure below

Generally, a P_d and P_{fa} are selected for a particular application. The required *SNR* is then determined from the graph. This *SNR* is then reduced by the effective integration gain $10\log_{10}(K)$ for coherent integration and $8\log_{10}(K)$ for non-coherent (also know as post-detection) integration where K is the number of pulses integrated before being plugged into the range equation.

Note that any fluctuations in the target RCS has a major influence on the P_d and P_{fa} or on the detection range. These effects are dealt with in Chapter 10.

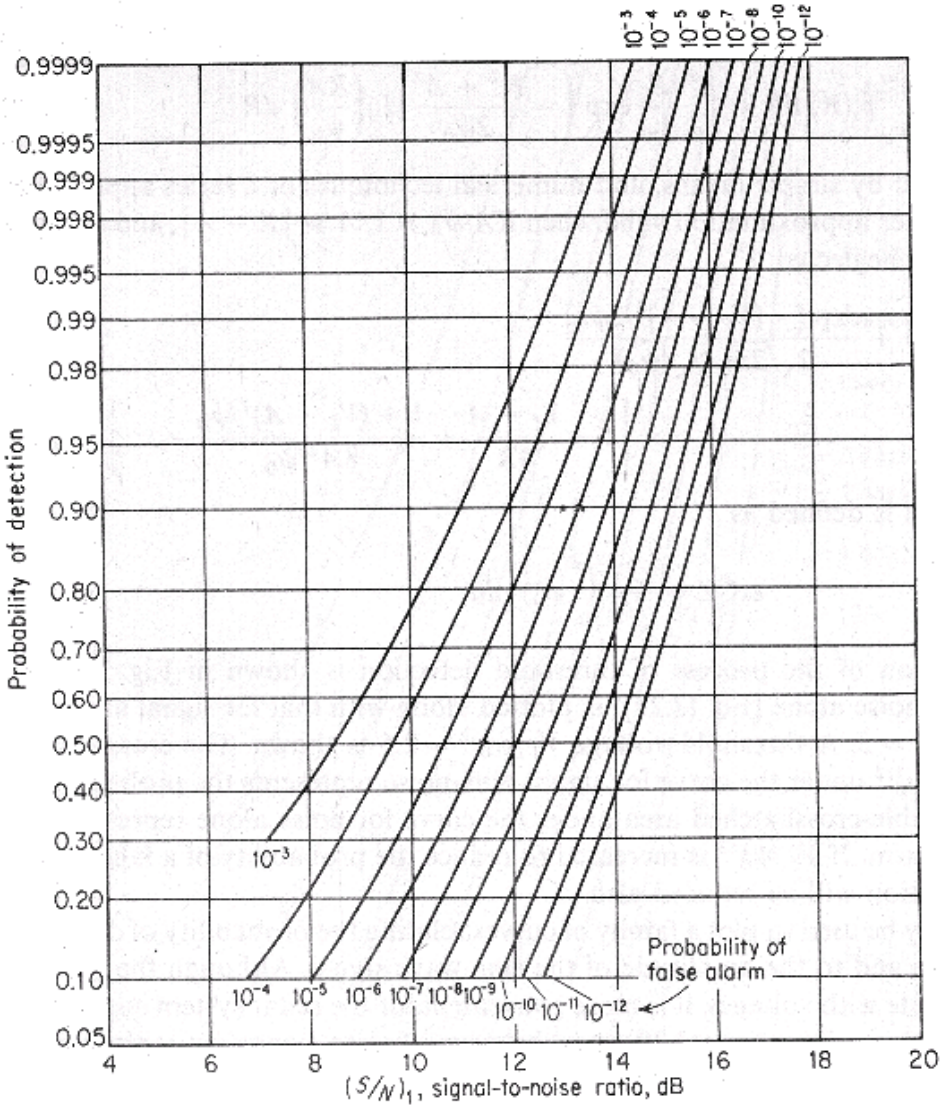


Figure 5.37: Detection probability curves for a sinusoidal signal in white noise

5.6. Issues with TOF Measurement

Electromagnetic waves travel very fast with δT = 6.6ns/m. This means that high data rates are possible but it is difficult to obtain measurement resolution much better than 1m. As discussed earlier in this chapter, for a single point target, techniques exist to achieve measurement accuracies of a few mm.

Acoustic signals travel much more slowly δT = 6ms/m (in the air). Data rates are much lower but good resolution and accuracy (<1mm) is possible with appropriate temperature compensation if the medium is still.

5.6.1. Range Ambiguity

Echoes from large targets at long range may be received after the next pulse has been transmitted. This problem can be eliminated by transmitting at different pulse repetition frequencies.

5.6.2. Beamwidth Ambiguity

Any object within the beam (or the sidelobes) will return an echo and as the actual target range may vary considerably across the beamwidth, there may be significant range uncertainty.

Targets with large cross sections will dominate the return echo even if they are not in the centre of the beam.

Compared to the possible range resolution, angular resolution of typical sensors is poor. It is determined by the effective aperture (in wavelengths) of the radiating antenna, and so is determined to a large extent by the transmission frequency.

Synthetic aperture, monopulse or interferometry can also be used to improve the angular resolution using techniques that are discussed in Chapter 12.

5.7. Examples of Range Measurement Radar

5.7.1. Radar Altimeter

Under certain conditions the radar cross section of a target may be a function of range.

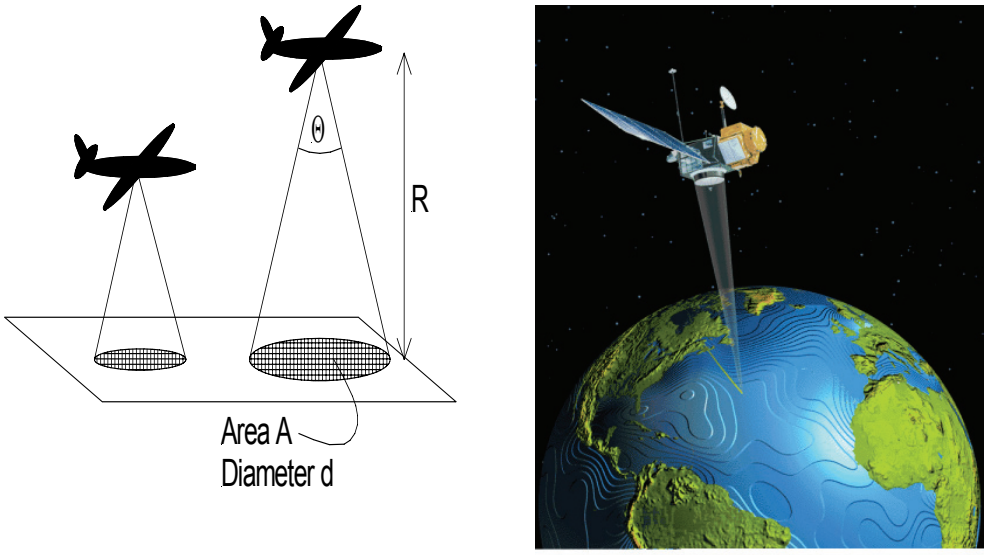


Figure 5.38: Variations in the target cross section for a radio altimeter

In the case of a radar altimeter (shown above), the cross section, σ, is proportional to the area of the beam footprint on the ground

For a beamwidth θ, the area will be

$$A = \frac{\pi d^2}{4} = \frac{\pi (R\theta)^2}{4}. \qquad (5.32)$$

The scale factor σ^o is called the target reflectivity and is defined as the radar cross section per unit area. The RCS is found by taking the product of the reflectivity and the area as discussed in detail in Chapter 9.

$$\sigma = \sigma^o A = \frac{\pi(\theta R)^2 \sigma^o}{4}. \tag{5.33}$$

Substituting into the basic radar range equation reduces the received signal power to a function of R^{-2}

$$S = \frac{P_t G^2 \lambda^2 \theta \sigma^o L}{4^4 \pi^2 R^2}. \tag{5.34}$$

Because typical radar cross sections are large, and ranges are short, altimeters can transmit low average powers (100-500mW), and if required operate using low probability of intercept (LPI) spread spectrum waveforms

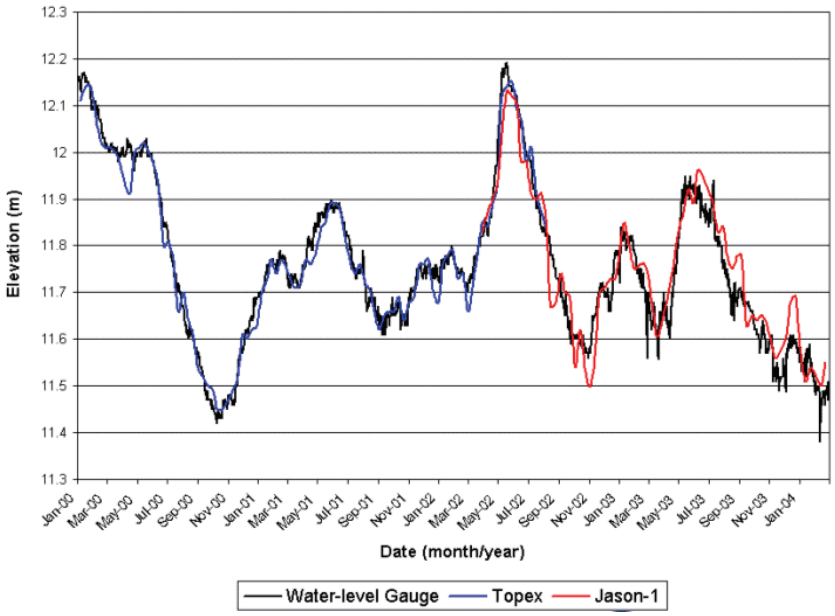


Figure 5.39: Measurements made of Lake Victoria comparing radar altimeter data with that of a water level gauge

5.7.2. Cruise Missiles and Standoff Weapons

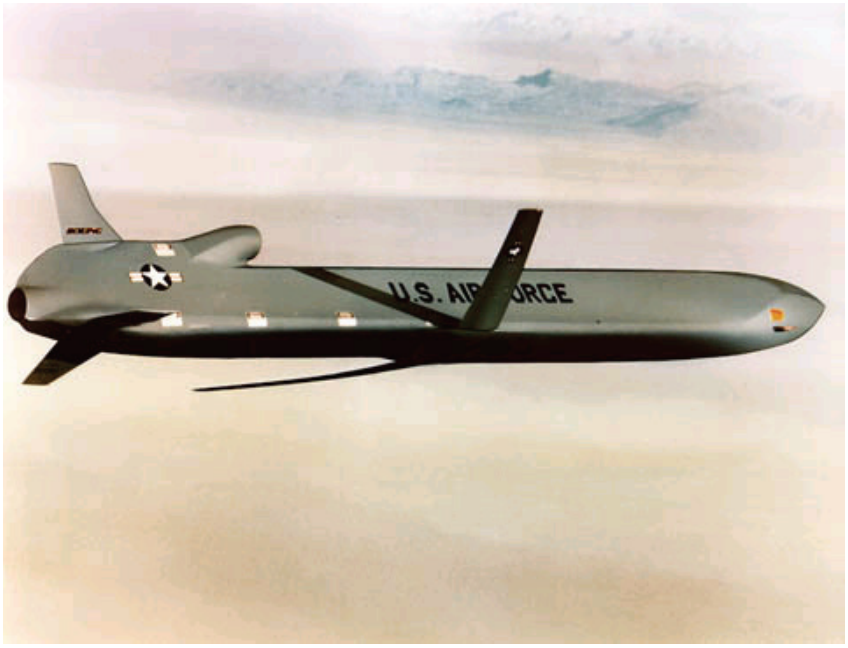


Figure 5.40: Cruise missile in flight

The formal definition of a cruise missile or standoff weapon is a dispensable, pilotless, self-guided, continuously powered, air breathing warhead delivery vehicle that flies like a plane.

The Boeing AGM-86B/C air-launched cruise missile, shown in the figure above, is a long-range subsonic, 1500kg self-guided missile carried by a B-52 bomber at high and low altitudes. Armed with a nuclear warhead, it is designated ALCM. With a conventional warhead, it is designated CALCM. The missile electronically "sees" the terrain over which it flies and can travel more than 2500km to hit the target.

Standoff weapons generally operate over much shorter ranges, hundreds of kilometres, than cruise missiles that can operate over thousands of kilometres.

Unlike ballistic missiles, cruise missiles require continuous guidance since both the velocity and direction of flight can be unpredictably altered by local weather conditions or unexpected changes in the performance of the guidance system.

A ballistic missile is guided for the first 5 of twenty minutes that it takes to fly 5000km; a cruise missile that flies at a subsonic speed would require at least 6 hours of continuously guided flight to cover the same distance.

Guidance errors that integrate with time would thus be 100 times larger for a cruise missile than for a ballistic missile operating at comparable range.

To obtain the necessary location information accuracy, a cruise missile generally uses a combination of inertial guidance that can be augmented by GPS when it is available and some form of map stored in its memory that can be used to correlate its position.

The earliest cruise missile was the German V-1 “buzz bomb” of World War II. However it and the US Matador, Regulus and Snark, and the Russian Shaddock missiles had no way of correlating their positions after launch.

Modern cruise missiles generally use one of two correlation techniques

- Terrain contour matching (tercom)
- Area correlation

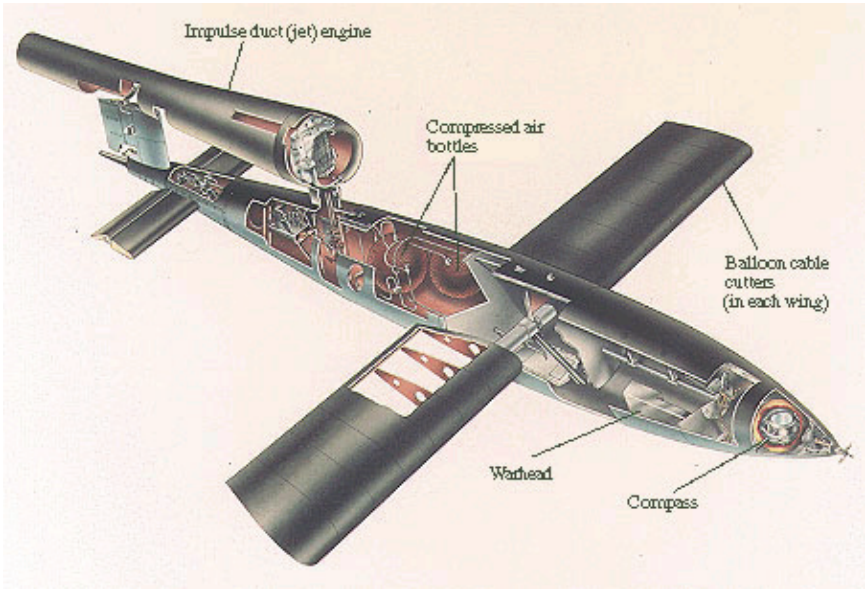


Figure 5.41: The V-1 (Doodle Bug) the original standoff weapon

5.7.3. Terrain Contour Matching (Tercom)

Patented in 1958 it relies on the simple fact that the altitude of the ground above sea level varies as a function of location.

A terrain altitude map with 2km×10km with a cell size of 100×100m would require only 2000 points. So even if 20 such maps were required, the memory requirements are miniscule.

As the missile flies into one of the mapped areas, altitude readings obtained from an on-board radio altimeter are compared with the stored data to determine the missile position to the accuracy of one cell. The autopilot is then instructed to make the appropriate correction.

Inferences about drifts in the inertial system and prevailing winds etc. can be made at each stage, so subsequent maps may be smaller and higher resolution, leading to pinpoint targeting.

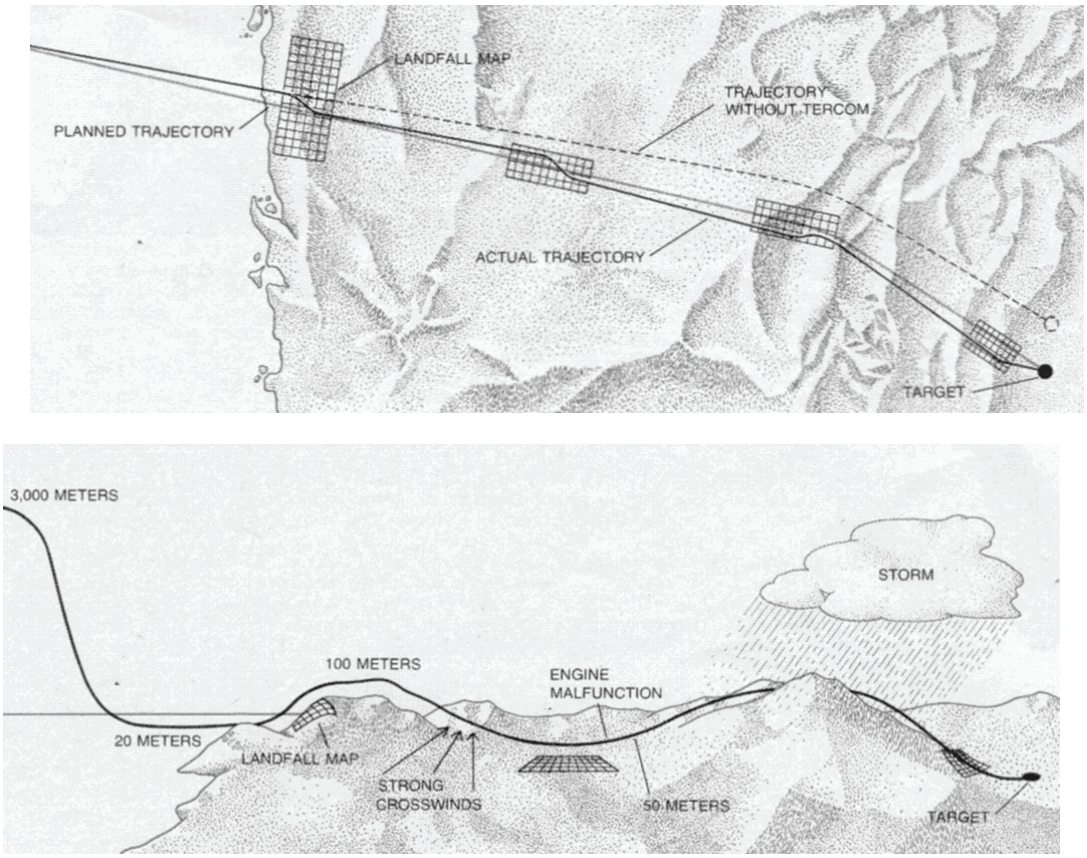


Figure 5.42: Cruise missile map matching

5.8. References

[1] K.Tsipis, Cruise Missiles, *Scientific American,* February 1977.
[2] http://physics.iop.org/Physics/Electron/Exhibition/section5/magnetron.html, 21/02/2001.
[3] http://gallawa.com/microtech/mag_test.com, 21/02/2001.
[4] Hughes Millimeter-Wave Products for 1987/1988
[5] M.Skolnik, *Radar handbook,* McGraw-Hill, 1970.
[6] M.Skolnik, *Introduction to Radar Systems,* McGraw-Hill, 1980.
[8] AGM-86B/C Air-Launched Cruise Missile, www.boeing.com/history/ boeing/alcm.html

Chapter 6.

Active Imaging Sensors

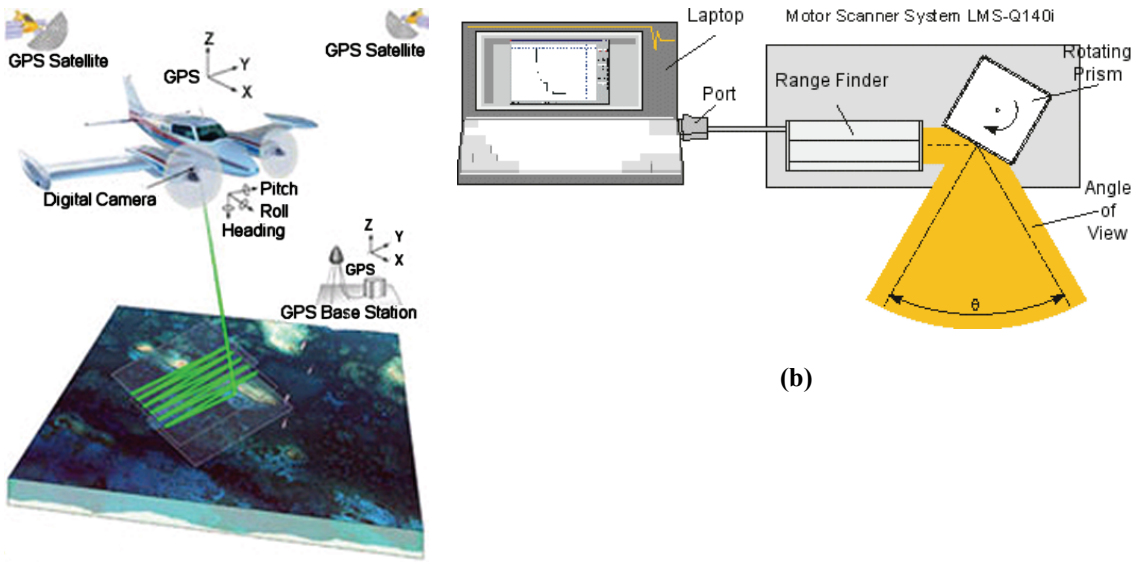
6.1. Imaging Techniques

The two time-of-flight imaging techniques that are commonly used are as follows:

- 1D line scanner with forward velocity to produce a push-broom raster scan
- 2D scanner with orthogonal scan axes

6.1.1. Push Broom Scanner

A rotating prism scans the laser (or radar) beam at right angles to the direction of travel. The laser produces between 2000 and 8000 pulses every second. Each pulse strikes the ground and because it is rough, some power is reflected back to the receiver where it is detected. By registering the forward motion of the aircraft using GPS/INS and the beam angle, a 2D raster is produced as shown in the figure below. Range and /or reflected signal amplitude are logged to produce an image of the ground.



(a) (b)

Figure 6.1: Push broom scan principles showing (a) generation of a raster image and (b) rotating prism mechanism to produce the scanned beam

6.1.2. 2D Scan Principles

The beam from a laser range finder is scanned across a surface using either

- Standard pan tilt unit
- Pan rotating prism unit
- Pan rotating mirror

Pan Rotating Prism Unit

The fast scan axis is implemented by rotating a multi faceted prismatic mirror as shown in the diagram, with this the slow scan axis is implemented by rotating the whole prism scanner module. As with the previous method, range, amplitude or both are logged to produce an image.

High resolution, high accuracy images require many minutes or even hours to construct.

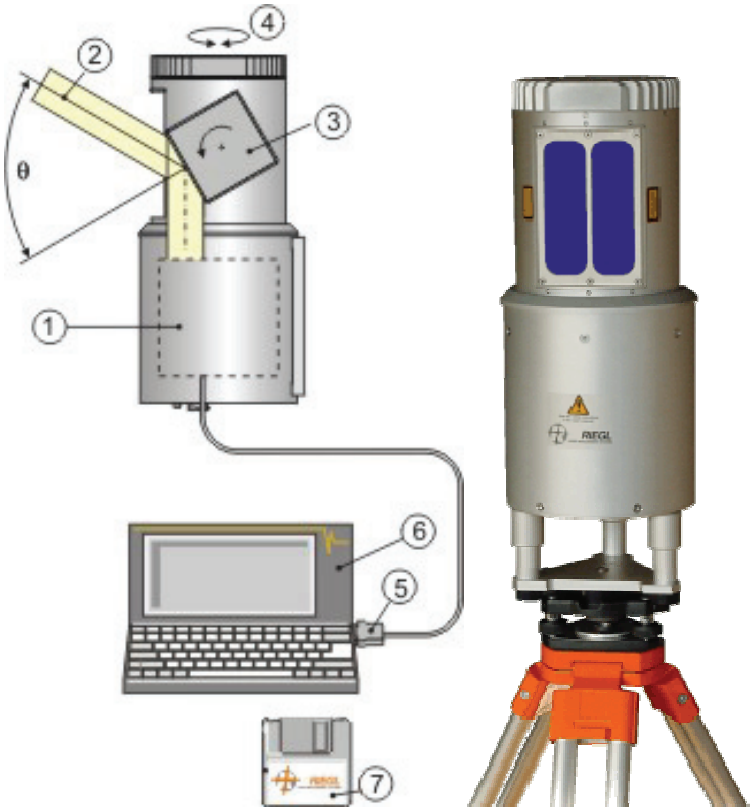


Figure 6.2: Pan prism scanner

1. **Laser Range Finder**
2. **Beam**
3. **Scanned Prismatic Mirror**
4. **Pan Axis**
5. **External Interface**
6. **Display Computer**
7. **Display Software**

Rotating Mirror Scanner

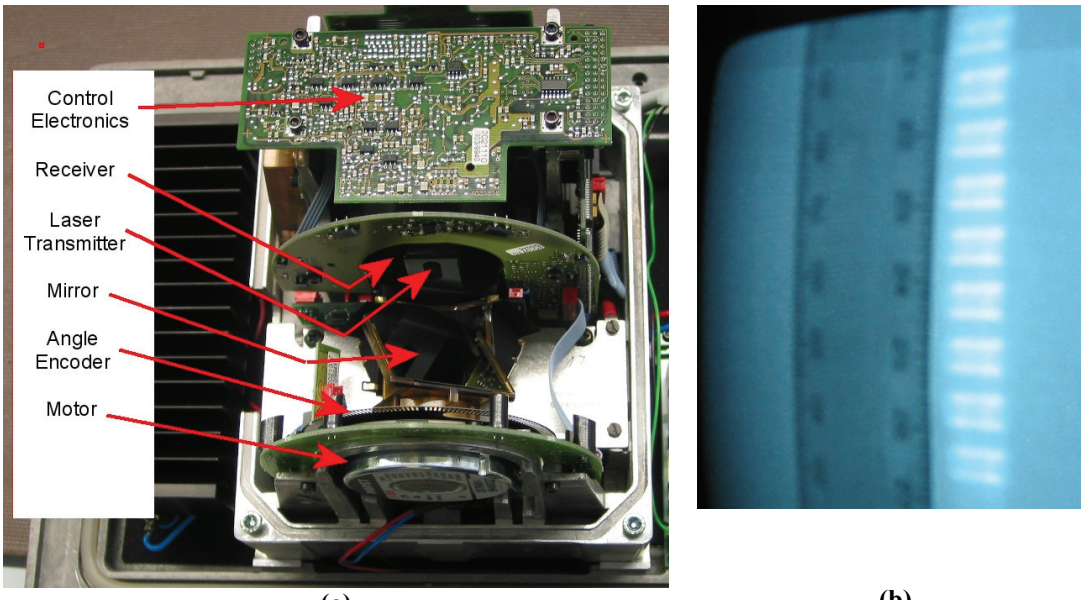
One of the most common imaging systems used in robotics is the SICK laser scanner, an example of which is shown in the figure below.



Figure 6.3: SICK laser scanner unit

The dismantled SICK laser shown in the photographs uses a single mirror mounted on the motor shaft at an angle of 45° to the fixed laser-beam axis to produce a scanned beam.

A mirror behind the laser diode collects the received energy and focuses it onto the receiver diode.



(a) **(b)**

Figure 6.4: Photographs showing (a) interior of a SICK laser unit and (b) an IR image of the laser pulses generated by the unit

6.2. Laser Radar Performance

6.2.1. The Laser Radar Range Equation

Referring to the geometry shown below, the signal power received by the laser range finder at range R is as follows:

$$S = \frac{P}{4\pi R^2} \cdot \frac{4\pi}{(\pi/4)\theta_{BW}^2} \cdot \frac{\pi}{4} \left(R\theta_{BW}\right)^2 \cdot \rho \cdot \frac{1}{2\pi R^2} \cdot A \cdot \tau_o \qquad (6.1)$$

(1) (2) (3) (4) (5) (6) (7)

where S – target signal power at the laser radar (W)
P – transmitted pulse power (W)
R – range to the target (m)
θ_{BW} – laser beamwidth (rad)
ρ - target backscattering coefficient
A – lens aperture area (m^2)
τ_o – optical efficiency

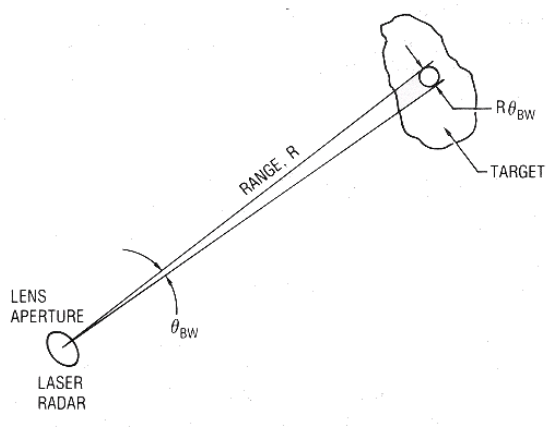


Figure 6.5: Laser radar target area

(1) Point source power over a spherical area $4\pi R^2$. The power of a point source without the benefit of a lens that focuses the power in a given direction.
(2) Measure of the focussing effect of the lens where it is used to direct the radiated power in a given direction. The gain is the ratio of the spread of the power over a sphere of 4π sr and the laser beamwidth in sr.
(3) The area of the target. In this case, the laser spot is smaller than the target area.
(4) The backscatter coefficient of the target. This depends on the material reflectivity and the surface roughness.
(5) The power from the target is scattered equally over the forward hemisphere of 2π sr. The resulting power density back at the laser will be $1/2\pi R^2$.An alternative is to use the Lambertian scattering assumption which is that the reflected flux per solid angle is proportional to the cosine of the angle between the normal to the surface and the reflection angle. In this case the spread is π.
(6) The target return power that is intercepted by a lens with area A.
(7) The optical efficiency of the laser radar transmission chain from the front aperture of the lens.

The formula can be simplified to the following for a spherical target power distribution.

$$S = \frac{PA\tau_o\rho}{\pi R^2} \text{ W.} \tag{6.2}$$

If the laser beam is larger than the target size, then term (3) should be substituted for by the laser radar cross-section σ and the equation becomes.

$$S = \frac{2PA\sigma\tau_o}{\pi^2 R^4 \theta_{BW}^2} \text{ W,} \tag{6.3}$$

where: θ_{BW} - λ/D,
λ - Wavelength (m),
D – Lens diameter (m),

In contrast to the 3dB half-power beamwidth definition used in microwave radar analysis. In optical systems the 1/e=0.367 level is used which equates to a beamwidth of $1.05\lambda/D$, where D is the lens diameter.

Substituting for θ_{BW} in terms of the lens area and the wavelength

$$\theta_{BW}{}^2 = \frac{\lambda^2}{D^2} = \frac{\lambda^2\pi}{4A}. \tag{6.4}$$

The final equation is

$$S = \frac{8PA^2\sigma\tau_o}{\pi^3 R^4 \lambda^2} \text{ W.} \tag{6.5}$$

6.2.2. Laser Receivers

Direct detection laser receivers convert the echo directly into a voltage or current using PIN or Avalanche photodiodes. Heterodyne receivers down-convert the received signal to a lower frequency by mixing with the output of a stable local oscillator. The signal can then be amplified and filtered to enhance the detection process. Because phase information is maintained, such receivers can be used to measure speed by Doppler processing.

Direct Detection

The detector noise for direct detection is given by

$$N = \frac{(A_d \Delta f)^{1/2}}{D^*} \text{ W,} \tag{6.6}$$

where: N – Noise level (W),
A_d – Detector area (cm^2),
Δf – Receiver bandwidth (Hz),
D^* - Detectivity ($cm\text{-}Hz^{1/2}W^{-1}$).

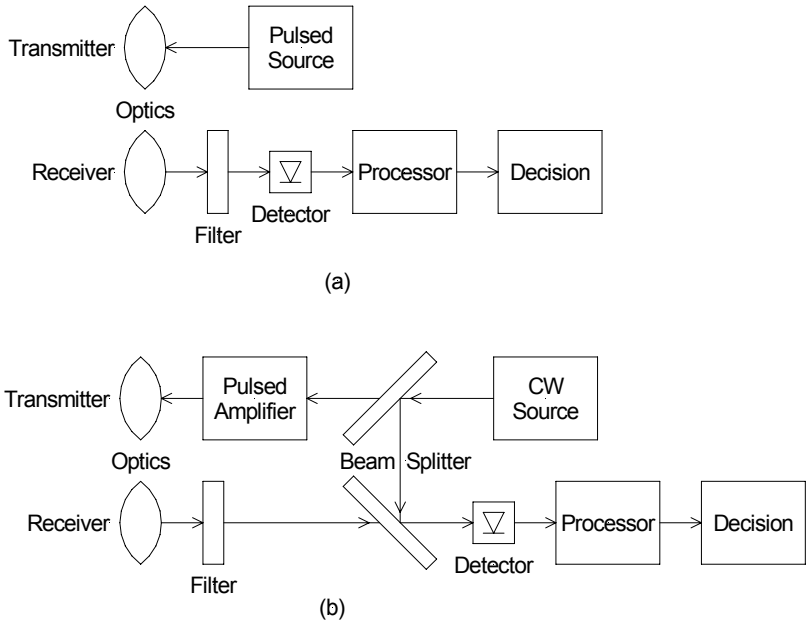


Figure 6.6: Laser receivers using (a) direct detection and (b) heterodyne techniques

The detector area for a square detector is related to the lens diameter and the focal length as follows

$$A_d = \left(\alpha_d^{1/2} f_1\right)^2 = \alpha_d f_1^2 , \tag{6.7}$$

where A_d – Detector area (cm^2),
$\alpha_d^{1/2}$ - Instantaneous detector field of view (IFOV).

The focal length f_1 can be written as the product of the lens diameter (D) and the lens focal number (f/#)

$$f_1 = (f/\#)D . \tag{6.8}$$

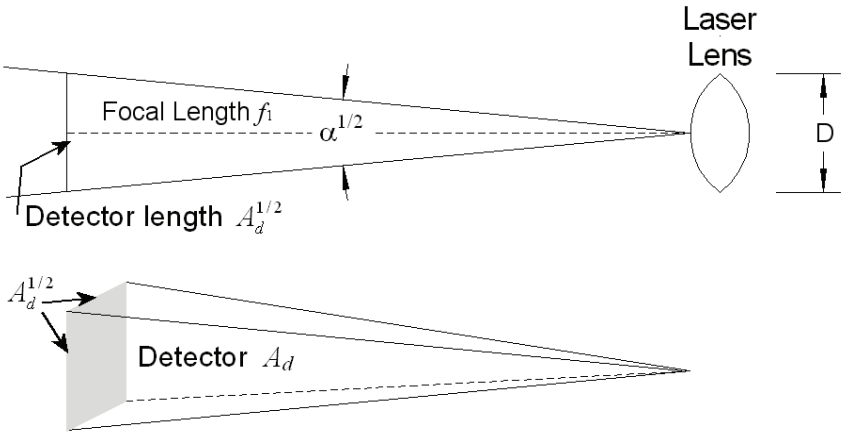


Figure 6.7: Detector area relationship to focal length

Direct Detection Photodiodes

A p-n diode's deficiencies are related to the fact that the depletion area (active detection area) is small and many electron-hole pairs recombine before they can create a current in the external circuit. In the PIN photodiode, the depleted region is made as large as possible. A lightly doped intrinsic layer separates the more heavily doped p-types and n-types. The diode's name comes from the layering of these materials positive, intrinsic, negative — PIN. Figure a below shows the cross-section and operation of a PIN photodiode. The conversion efficiency of these devices is between 0.5 and 1 A/W of incident power.

The avalanche photodiode (APD) operates as the primary carriers, the free electrons and holes created by absorbed photons, accelerate, gaining several electron Volts of kinetic energy. A collision of these fast carriers with neutral atoms causes the accelerated carriers to use some of their own energy to help the bound electrons break out of the valence shell. Free electron-hole pairs, called secondary carriers, appear.

Collision ionisation is the name for the process that creates these secondary carriers. As primary carriers create secondary carriers, the secondary carriers themselves accelerate and create new carriers. Collectively, this process is known as photomultiplication. Typical multiplication ranges in the tens and hundreds. For example, a multiplication factor of eighty means that, on average, eighty external electrons flow for every photon of light absorbed.

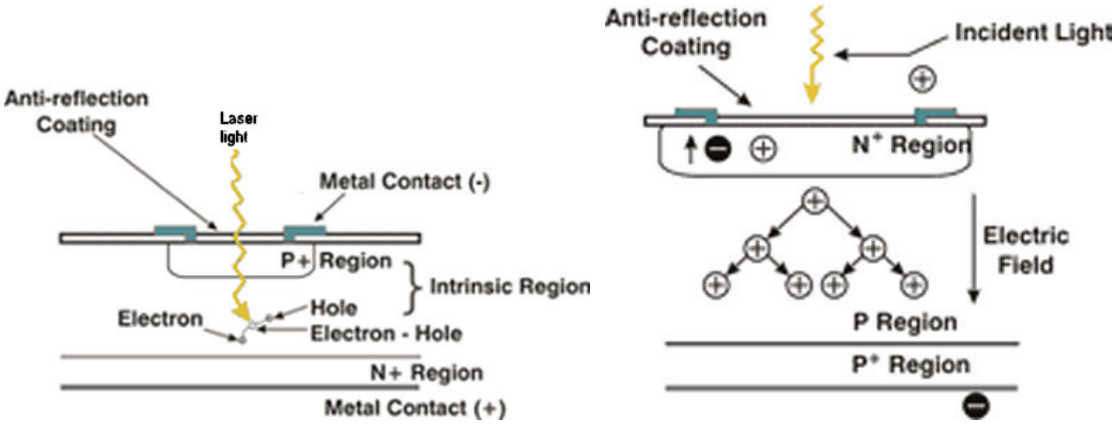


Figure 6.8: PIN (a) and (b) avalanche photodiode (APD) operational principles

APDs require high-voltage power supplies for their operation. The voltage can range from 30 or 70 Volts for InGaAs APDs to over 300 Volts for Si APDs. This adds circuit complexity. Also, APDs are very temperature sensitive, further complicating circuit requirements. Because of the added circuit complexity and the high voltages that the parts are subjected to, APDs are always less reliable than PIN detectors.

The conversion efficiency of typical APDs varies between 0.5 and 100 A/W.

Heterodyne Detection

The noise spectral density of an ideal amplifier is given by the following formula:

$$\psi(f) = \frac{hf}{e^{hf/kT} - 1} + hf \quad \text{W/Hz}, \tag{6.9}$$

where: $\Psi(f)$ – Spectral density (W/Hz),
h – Plank's Constant 6.6256×10^{-34} (Ws^2),
f – Frequency (Hz),
k – Boltzmann Constant 1.38×10^{-23} (Ws/K),
T – Absolute Temperature (K).

Substituting for the wavelength $\lambda = c/f$ and plotting in the figure below. Also plotted is $\mu(f) = hf$ and $\gamma(f) = kT$.

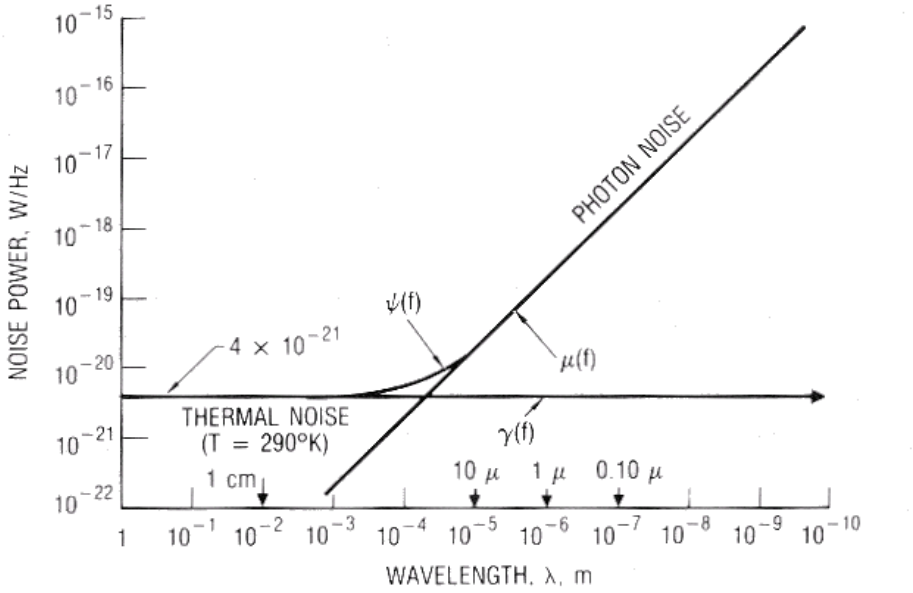


Figure 6.9: Laser radar noise floor as a function of wavelength for heterodyne detection

For microwave radars the noise power density is determined by the thermal noise floor and is approximately $\gamma(f) = kT$ while in the infrared, the noise power density is determined by the photon noise $\mu(f) = hf$.

The noise level of an heterodyne receiver can therefore be written as follows:

$$N = \frac{hfB}{\eta} \quad \text{W}, \tag{6.10}$$

where: η - Quantum efficiency (0.3 to 0.5) (how many photons are required to produce one photo-electron),
B – Receiver bandwidth (Hz).

It is easy to show that at 290K, the noise floor for a microwave receiver is about 10dB lower than a photon noise limited laser receiver operating at 10.6μm with a 50% quantum efficiency.

6.2.3. Signal to Noise Ratio and Detection Probability

Glint targets represent returns from corner reflectors or normal surfaces (such as the ground) where there is a single dominant scatterer. Returns are normally fairly constant from pulse to pulse.

Using the signal to noise ratio for a glint target calculated using the formulae derived above, the detection probability P_d and false alarm probability P_{fa} can be determined from the graph below which give the probability of detecting a sinusoidal signal of constant magnitude in Gaussian noise.

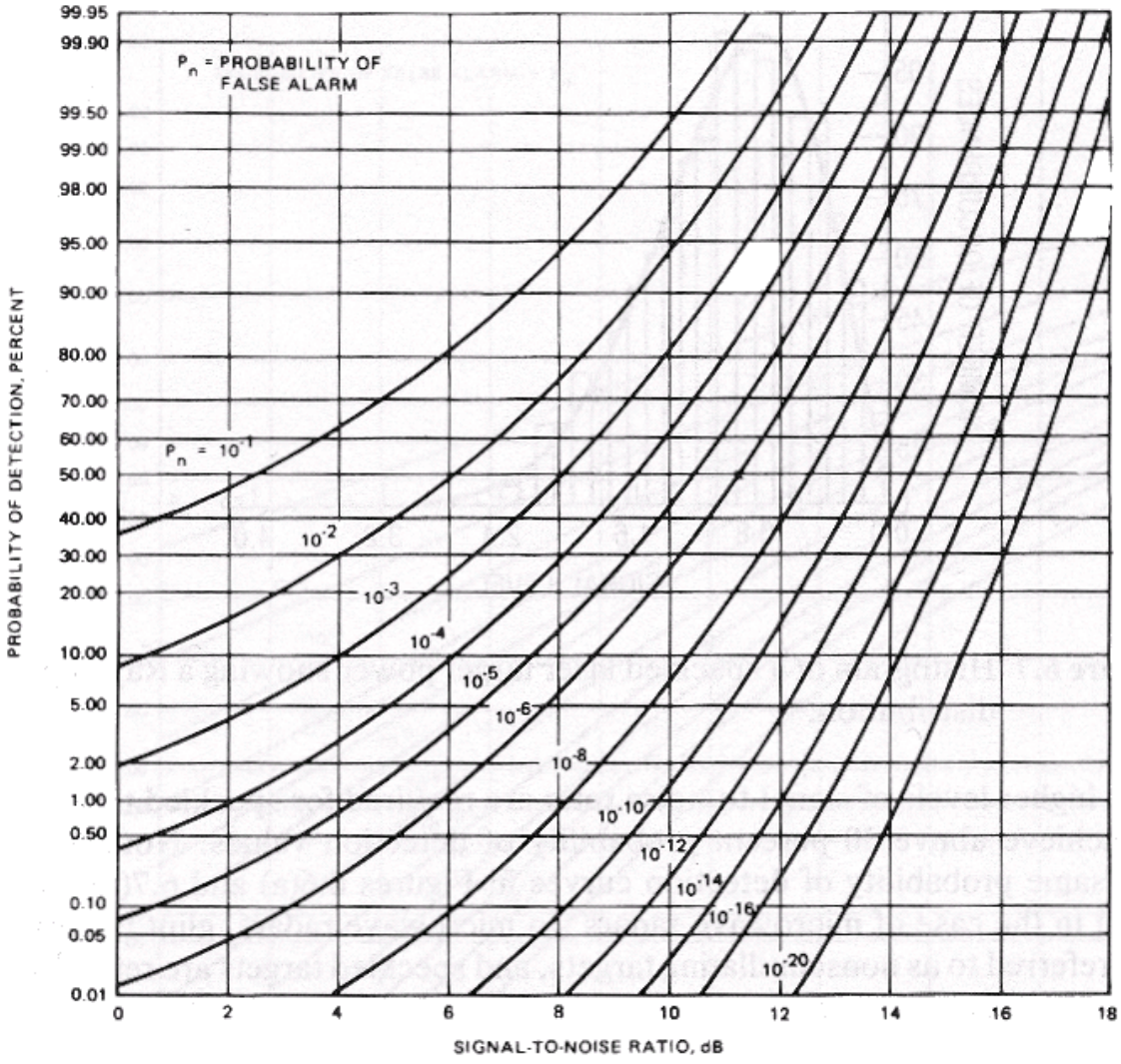


Figure 6.10: Probability of detection of a LIDAR for fluctuating targets

6.2.4. Example of a Laser Radar

An earth bound CO_2 laser operating at a wavelength of 10.6μm radiates through a collimating lens with a diameter of 500mm. If it produces 500W pulses each of duration 0.1s answer the following questions.

Mean distance to the moon is 384400km

The power is $P_{dB} = 10\log_{10}(500) = 27\text{dBW}$

The 1/*e* beamwidth is:

$$\theta_{BW} = \frac{1.05\times\lambda}{D} = \frac{1.05\times10.6\times10^{-6}}{0.5} = 22.3\mu rad$$

The antenna aperture is

$A = \pi D^2/4 = 0.196\ \text{m}^2$

a) What would the diameter of the footprint be on the moon

$d = R\theta_{BW} = 3.844\times10^8\text{x}22.3\times10^{-6} = 8556\text{m}$

The area of the footprint on the moon is

$A_{foot} = \pi d^2/4 = 57.5\times10^6\ \text{m}^2\ [77.6\text{dBm}^2]$

b) Ignoring atmospheric effects what would the power density on the moon be in W/m^2

$S_I = P/A_{foot} = 500/57.5\times10^6 = 8.7\mu\text{W/m}^2$

$S_i = P_{dB} - A_{foot} = 27\text{-}77.6 = -50.6\text{dBW/m}^2$

c) A retro-reflector with a diameter of 10cm and a reflectivity of 0.99 reflects some of the power back to earth. What is the received power density.

The effective cross section of the retro-reflector is (see Chapter 9)

$$\sigma = 0.99\frac{4\pi D^4}{3\lambda^2} = 0.99\frac{4\pi\times0.1^4}{3\times\left(10.6\times10^{-6}\right)^2} = 3.7\times10^6 m^2 = 65.7dBm^2$$

The power density back on the earth is given by the following formula

$$S_R = \frac{2P\sigma}{\pi^2R^4\theta_{BW}^2} = \frac{2\times500\times3.7\times10^6}{\pi^2\times\left(3.844\times10^8\right)^4\left(22.3\times10^{-6}\right)^2} = 3.45\times10^{-17}W/m^2$$

d) Is the reflected power density from the moons surface back on the earth (backscatter coefficient 0.2) larger or smaller than that returned by the retro-reflector.

The power density back on earth is just

$$S_R = \frac{P\rho}{\pi R^2} = \frac{500 \times 0.2}{\pi\left(3.844 \times 10^8\right)^2} = 2.15 \times 10^{-16} W/m^2$$

Which is 10× higher than that received from the corner reflector

e) If an heterodyne receiver uses the same size lens, what is the single pulse signal to noise ratio that we could expect

For a receiver bandwidth matched to the pulsewidth $\beta = 1/\tau$ = 10Hz, and a quantum efficiency $\eta = 0.5$

$$N = \frac{hf\beta}{\eta} = \frac{hc\beta}{\eta\lambda} = \frac{6.625 \times 10^{-34} \times 3 \times 10^8 \times 10}{0.5 \times 10.6 \times 10^{-6}} = 3.75 \times 10^{-19} W$$

Assuming that the optical efficiency is 100%, the received power is the product of the power density and the receive aperture

$S = S_R A = 2.15 \times 10^{-16} \times 0.196 = 4.21 \times 10^{-17}$ W

The single pulse signal to noise ratio is, S/N = 112 (20.5dB)

6.3. Digital Terrain Models

Laser altimeters measure height by transmitting a narrow pulse of IR radiation in a narrow beam towards the ground. Because the ground is rough (compared to the wavelength of the light), the laser pulse is generally scattered in all directions including back up to the aircraft.

Both the roundtrip time and the amplitude of the received echo can be logged, and by using the push-broom technique, discussed earlier, strip-map images of the ground can be generated. Because the swathe painted on the ground is quite narrow (it depends on the height and the scan angle), it is generally necessary to produce composite images produced by successive passes over adjacent, overlapping areas.

To produce well registered, accurate images, the aircraft position is logged using a combination of GPS and inertial navigation systems. Using this technique, the position of a spot on the ground can be determined in this way to an accuracy of 10cm (vertically) and 50cm (horizontally).

6.3.1. Surface Models

A digital image is a rectangular array of cells where each cell contains a single value.

- A topological image is produced if the height information is used
- A reflectivity image is produced if the echo amplitude information is used.

Though the points measured on the ground usually have a non-linear spacing, the cells in the image are spaced regularly (using interpolation) to facilitate processing.

A digital elevation model (DEM) is a continuous mathematical representation describing the shape of the surface where elevation is a function of longitude and latitude.

- Digital surface models (DSM) contain elevation values of the air/surface interface, it includes trees and buildings etc.
- Digital terrain models (DTM) reflect pure terrain information as it is represented on contour maps. These are produced by filtering the DSM information.

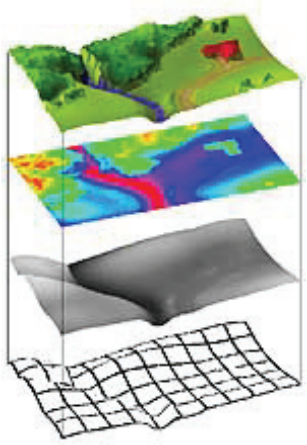


Figure 6.11: Digital model representations

6.3.2. Digital Landscapes

Digital landscapes are DSMs with additional information like surface colour and texture or vegetation types to allow a more realistic representation to be produced. Both DEMs and DSMs are considered to be 2½ -D representations of the earth's surface because only a single elevation value is given for each point on the surface, whereas in reality each point may have a multitude of surfaces (tree canopy, building roof and ground).

If the fine structure of the pulse echo is examined, it yields information about the vertical structure of the surface, roughness, height and shape of manmade objects can be determined as well as canopy density and height of trees. In addition the reflectivity properties can be analysed to produce images similar to those available from infrared cameras (albeit with lower resolution)

6.3.3. Combining Image Types

The usual approach in commercial systems is to use high resolution digital camera in parallel with the laser altimeter. This has the advantage of producing a higher resolution image, and the ability to image across different spectral bands. However, unlike the laser images, it required an external source of illumination which will probably be at an angle to the camera and thus produce shadows.

This first image shows a transformed topological image produced by the scanning laser system. It is possible to see that this image has been manipulated to include shadows which would not occur in the original image which is produced from directly above.



Figure 6.12: Transformed topological image

Individual buildings are resolved to an accuracy of 0.5 to 2m which is sufficiently good to resolve the heights, footprints and even the roof shapes of most buildings.

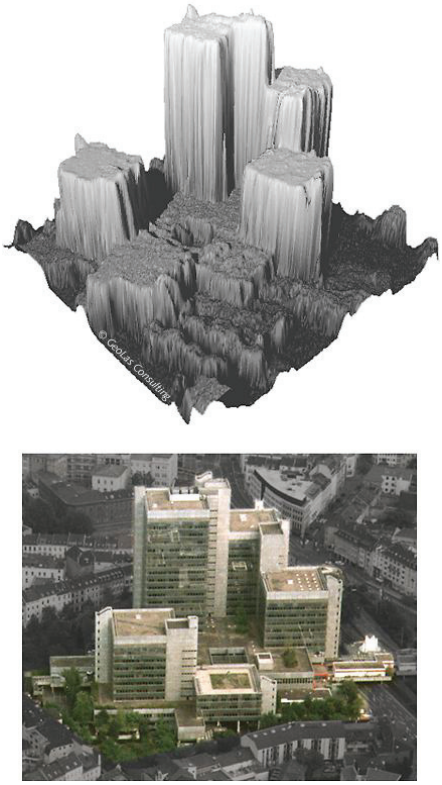


Figure 6.13: Topology image of a cluster of buidings

The high vertical accuracy available from a digital surface model allows simulations of flooding to be conducted. The following image shows the effect of the Rhine's flooding on the city of Bonn.

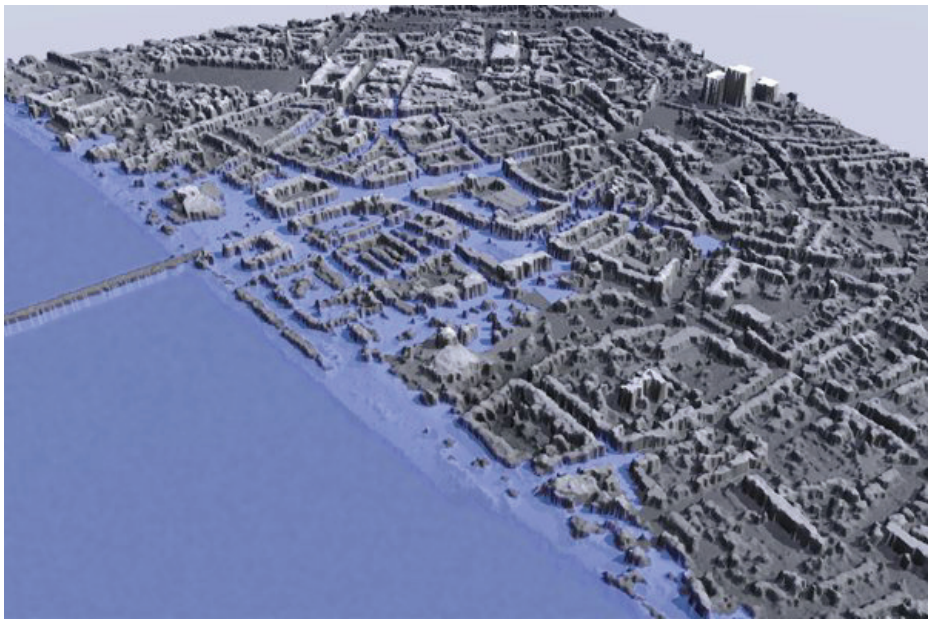


Figure 6.14: Application of topological images for flood analysis

6.4. Laser Based Sea-Bottom Profiling

The imaging principles developed for ground mapping can be extended to include sub-sea mapping as shown in the image generated using the Laser Airborne Depth Sounder (LADS) system

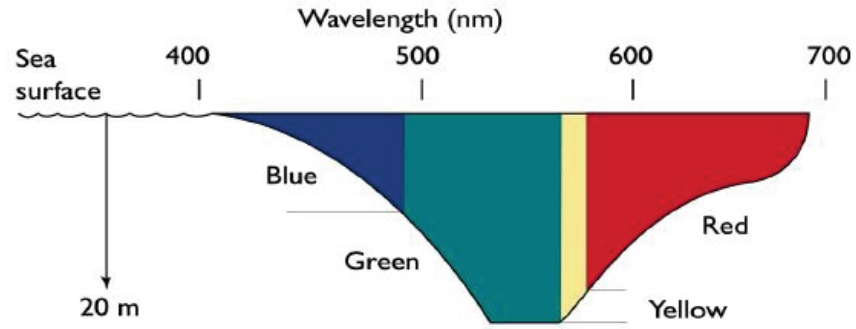


Figure 6.15: Inshore water penetration by light of different colours

Differential GPS and INS are used together to pinpoint the position of the aircraft while a laser altimeter measures the height of the aircraft above the sea and a scanned blue-green laser with good water penetration characteristics measures the height of the aircraft above the sea bottom across a swathe

The process can produce depth information at a rate of 64 square km per hour with a spatial resolution of 5m.

- Sounding rate 900Hz
- Water depth between 0.5 and 70m
- Swath width 240m
- Resolution normally 5×5m but with a 2×2m option
- Position accuracy 5m CEP 95%

Areas of cloud cover or high water turbidity that could compromise the performance of the LADS system are identified using satellite images provided by NOAA, and those areas are avoided during the survey.

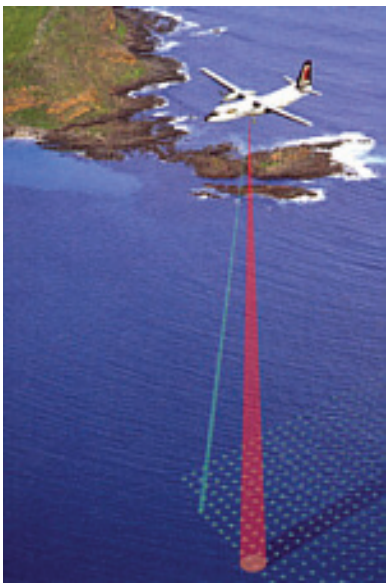
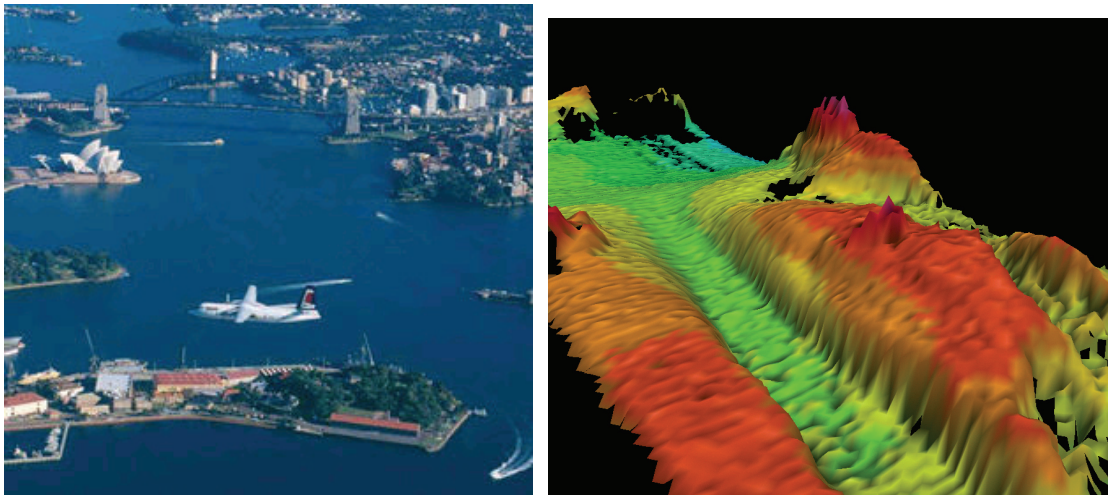


Figure 6.16: Laser based sea bottom profiling

The Royal Australian Navy's Laser Airborne Depth Sounder (LADS), mounted on a Fokker F27-500 performed a survey of Sydney Harbour in May 2002 as illustrated in the figure below.



(a) (b)

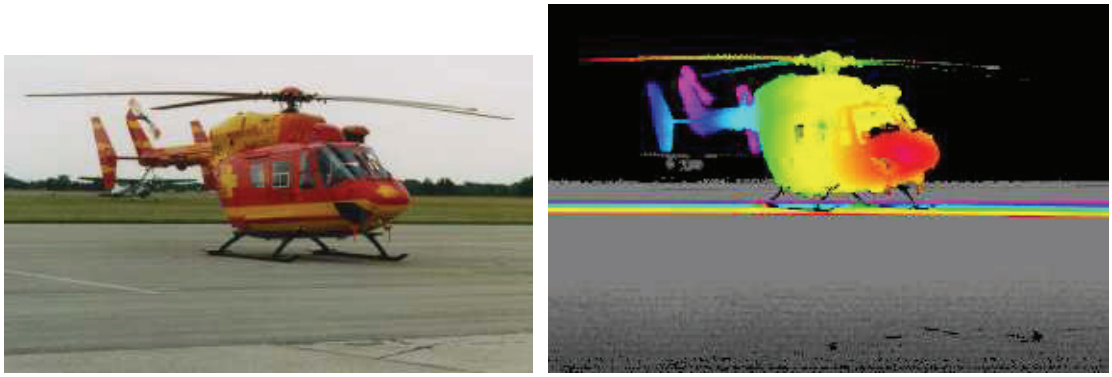
Figure 6.17: LADS Survey of Sydney Harbour showing (a) the aircraft and (b) the sow and pigs reef and the western channel

6.5. 2D Laser Scanners

Unlike the process shown above that relies on the forward motion of the vehicle to scan the one axis, 2D scanners involve scanning the laser range-finder using a raster to cover both elevation and azimuth angles as discussed.

With most scanners that operate over reasonably long ranges, the fast axis is usually scanned using a spinning mirror or prism, while the slow axis is scanned by rotating the whole scanner head.

This principle was used by Riegl to produce the range colour coded images shown below.



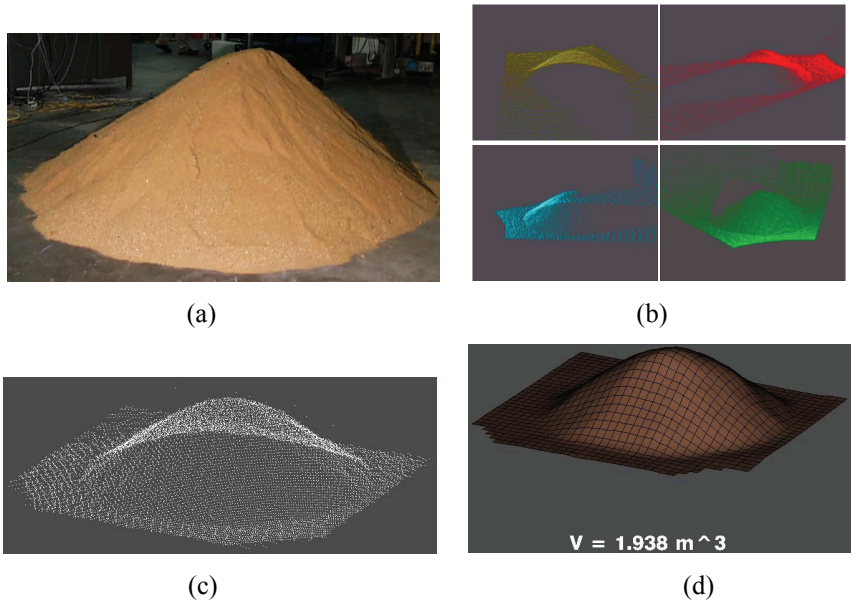
(a) (b)

Figure 6.18: 2D laser scanner colour coded images

6.6. 3D laser Scanners

The availability of low cost high resolution laser scanners has lead to a proliferation of industrial applications from volume estimation, architectural reproduction to CAD/CAM manufacturing.

For example, to determine its volume, the object is scanned from a number of carefully surveyed positions to produce individual views that are combined to generate a complete image point cloud image. Surface heights on a regular grid are computed from the point cloud data and the volume is calculated by integrating the height of the grid over the surface.



(a) (b)

(c) (d)

Figure 6.19: Process of estimating volume using a 3D laser scanner

In the CAD/CAM application, a component is scanned in a similar manner to produce a point cloud image which is then further processed to produce 3D digitised data of the surface. This is then used to reproduce the original component using a "rapid prototyping printer" or a CNC machine.



Figure 6.20: High resolution laser scanner produces a 3D model for computer aided manufacture

6.7. 3D Millimetre Wave Radar Scanners

Under adverse weather conditions or through dust or smoke, the performance of laser systems is degraded due to absorption and scattering (see Chapter 8). An alternative, albeit one with poorer angular resolution, is to use a millimetre wave radar to produce images.

Millimetre wave frequencies are used because their wavelength is sufficiently long compared to the sizes of dust, smoke or mist that the signal is hardly attenuated. Scanners based on standard pan-tilt technology (slow), or moving mirror (fast) have been developed at the ACFR.

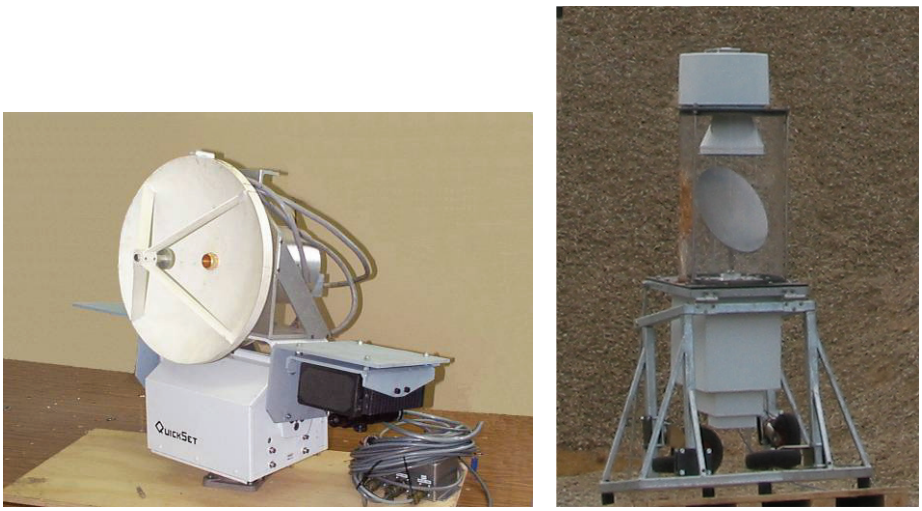


Figure 6.21: Millimetre wave radar imagers based on (a) pan and tilt and (b) mirror scanner

In addition to being relatively immune to weather conditions, because of the wide beamwidth, data produced by radar scanners can illuminate multiple targets simultaneously. This allows the unit to "see" through trees and to obtain foliage density information

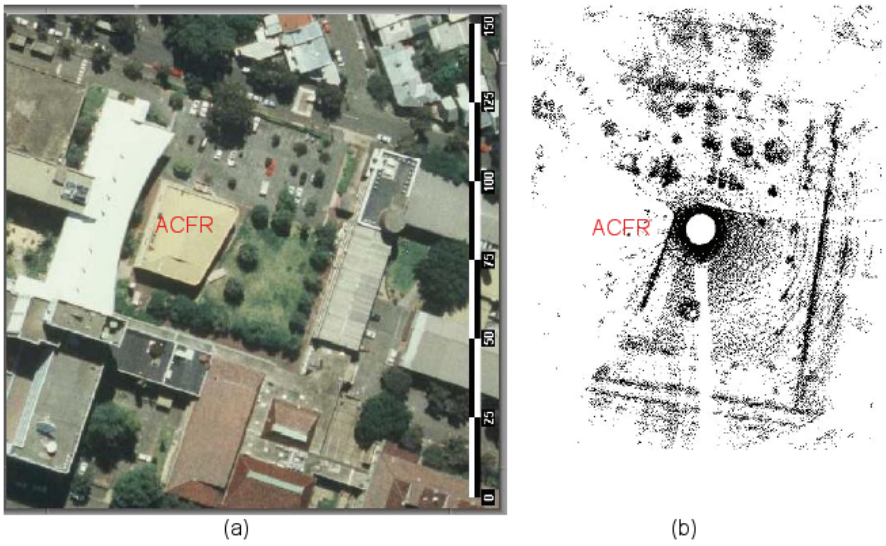


Figure 6.22: Radar image and photograph of the ACFR and surrounds - plan view



Figure 6.23: Radar image of the quad adjacent to the ACFR – 3D perspective view

6.8. Scanning Acoustic Microscopes

The Scanning Acoustic Microscope (SAM) produces images by scanning a focussed beam of acoustic energy (sound) across a sample to measure its elastic properties

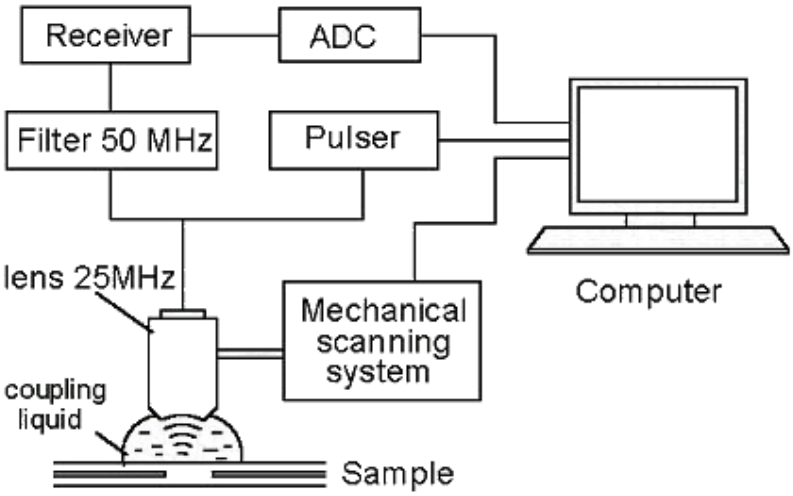


Figure 6.24: Schematic diagram showing the components of a scanning acoustic microscope

In principle this process is very similar to that used by the medical ultrasound scanning technology discussed in Chapter 12. The primary difference in this application is the use of mechanical scanning techniques while most medical systems use electronic scanning through one plane.

6.8.1. The Acoustic Lens

The principles involved are similar to those of an optical microscope, that is, different propagation velocities through different media are used to focus a beam so as to resolve small objects

Unlike its optical counterpart, where the relative velocities between the two media seldom exceeds 1.9 (the refractive index difference), with the acoustic microscope, the velocity can decrease by a factor of 10 or more when traversing a suitable solid-liquid interface.

A suitable lens such as the one manufactured for the original Stanford University microscope consisted of a concave spherical sapphire lens with a radius of approximately 40μm immersed in water.

At the interface, the sound waves are bent sharply inward as they encounter the liquid, to focus close to the centre of curvature of the sphere.

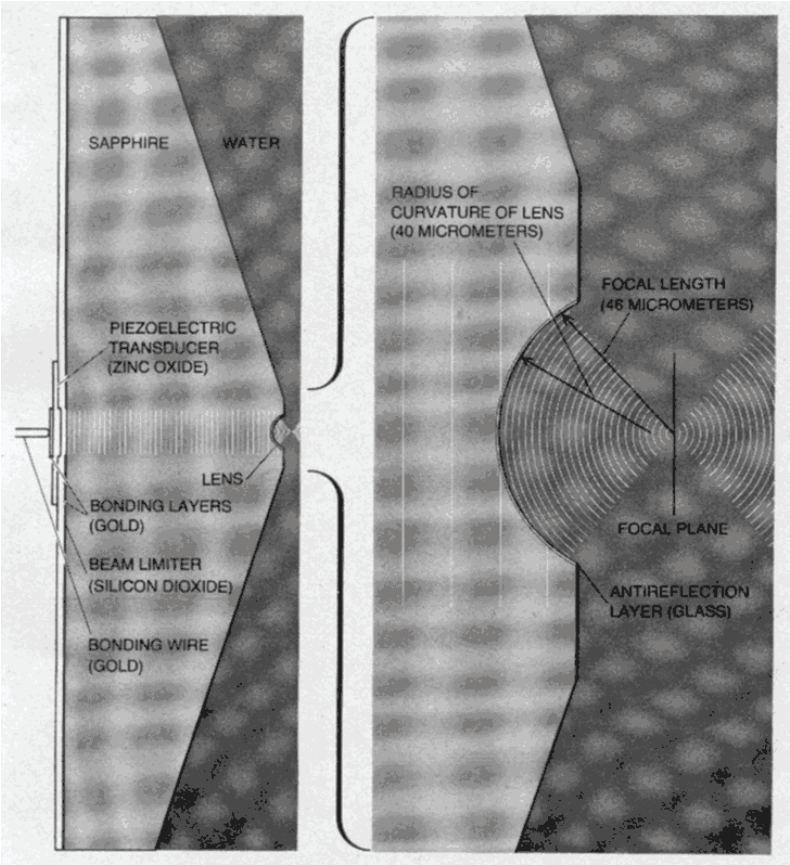


Figure 6.25: Schematic diagram of an acoustic lens

The first step in the imaging process is to convert the high frequency electrical signal to an acoustic one by means of a piezoelectric film deposited on the back of the

sapphire lens. This generates an acoustic wave that travels across the lens material until it strikes the back of the curved section where the focussing takes place.

The different elastic properties of the material distort the phase of the signal reflected from the focal point which then follow the reverse path and are detected by the same piezoelectric film which acts as a microphone.

Modern SAMs use short ultrasound pulses at a frequency ranging between 15 and 180MHz to separate the received and transmitted signals as well as time gating to image specific depths in the medium.

6.8.2. The Scanning Mechanism

An image is formed by scanning the focal point across the object in a raster fashion, storing the amplitude of the received signal along with the spatial co-ordinates and then displaying the complete picture. In the original microscopes as shown in the figure below this process could take a couple of seconds.

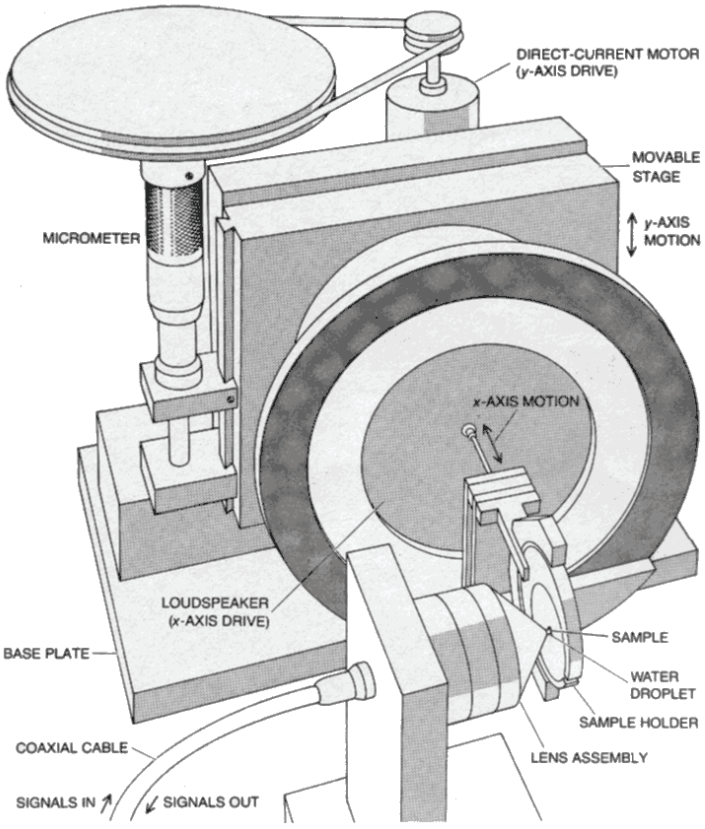


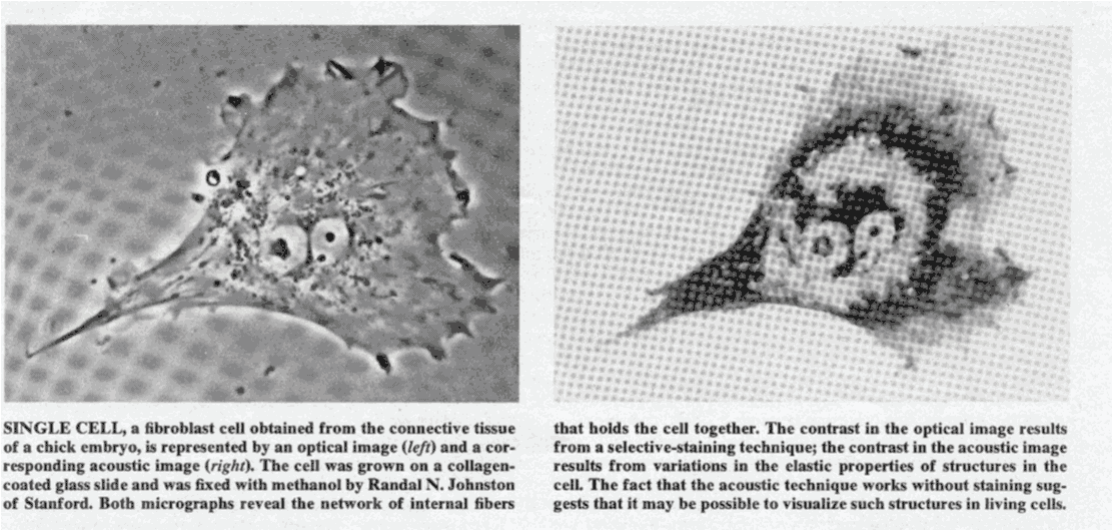
Figure 6.26: Schematic diagram of the scanning mechanism of an early SAM.

The original SAM, shown in the figure used a loudspeaker moving driven at 60Hz for the horizontal translation stage, and a DC servo motor driving a micrometer for vertical translation. An image was formed by synchronising this raster scan with a TV monitor the intensity of which was controlled by the amplitude of the reflected acoustic signal.

With more modern microscopes based on commercial scanning probe microscope (SPM) technology, the process is much faster.

6.8.3. Images

Because the process takes place in water which is compatible with living biological specimens, it is possible to make images of living cells. In the following example, however, the cell was fixed and stained so that a comparison could be made between the acoustic and optical images.



SINGLE CELL, a fibroblast cell obtained from the connective tissue of a chick embryo, is represented by an optical image (*left*) and a corresponding acoustic image (*right*). The cell was grown on a collagen-coated glass slide and was fixed with methanol by Randal N. Johnston of Stanford. Both micrographs reveal the network of internal fibers that holds the cell together. The contrast in the optical image results from a selective-staining technique; the contrast in the acoustic image results from variations in the elastic properties of structures in the cell. The fact that the acoustic technique works without staining suggests that it may be possible to visualize such structures in living cells.

Figure 6.27: Comparison of an acoustic image and an optical image of a single cell

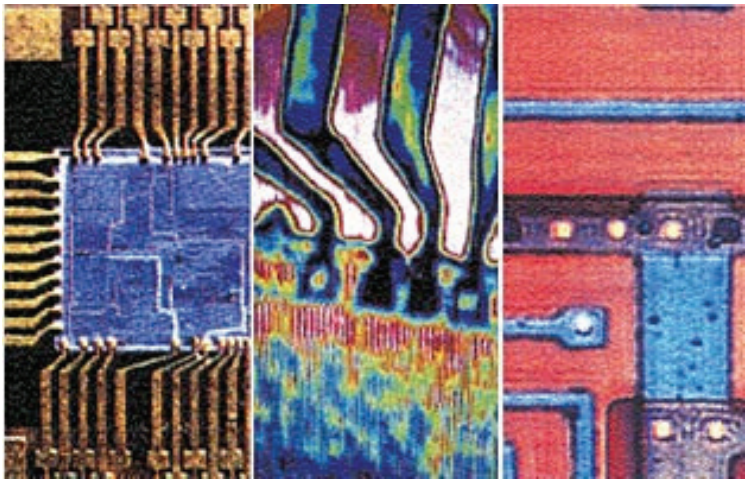


Figure 6.28: Acoustic images made of the interior of a potted IC. resolution increases from left

6.8.4. Near Field Microscopes

The scanning acoustic microscope has a limited resolution determined by the Rayleigh diffraction limit of the focussed beam.

To overcome this problem, the so-called near-field acoustic microscope was also developed using the fine mechanisms that have become available for (SPM's).

An acoustic wave is produced in the tiny area near the probe tip of the microscope (near-field area) and the interactions with the material are measured with a resolution that is not dependent on the wavelength.

6.8.5. Identification of Defects

During the propagation on the surface the surface acoustic wave interacts with the inhomogeneities like cracks, voids and topographic structures. Interference between the incident and reflected surface waves allows the detection of cracks that are an order of magnitude smaller than the spatial resolution of the microscope (approx. 1.5 µm at 1 GHz)

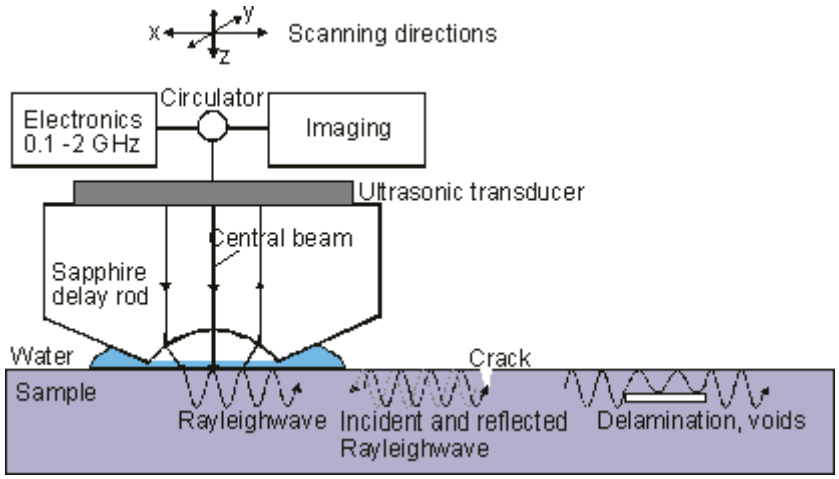


Figure 6.29: Mechanism to detect material defects using an acoustic microscope

6.9. References

[1] Imaging Laser Altimetry, http://www.geolas.com/Pages/laser.html, 28/02/2001.
[2] LADS, http://www.aims.gov.au/pages/facilities/remote_sensing/pages/rs_lads.html, 22/02/00.
[3] Depth Finder, *New Scientist,* 2 September 2000.
[4] S.Hovanessian, *Introduction to Sensor Systems,* Artech House, 1988.
[5] E. Brookner, *Radar Technology, Artech House, 1977.*
[6] M. Sinclair, *Laser Hydrography- Commercial Survey Operations*, Hydro 99,
[7] S. Scheding et. Al. *Terrain Imaging Millimetre Wave Radar.* ICARCV 2002, Singapore 2002.
[8] C.Quate, The Acoustic Microscope, Scientific American, October 1979
[9] Industrial Physicist "Seeing with Sound", www.aip.org/tip/INPHFA/vol-10/iss-3/p14.html
[10] Visualization and Volume Calculation of a Sand Pile by Laser Range Scanning (LIDAR), cic.nist.gov/lipman/sciviz/scan/
[11] Speed Up Your Production Projects and Save Money with 3D Laser Scanning, Digitzing and Reproduction Services by Cyber F/X ®, cyberfx.com/3d_laser_scanning.htm
[12] http://www.fiber-optics.info/articles/detector.htm
[13] Shallow Survey 2003, www.shallowsurvey.com/info_files/images3.html

Chapter 7.

Range Measurement Applications

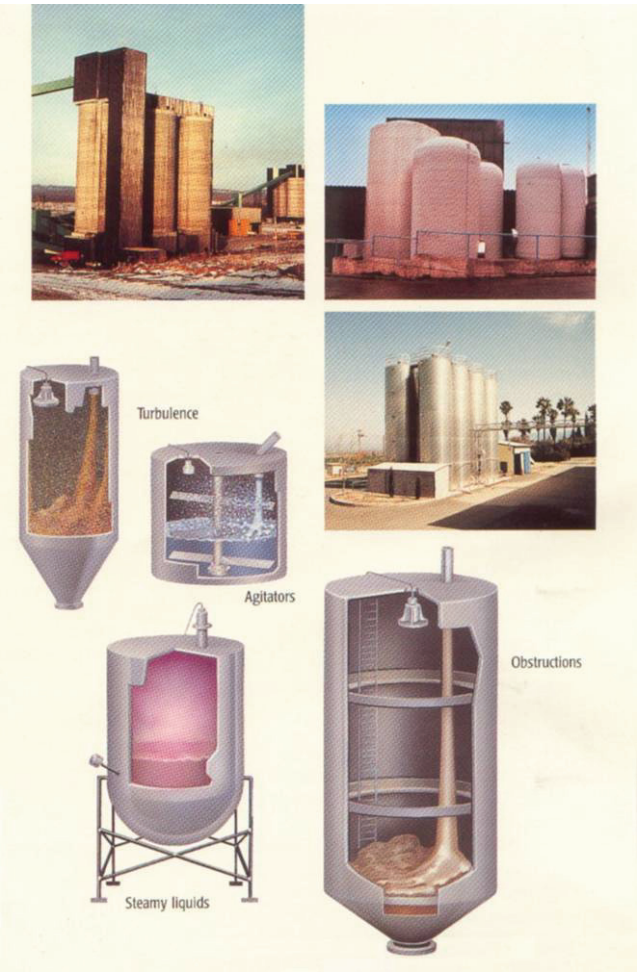


Figure 7.1: Industrial range measurement applications

Probably one of the greatest visions of the process industry has been a truly wire-and-retire non-contact, non-intrusive continuous level measurement instrument, a single technology that can be used in every application, a device that is self-calibrating and maintenance-free, that is easy to install onto any vessel with any process connection. At the same time this device should offer an accuracy to within 1mm, it must be low-cost and capable of paying for itself in under three months while able to operate in excess of 20 years. SA Instrumentation and Control, May 1998

7.1. Introduction

In the past non-intrusive measurement technologies struggled to cope with common industrial situations:

- Dust
- Fumes and vapours
- Air currents
- Angle of repose
- Foam
- Fixed vessel intrusions
- Agitator blades

7.2. Acoustic Level Measurement

This is also known as *ultrasonic* level measurement even when the frequency of operation is within the audible range. Operation depends on measuring the elapsed time between sending a sound pulse and receiving an echo and is probably the most widely accepted non-contact technology in use today

Applications range from levels in silos, flow in open channels, blocked chute detection to liquid level in tanks. However performance is limited by the presence of changing concentrations of fumes and vapours, pressure changes, vacuum, high temperatures, large temperature changes, excessive dust and foam on a liquid surface.

7.2.1. Propagation Velocity and Measurement Accuracy

Because the accuracy of ultrasonic technology relies on a knowledge of the speed of sound in the medium, every unforseen change in that speed affects the accuracy of the measurement.

In air at 20°C, the speed of sound is 344m/s and it changes by 0.17% for every 1°C change in temperature. Most measurement systems incorporate a temperature sensor that is used to compensate automatically for this variation.

The relationship between the molecular weight and the speed of sound is as follows

$$c = \sqrt{\frac{\gamma R(273+T)}{M}} \tag{7.1}$$

where: c- Velocity of sound (m/s)
R – Universal gas constant 8134.3 (J/Kmol)
T – Temperature (°C)
M – Molecular weight (kg/Kmol)
γ - Adiabatic exponent

The Molecular weight of the gas is calculated from its chemical formula and the atomic number of its constituent elements. For example Toluol (C_7H_8) has a molecular weight $M = 7\times12 + 8\times1 = 92$ kg/Kmol

The adiabatic exponent γ can be estimated as follows:

- 1.66 for monatomic gases (He, Ne, Ar)
- 1.40 for diatomic gases (H_2, O_2, N_2)
- 1.33 for triatomic and more complex gases (NH_3, CH_4, C_7H_8)
- 1.286 for very long molecules

For Toluol at 50°C, the speed of sound is

$$c = \sqrt{\frac{1.33 \times 8314.3 \times (273 + 50)}{92}} = 197 m/s .$$

The speed in mixtures of gases can be calculated using the molecular weight of the gas mixture.

Problems still arise if the medium is not homogeneous in which case ultrasonic technology is the incorrect choice for that application. For example a change from 0% to 100% relative humidity produces a speed change of 0.3% at 20°C, and a change in pressure of 30bar similarly produces a speed change of 0.3%.

7.2.2. Absorption

Absorption loss is a complex function of frequency and will be discussed in Chapter 9. As a rule of thumb a 3dB decrease in signal level occurs every 2m at 45kHz and only every 100m at 10kHz. To cater for this, long-range transducers have been developed that operate at frequencies as low as 5kHz.

Attenuation is greater in some gases than in others with CO_2 being particularly bad. Mixtures of gases will generally exhibit an attenuation that is proportional to their respective concentrations. Attenuation is also proportional to humidity, but this is generally solved for all but the most marginal cases by the selection of an appropriate transducer.

Attenuation by dust is dependent on its distribution and density. Light dust distributed evenly throughout a long measuring range may be much more detrimental than heavy dust confined to a small part of the range.

Decreases in pressure reduce the sound intensity and transducer performance due to mismatch losses and thus reduces performance, In contrast to this, with an increase in pressure, the increased mismatch losses are partially compensated for by the increased sound intensity. Hence most acoustic systems can tolerate increases in pressure better than they can tolerate a decrease.

7.2.3. Obstructions

Fixed obstructions such as support members can produce high-strength echoes that can cause some instruments to malfunction. However, most modern instruments allow false echoes to be identified and marked during commissioning.

Some rejection schemes depend on blanking segments of the span, while others form a time varying sensitivity profile with low sensitivity at the false echo regions. The

latter technique is more reliable as it does not hide the true target echo. However it is difficult to strike a balance between false readings and detection probability.

Modern instruments are generally capable of producing a database of echoes when the vessel is empty. This database is continuously updated and is used as a template to identify the true target.

7.2.4. Air Currents

Since the medium is the carrier of the acoustic wave, bulk movement of the medium will displace the acoustic wave. In open environments, air currents can cause the beam to be deflected, and an incorrect path length to be measured, while in confined environments, air currents are generally circulatory, and so will not cause sustained bending. If, however, the flow becomes turbulent, significant disruption of both the transmitted pulse and the echo can result in severe attenuation.

Doppler shift due to fluctuations in the air flow velocity can distort the echo phase and result in significant mismatch with the transducer resulting in reduced sensitivity.

7.2.5. Vibration

Low frequency vibration can cause shifts in the carrier frequency that result in reduced sensitivity.

Vibration frequencies close to the transducer resonant frequency can cause severe degradation of the signal quality if the vibration is transmitted to the sensing element of the transducer as it can mask echoes.

Vibration damping is generally employed to isolate acoustic transducers if they are mounted on moving structures.

7.2.6. Target Properties

All materials will partially reflect, partially absorb and partially transmit the incident acoustic pulse.

The proportion of energy reflected is a function the ratio of the characteristic impedance of the solid target to the “air”. Because this is related to the propagation velocity, hard dense targets tend to reflect well (as their propagation velocity is high), while soft light targets tend to transmit or absorb

$$Z = \rho.c\,, \tag{7.2}$$

where ρ - Material density (kg/m^3),
c- Speed of sound in the material (m/s).

The following list gives the acoustic impedance of a few common materials.

$Z_{air} = 400\ \Omega$
$Z_{water} = 1.4\times10^6\ \Omega$
$Z_{glass} = 13.1\times10^6\ \Omega$

While material properties are important at a microscopic level, the acoustic pulse interacts with a relatively large area of the target, so the strength and quality of an echo from the target will depend on its geometry.

With regard to geometry, there are two characteristics that are important:

- Small scale granularity,
- Large scale angle of repose and undulation.

Granularity

Granular particles scatter the reflected wave in all directions which is essential for an echo return if the material is lying at an angle to the normal. If, however, the particle size is comparable to $\lambda/4$, then significant cancellations can occur.

As a rule of thumb, the acoustic wavelength should be chosen to exceed the grain size by a factor of four

Angle of Repose and Undulations

If the material surface lies at an angle to the incident acoustic wave, the echo will be reflected away from the transducer towards the walls of the vessel. This can result in the echo return following a zig-zag path and an incorrect range reading.

In general, however, surface granularity effects with solids ensure that sufficient energy is scattered back in the direction of the transducer to obtain an accurate reading. For targets with steep angles of repose, the width of the beam that strikes the target can include will cover a wide range of distances, and so it is difficult to decide on the correct one. In this instance, it is important to understand the target material, and to use the highest possible frequency to minimise the beamwidth and hence spot size on the target.

7.2.7. Transducer Effects

Most systems use a single piezoelectric transducer to perform the transmit and receive function as the cost of the transducer represents a significant portion of the system price.

Modern systems apply a high voltage (>100V) sinusoidal signal to generate the transmit pulse. This allows precise control of the pulse and improved efficiency.

Transmitter frequency selection follows the following basic principles:

- Liquids &simple solids 30kHz
- Agitated liquids & dust free solids 20kHz
- Steam, foaming liquid, dusty solid 10kHz
- Steam, foaming liquid, powders etc 5kHz

Amplitudes of received pulses vary between about one volt down to fractions of micro-volts depending on the target range and losses.

The received signal is amplified, demodulated (detected) and filtered to produce an envelope which is further processed to identify a target echo.

The emitted pulse envelope is generally rectangular, however, it takes a finite time for the transducer to stop oscillating. This is known as the ring-down time.

During this time, the high amplitude oscillations would mask any echoes, so there is a period after transmission during which no target can be reliably detected. This is known as the blanking distance and is typically between 1 and 10ms (0.17 to 1.7m range)

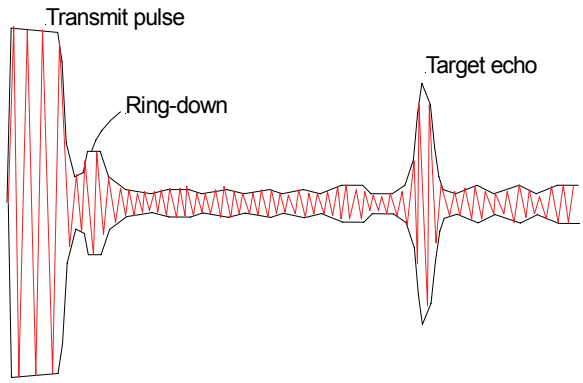


Figure 7.2: Salient features of an acoustic pulse

7.2.8. Transducer Mounting and Placement

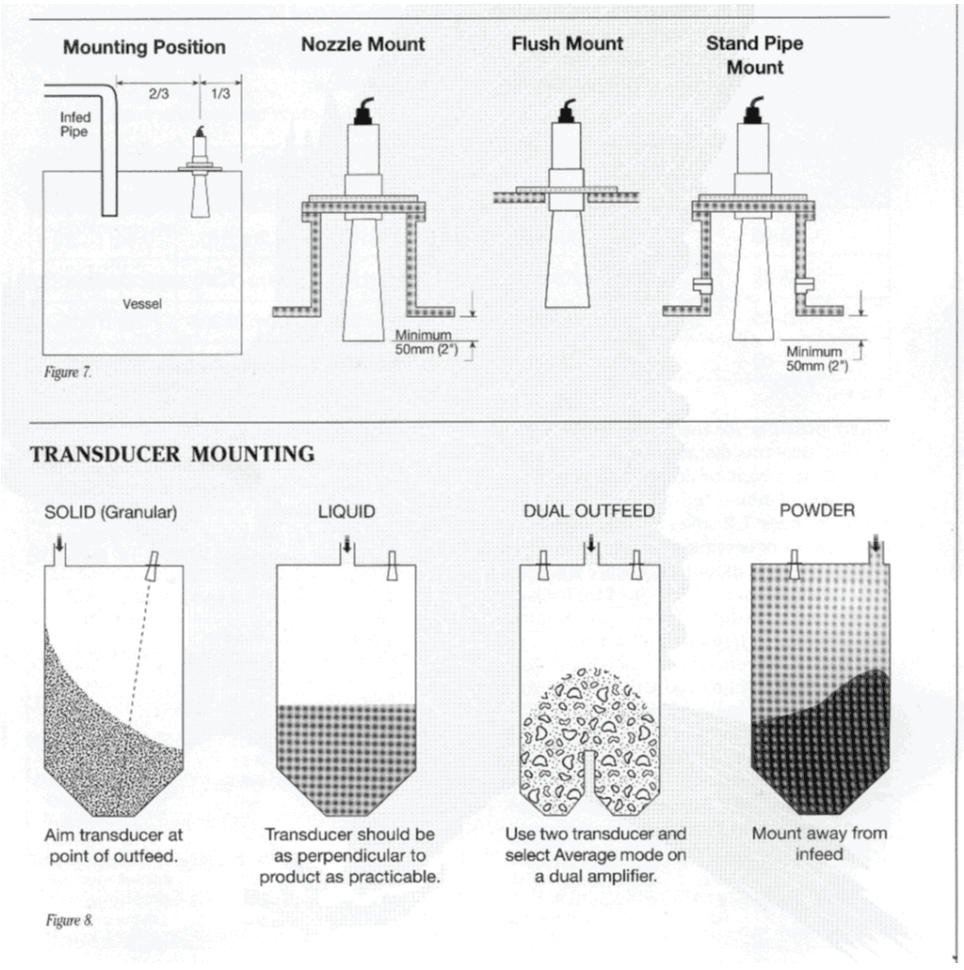


Figure 7.3: Transducer mounting configurations

7.3. Acoustic Systems

The following section includes a few of the major producers of acoustic measurement systems and their specifications where applicable.

7.3.1. Hawk Range Master System Specifications



Figure 7.4: Hawk acoustic measurement system

Amplifier and Transducer selections are made according to the maximum operational range required from the unit

Table 7.1: Hawk amplifier and transducer selection

Maximum Range (m)	Transducer Model	Amplifier Model & Operating Range	Blanking Distance (m)
10	TD-30	RMA 10 (0-15m)	0.3m
20	TD-20	RMA 20 (0-30m)	0.4m
75	TD-10	RMA 100 (0-100m)	1.0m
125	TD-05	RMHA 125 (0-125m)	1.2m

System accuracy is 0.2% of full range

System resolution if 0.1% of full range

7.3.2. Milltronics AiRanger System Specifications



Figure 7.5: Milltronics acoustic measurement system

Amplifier and Transducer selections are made according to the maximum operational range required from the unit

Table 7.2: AiRanger specifications

Maximum Range (m)	Transducer Model	Frequency (kHz)	Beamwidth (deg)	Blanking Distance (m)
7.5	ST-25 C	44	12	0.3m
15	ST-50	44	5	0.3m
30	LR-21	21	5	0.9m
60	LR-13	13	5.5	1.2m

7.3.3. Vega Vegason System Specifications



Figure 7.6: Vega acoustic measurement system

Vegason-50 Range 0.25 to 15m
Vegason-70 Range up to 30m
Vegason-80 Range up to 60m

7.4. Short Range Radar Level Measurement

For short range level measurement (R<30m), microwave radar sensors are very common as they will operate through pressure windows (typically vacuum to 64bar) into tanks.

Applications involving high pressure and temperature usually involve measuring liquid levels and not solids or slurries, so very few instruments are designed to measure the latter.

7.4.1. Propagation Velocity and Measurement Accuracy

It can be assumed that the propagation velocity is both known and constant.

$$v = \frac{c}{\sqrt{\varepsilon}} \text{ m/s,} \tag{7.3}$$

where: v – Velocity (m/s),
c – Speed of light 2.997925×10^{8} (m/s),
ε - Relative dielectric constant.

In contrast to ultrasonics, errors caused by changes in propagation velocity due to variations in temperature, pressure or medium are almost non-existent for radars.

Table 7.3: Vapour content effect on velocity

Vapour Content	Temperature (°C)	Dielectric Constant	Velocity (10^8 m/s)	Error at 30m (m)
Air	0	1.000590	2.997925	0
Helium	140	1.000068	2.997823	+0.00104
Hydrogen	100	1.000264	2.997529	+0.00403
Oxygen	100	1.000523	2.997141	+0.00797
Nitrogen	100	1.000580	2.997055	+0.00885
Ammonia	0	1.007200	2.9871904	+0.10953
Benzene	400	1.002800	2.993736	+0.04265
Carbon Dioxide	100	1.000985	2.996449	+0.01501
Water	100	1.007850	2.986226	+0.11940

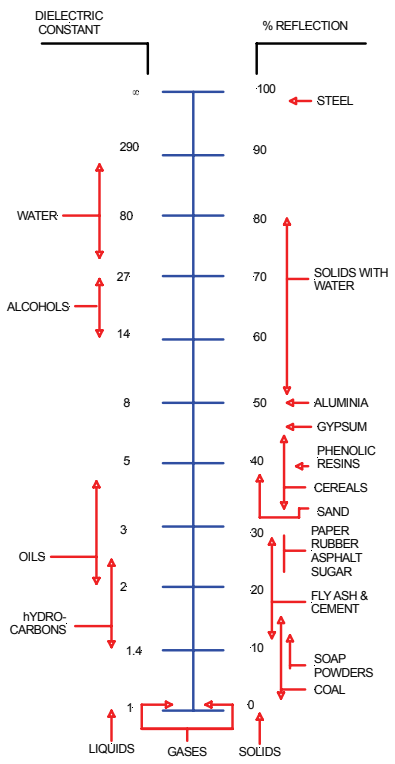
7.4.2. Absorption

The absorption of electromagnetic radiation by the gaseous medium is very small and can be ignored for most industrial applications.

Particles suspended in the medium such as water droplets or dust can however have a significant effect depending on their size (compared to the wavelength) and their dielectric and conductivity properties. This will be examined in a later lecture

Absorption effects are proportional to frequency, and become particularly severe as the wavelength approaches the size of the suspended particle. This is generally only a problem for laser systems.

7.4.3. Target Properties



- The radar reflectivity characteristic is inversely related to its relative dielectric constant
- Reduced reflection from low dielectric materials allows the radar to penetrate foam layers above liquids. It also allows the tracking of water levels in tanks containing hydrocarbons
- As with acoustics, for solid targets, the particle size and angle of repose will have an effect on the echo strength, so most of the discussion in the section above is applicable here.
- Liquid level radars often rely on the fact that only one smooth high reflectivity target will be visible to measure ranges to sub millimetre accuracy. This is useful in custody transfer applications (petrol & oil).
- Pulsed radars are good for high dielectric constant materials ε_r >8

Figure 7.7: Dielectric effects at X-band

7.4.4. Transducer Effects

Unlike the acoustic devices, radar units do not use a common transducer for the transmitter and receiver, though they generally use a common antenna.

Most existing short range sensors operate at 5.8 or 10GHz. However, the next generation of radar systems at 24GHz offer the advantages of smaller size and narrower beamwidth.

The transmitter is generally based on a solid state oscillator (FET or HEMT) with the whole circuit (transmitter and receiver) built on microstrip line. Some low-cost modules still use iris coupled cavity based Gunn oscillators and diode mixers.

For long range applications (>100m), the frequency of choice will be even higher; at 35, 77 or 94GHz as a narrow beamwidth becomes even more important. In this case the circuitry is still brass block and waveguide, though MMIC technology is starting to appear at 77GHz.

Horns are the most common antennas and are mounted within the pressure vessel beyond a "transparent" pressure window in the throat. The use of inert dielectric rod (PTFE) antennas in clean industries such as dairy is also quite common, and parabolic reflector antennas are available from some manufacturers for specialist applications.

Major manufacturers of short-range time-of-flight radar equipment with moderate accuracy include Milltronics, Endress+Hauser and Vega, while SAAB, Enraf and Krohne make high accuracy frequency modulated continuous wave (FMCW) radar units for custody transfer applications.

7.4.5. Milltronics IQ Radar Specifications



- Operates at 5.8GHz (USA 6.3GHz) and transmits a 1.5ns pulse every 2us.
- It will take reliable measurements of liquids and slurries with ε_r > 3 at ranges from 1 to 15m.
- Temperature –40 to +200°C, Pressure 1-16bar
- Accuracy +/-0.3% of range
- Repeatability +/-10mm
- Time transformation is used

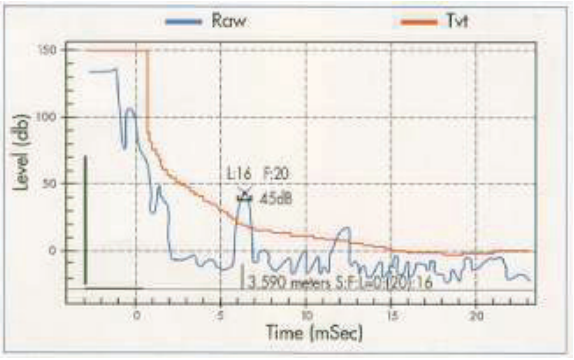


Figure 7.8: Milltronics radar

7.4.6. Endress+Hauser Micropilot FMR-130 Radar Specifications



- Pulsed time of flight principle operating at 5.8GHz.
- Effective radiated power (ERP) 1μW avearge.
- Maximum range 18m (with a 33m option)
- Typical accuracy +/-5mm
- Repeatability +/-3mm
- Processing speed 44 samples per second.
- Beamwidth rod antenna 23°
- Beamwidth horn antenna 45°

Figure 7.9: Endress radars

7.4.7. Vega Vegapuls Radar Specifications



- Pulsed time of flight principle
- Operational frequency 5.8GHz
- Pulse width 1ns
- Pulse repetition frequency 3.6MHz
- Accuracy <0.1%
- Uses time transformation processing

Figure 7.10: Vega radar

7.4.8. SAAB TankRadar PRO Radar Specifications



- Frequency Modulated Continuous Wave (FMCW) principle
- Centre frequency 10GHz
- Swept bandwidth 1GHz
- Self calibrating 6 times per second with internal delay line reference
- Range 0 to 50m
- Accuracy +/-5mm
- MIP mode measures phase shift as the surface changes to improve accuracy to +/-0.1mm

Figure 7.11: SAAB radar

7.4.9. Krohne BM70A Radar Specifications



- FMCW perinciple
- Operates at 9GHz
- Swept band 8.5 to 9.9GHz
- Linearity correction using oscillator reference. Correction to 98%
- Accuracy not specified (BM70 specified as <0.5% of measured value)
- Range 0.5 to 40m (options up to 100m)
- Repeatability < 0.5 error of measurement
- Resolution 1mm
- Permittivity ε_r >=1.5
- Pressure up to 64bar (option 400 bar)

Figure 7.12: Krohne radar

7.4.10. Other radars

Apex

- Operates using the FMCW principle
- Centre frequency 25GHz
- Sweep band 2GHz (24-26GHz)
- Accuracy +/-5mm over range 0.5 to 10m
- Accuracy +/-0.05% over range 10 to 30m
- Repeatability +/-1mm
- Resolution +/-0.4mm
- Beamwidth 22.9°, 13.7° and 10.5° for different horn antennas
- Full vacuum to 10bar

Enraf Smart radar

- Based on a combination of pulsed and phase shift methods
- Operational frequency 10GHz
- Synthesised pulse (phase shift at different frequencies to obtain superior results). Accuracy <+/-1mm
- Designed for Tank farm operations

Trolex

- Range to 20m
- Resolution 1mm

TN-Technologies RCM

- FMCW mode of operation
- Range 0.3 to 34m
- Accuracy +/-3mm
- Repeatability +/-3mm

7.5. Long Range Radar Level Measurement

For longer range operation (100 to 400m) in dusty or humid environments, millimetre wave radar offers the only viable option for two reasons.

- Dust and vapour penetration is superior to laser or ultrasonic devices
- The beamwidth is sufficiently narrow to avoid illuminating the walls and so superior to microwave-radar devices

Most of the characteristics of millimetre wave radar are similar to those of microwave radar, and so will not be repeated here.

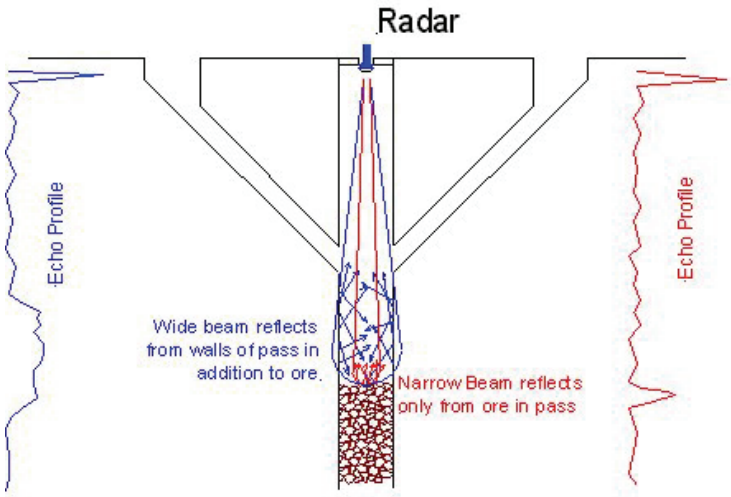


Figure 13: Beamwidth effects on echo shape

The Dusty Ranger is a W-Band (94GHz) radar developed by us at AMS in South Africa primarily to measure range in dusty orepasses and silos

The photograph gives an indication of the dust level in a typical orepass

Note the dust that has accumulated on the radar in less than a week



Figure 7.13: Dust accumulation on a radar

Table 7.4: Specifications of the orepass radars

Short Range Version	Long Range Version
Pulsed FM principle	FMCW principle
Range 7 – 120m	Range 5 – 350m
Frequency 94GHz	Frequency 94GHz
Transmit power 10mW	Transmit power 10mW
Pulswidth 30ns	Swept bandwidth 150MHz
Antenna beamwidth 1.5°	Antenna beamwidth 0.75°
Resolution 4.5m	Resolution 1m
Accuracy +/-1m	Accuracy +/-1m

A new radar has been developed at the ACFR which replaces both the short and long range units with a single FMCW radar that can be configured for either requirement.

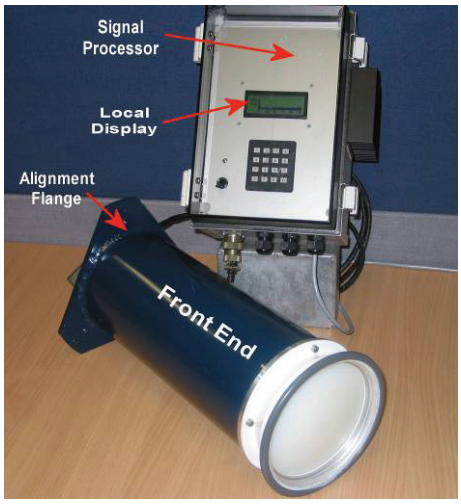


Figure 7.14: Orepass radar developed at the ACFR

7.5.1. Other Long Range Radar Developments

A low frequency radar was developed for LKAB (Sweden) that utilised the waveguide characteristics of a narrow pass to propagate the EM wave more than 400m. However this technique required that the radar frequency be tuned for every pass.

The University of Cape Town in South Africa developed an X-Band (10GHz) orepass radar that was unsuccessful because of clutter returns from the sides of the pass.

A Russian company ELVA-1 also has a 94GHz radar on the market. It operates using the FMCW principle and has specifications very similar to the ACFR unit.

7.6. Laser Level Measurement

Using low-cost mature technology, laser range finders provide the most cost-effective method to measure long range in benign environments.

Because the operating wavelength is about 1μm, even a small aperture (50mm) can produce a beam with a divergence of <0.1°, this allows for high angular resolution and long range measurements to be made with low effective radiated power (ERP).

High-speed modulation of laser diodes is possible so good range resolutions can also be achieved using short pulses and the split-gate discriminator discussed in Chapter 5.

7.6.1. Propagation velocity and measurement accuracy

Propagation velocity will be similar to that of the lower frequency EM sensors. However, as laser systems are generally used in air, it can be assumed that the velocity will be a constant 2.997925×10^8 m/s

Measurement accuracy is a function of the sensor electronics rather than the environment

7.6.2. Absorption

The maximum range achievable with a laser range finder depends strongly on the visibility.

Range performance is generally specified for clear air (20km visibility), while at lower visibility, the maximum range is reduced due to atmospheric attenuation. This is shown for Riegl lasers in the graph.

Absorption is a function of both the material type and the size of particles (this is dealt with in more detail in Chapter 8).

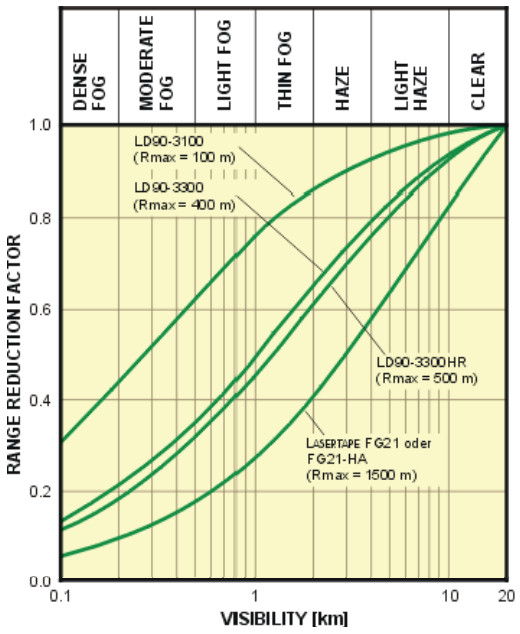


Figure 7.15: Effect of mist and fog on laser radar detection range

These visibility curves are calculated for water, however, as a first approximation they can be used for suspended dust particles if the particle diameters are similar. As a rule of thumb, the performance of IR lasers is similar to sensors operating in the visible region – If you can see a target, the laser can probably measure its range.

7.6.3. Target properties

The amount of light that is returned from a target's surface is characterised by its reflection coefficient and its surface properties.

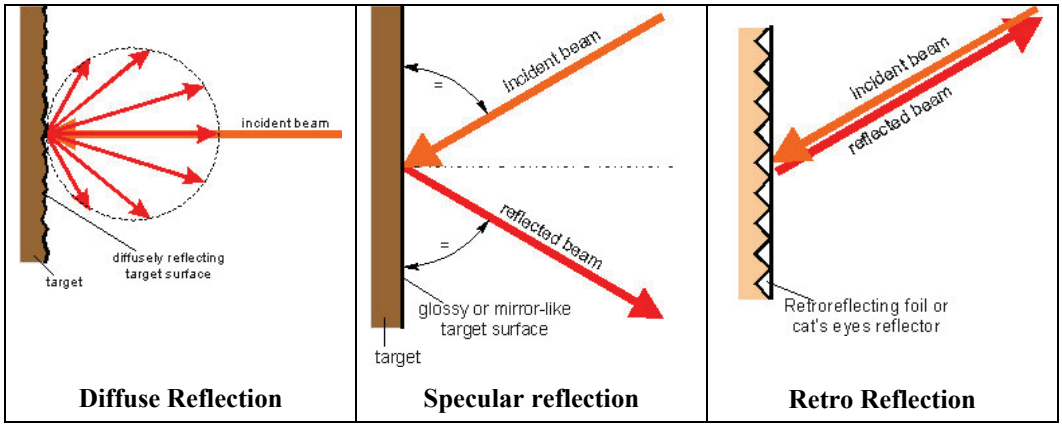


Figure 7.16: Target reflective characteristics

The reflection coefficient is a function of frequency, so the tables reproduced later for microwave and 10µm infrared will not be the same as those for 0.9µm infrared shown in the table.

For a diffuse scatterer, the reflection coefficient cannot exceed 100%, but for a specular scatterer, the reflection coefficient can be many times this value.

Table 7.5: Reflectivity values for various materials

Diffusely Reflecting Material	**Reflectivity (%)**
White paper	Up to 100
Cut clean dry pine	94
Snow	80-90
Beer foam	88
White masonry	85
Limestone, clay	Up to 75
Newspaper with print	69
Tissue paper 2-ply	60
Deciduous trees	Typ 60
Coniferous trees	Typ 30
Carbonate sand (dry)	57
Carbonate sand (wet)	41
Beach sand and bare desert	Typ 50
Rough wood pallet (clean)	25
Smooth concrete	24
Asphalt with pebbles	17
Lava	8
Black neoprene	5
Black rubber tyre wall	2
Specular Reflecting Material	
Reflecting foil 3M2000X	1250
Opaque white plastic[1]	110
Opaque black plastic[1]	17
Clear plastic[1]	50

1 Measured with the beam perpendicular to the surface to achieve maximum reflection

The operational range of a laser sensor is generally specified for a target with 80% diffuse reflectivity. For other reflectivities it can be determined using the graph below. The mechanisms that cause this attenuation are considered in more detail in Chapter 8.

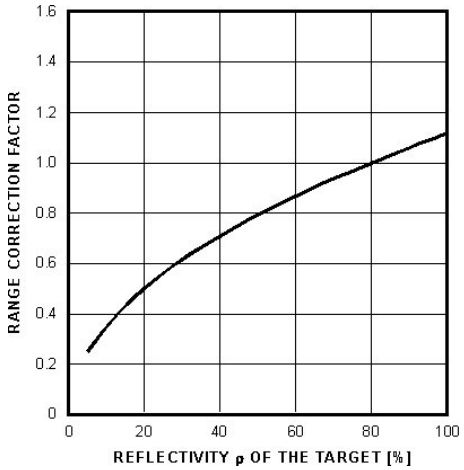
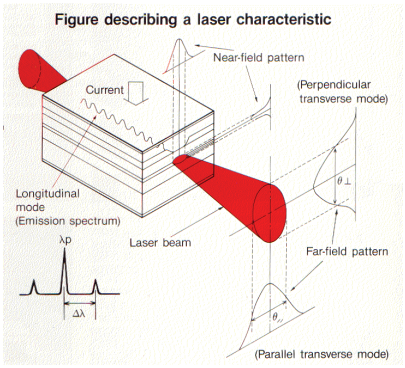
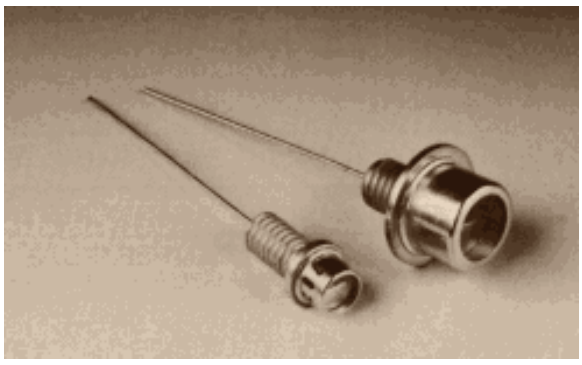


Figure 7.17: Effect of target reflectivity on laser radar range

7.6.4. Transducer effects

Most low cost laser range measurement devices operate using the pulsed time of flight principle. A low power (≈ 2mW) pulsed laser diode operating in the infrared (≈ 1μm) transmits a short pulse (≈ 10-20ns) through a collimating lens towards the target. The light is scattered by the target and a small portion is reflected back towards the sensor.





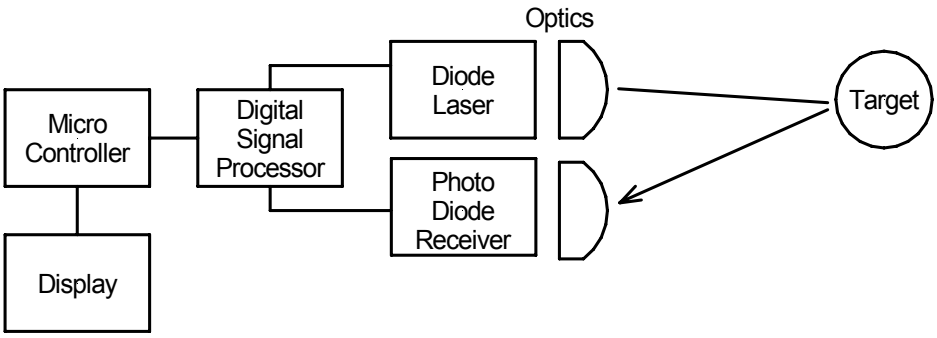


Figure 7.18: Schematic diagram of a laser radar

Generally, to limit receiver saturation (or even damage) a separate receive aperture focuses the reflected radiation onto a narrow band fast light sensitive diode (PIN diode or avalanche photodiode).

7.6.5. Last Pulse Processing

Under conditions of poor visibility, partial reflections may be received from a number of false targets before the true target range is reached. To cater for this eventuality, Riegl has introduced a processing scheme that allows the user to select the last or next to last return.

Targets can only be distinguished in range if they are separated by between 2 and 5m (depending on the echo size).

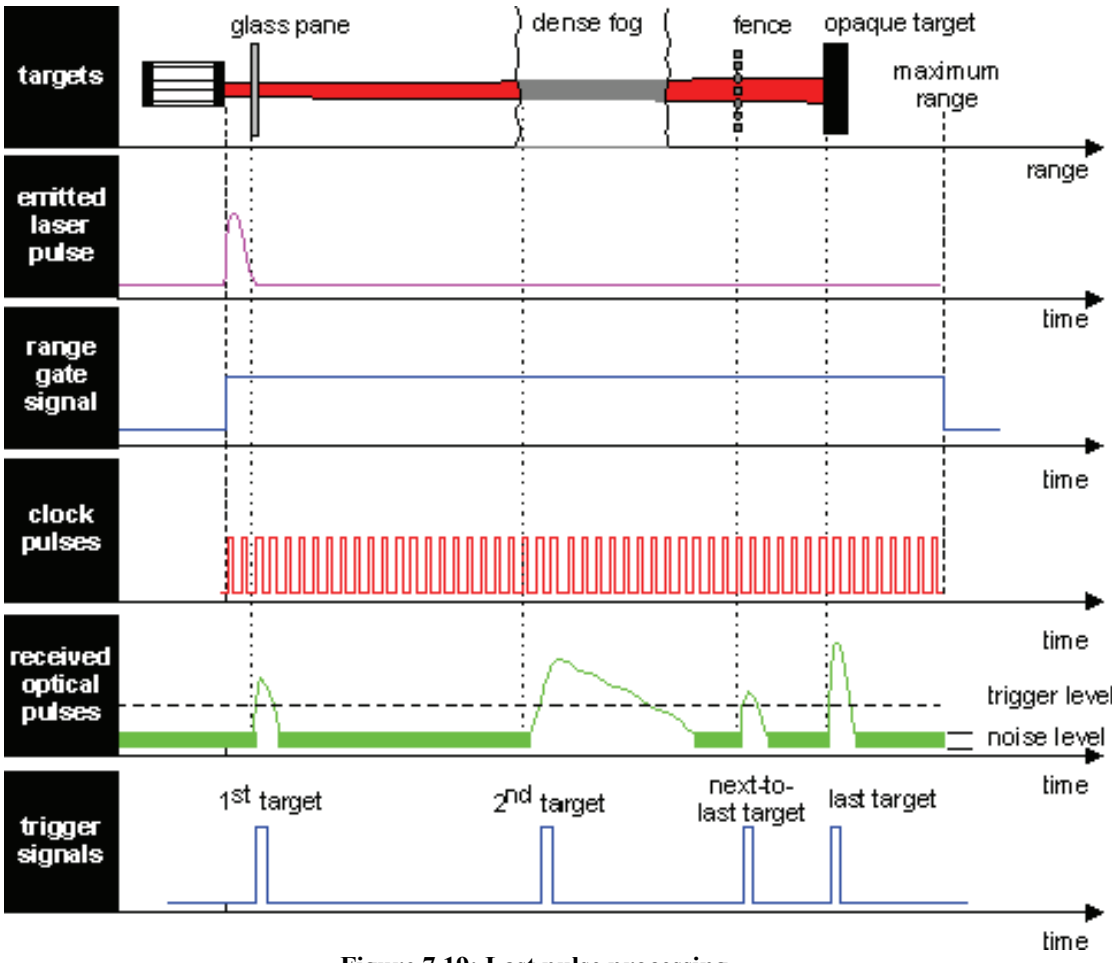


Figure 7.19: Last pulse processing

7.7. Industrial Laser Ranging Systems

There are many manufacturers of laser based industrial measurement sensors, so two manufacturers (Riegl and Laser-M) were chosen as being representative of the range of devices available

7.7.1. Riegl LD90 Industrial Distance Sensor



- Pulsed time of flight
- Range 150m (ρ>80%), 50m ρ>10%), 1000m (retro reflector)
- Accuracy +/-25mm
- Repeatability +/-50mm (175ms int time)
- Repeatability +/-10mm (2s integ time)
- Output resolution (quantisation) 5mm
- Divergence 2mrad (0.1°)

Figure 7.20: Riegl LD90

7.7.2. Riegl FG21 Laser Tape



- Pulsed time of flight
- Range: Masonry 2km, Trees 1.5km, Retro reflectors 3km
- Wavelength 0.9μm
- Accuracy +/-1m
- Resolution 1m
- Beam divergence 2mrad (20cm per 100m)
- Acquisition time 0.5s typ

Figure 7.21: Riegl FG21

7.7.3. Laser-M LM4-LR-120 Industrial Distance Sensor



- Pulsed time of flight
- Wavelength Infrared with visible alignment pointer
- Range 10-120m
- Resolution 0.6m (0.5% of max)
- Update up to 12 per sec
- Available in low and high power versions

Figure 7.22: Laser-M LR-120

7.8. Recreational Laser Ranging Systems

In the past few years, a number of low-cost laser range finders have become available for the recreational market (mostly golf and hunting). These systems are all based on pulsed time of flight techniques, and offer remarkable performance.

Table 7.6: Recreational laser range finder specifications

Manufacturer	NewCon Optiks [7]	NewCon Optiks	Nikon [8]	Nikon	Bushnell [9]	Bushnell	Bushnell
Model	LRM 1200	LRM 1500	Laser 600	Laser 400	Yardage Pro Legend	Yardage Pro Trophy	Yardage Pro Sport
Price (US$)	249	349	280	229	326	195	159
Targeting Device	scope	scope	scope	scope	scope	scope	scope
Minimum range	20m	20m	11 yards	11 yards	15 yards	5 yards	5 yards
Range to white target	1140m	1460m	600 yards	400 yards	930 yards	800 yards	800 yards
Output resolution	1m	1m	1 yard	1 yard	1 yard	1 yard	1 yard
Display	LCD	LCD	LCD	LCD	LCD	LCD	LCD
Data link	none	none	none	none	none	none	none
Power source	9V battery	9V battery	CR2 Lithium 3V	CR2 Lithium 3V	CR2 Lithium 3V	9V battery	9V battery
Size (mm)	122×120×60	122×120×60	96×72×36	96×72×36	100×80×45	102×84×46	100×79×43
Mass (g)	420	420	210	195	204	215	210
Additional features	10 reading recall. Target quality indicator	10 reading recall. Target quality indicator. Target speed	none	none	Continuous scan	Continuous scan	none



Figure 7.23: Bushnell Yardage Pro Sport and the image taken through the viewfinder

7.9. Selection of the Correct Sensor

The following should be considered when making a decision with regard to which sensor would be suitable for a particular industrial application.

Measurement Accuracy

- Rough, when to fill or empty only
- Accurate, volume or depth at any time

Conditions at Vessel

- Internal construction and obstructions
- Diameter
- Depth
- Wall material
- Heating coils
- Indoor or outdoor location
- Vibration
- Number of filling/emptying orifices

Measurement Medium

- Temperature
- Pressure
- Composition
- Steam or vapour
- Foam
- Fumes
- Dust

Target Characteristics

- Suspended solids
- Interfaces (water/oil)
- Corrosiveness
- Reflectivity
- Dielectric constant
- Conductivity
- Particle size
- Angle of repose

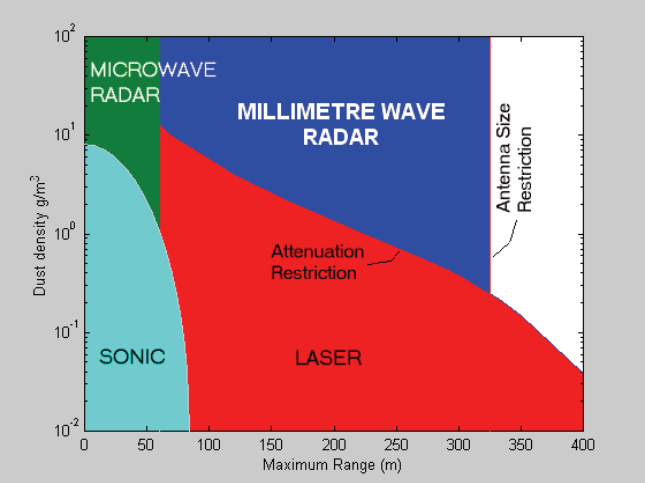


Figure 7.24: Cost effective sensor selection for an orepass

7.10.Short Range Sensors

7.10.1. The Polaroid/SensComp Ultrasonic Sensor

Probably the most common of the ultrasonic sensors used for robotic applications is the Polaroid 6500 ranging module and an appropriate transducer.

The “ping” is generated by supplying the electrostatic transducer with 16 low-high-low transitions between +200 and –200V at about 50kHz.

Under normal conditions the receiver is blanked for a short period (2.38ms) to reduce the possibility of false alarm. This defines the minimum range of operation.

The reflected signal excites the transducer which must have a resonance at about 50kHz, and it generates a small voltage which is fed into a stepped-gain amplifier.

The gain of the amplifier is increased exponentially to compensate for the $1/R^2$ propagation loss up to a maximum range of 10m

Threshold detection is used to detect an echo. This is output as a digital bit and the time of flight is determined by measuring the time from the initiation of the ping to the received echo.



Figure 7.25: Polaroid/SensComp ultrasound sensor

The current consumed by this sensor is quite low (<100mA) except when it is transmitting during which time the current drawn rises to 2A. This induces large transients on the DC power line that can cause problems.

The Polaroid 6500 ranging module can use a number of different transducers a selection of which are shown in this picture (Series 9000, Instrument Grade and Series 7000)



The Instrument Grade unit is the most accurate and offers the narrowest beamwidth. It operates at about 50kHz.

The Series 7000 has a slightly wider beam which can be useful for unscanned applications while the Series 9000 offers an oval beam pattern and is designed to withstand harsh environments where it may be exposed to water, salt etc. it operates at a frequency of 45kHz.

Navigation Application of Polaroid Sonar

Most indoor robots use ultrasound sensors as one of their localisation sensors because they are low cost, have a reasonable operational range and a good range resolution. Their main drawback is a wide beamwidth which results in poor angular resolution.



Figure 7.26: Various indoor robots showing the arrays of Polaroid sensors

From a navigation perspective, this poor angular resolution has a major impact on the performance of these sensors. Because many indoor walls, and other structures, are smooth in relation to the wavelength of the ultrasound they exhibit specular behaviour (see Chapter 8). This means that strong returns only occur if the beam is orthogonal to the surface or it is aiming into a corner.

Early researchers tried to construct line segments from which the internal structure of the space could be reconstructed as shown in the following figure, but because of the wide beam pattern and the specular behaviour this was not particularly successful in matching to external plans.

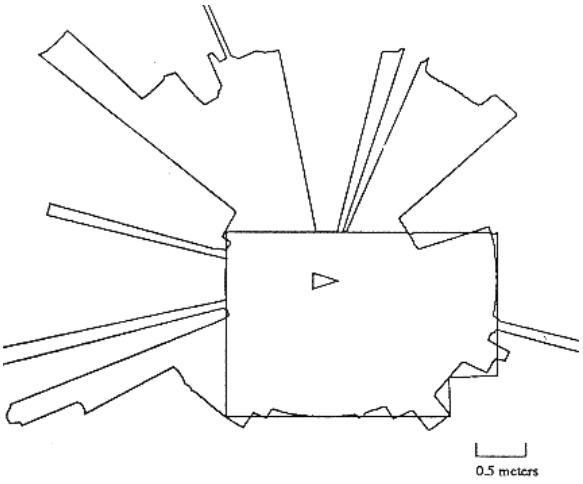


Figure 7.27: Scanned ultrasound image of a room

More robust methods of using sonar data include occupancy grids in which the sonar returns are used to confirm the occupancy of individual grid elements in a dense 2D array. Unfortunately, because of the relatively slow speed of sound, building up such grids is very time consuming.

Since the advent of high speed scanned LIDAR, the use of ultrasound has been relegated to low cost or niche applications.

7.10.2. The Micropower Impulse Radar

The microwave equivalent of Polaroid ultrasonic sensor is the Micropower Impulse Radar (MIR) which was developed by the Lawrence Livermore laboratory in 1993.

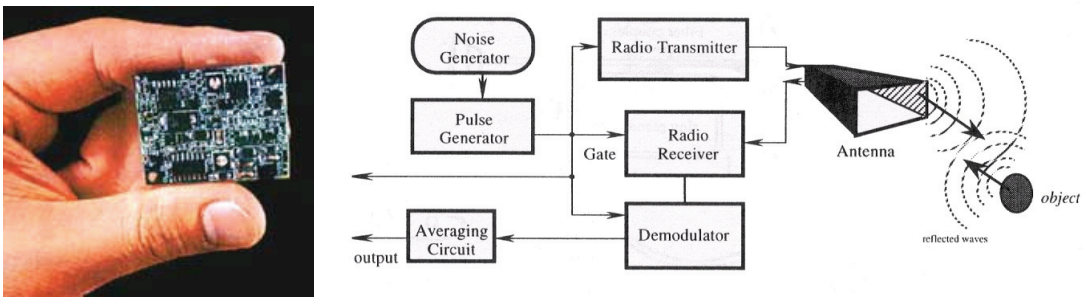


Figure 7.28: Micropower impulse radar module and schematic block diagram

A pseudo random noise generator generates randomly spaced pulses at an average PRF of 2MHz +/-20% with a Gaussian distribution. The interval between pulses can range from 200 to 625ns. The pulses have a constant width τ which on-off modulates a transmitter centred at either 1.95 or 6.5GHz.

Because the pulse width, τ, is very short, the approximate bandwidth of the radiated signal is very wide, about 500MHz at a centre frequency of 1.95GHz as shown in the figure below.

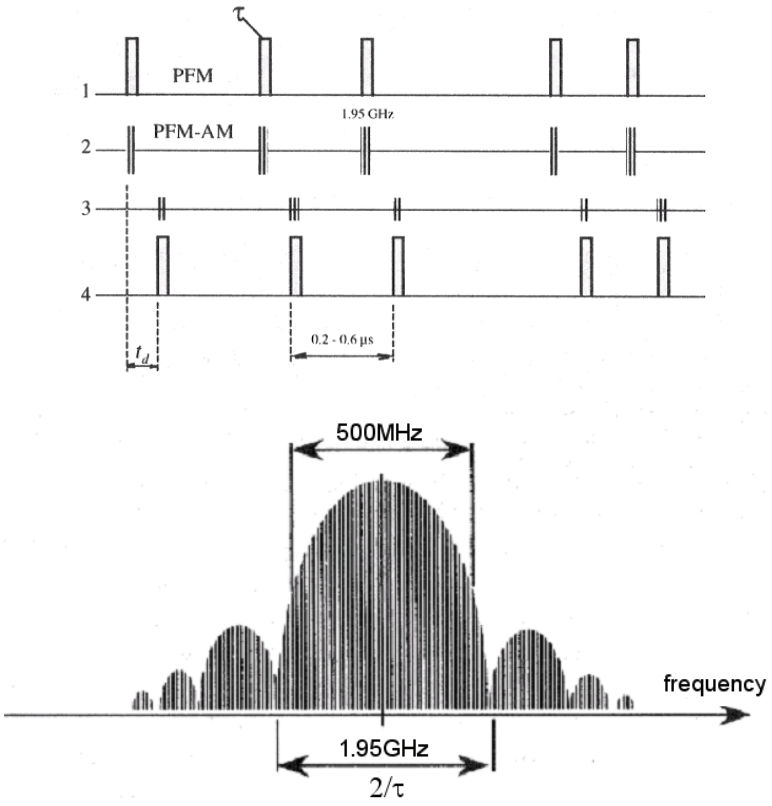


Figure 7.29: Micropower impulse radar timing diagram and spectrum

The same pulse generator that generates the transmit pulses is used to gate the receiver after a predetermined delay t_d. Only echoes received during that particular time window are detected. Because the average duty cycle of the transmitted pulses is <1%, and since the modulation spacing is random, any number of identical MIR sensors can be operated in close proximity without significant interference.

Integration of some 10000 received pulses is conducted prior to detection and ranging, so even if some interference is experienced it is unlikely to compromise the performance of the radar.

The low duty cycle of the radar ensures that the power consumption is very low (50μW) with the result that two AA batteries should power it for a number of years. In addition the effective radiated power (measured using a broadband bolometer) has been found to be about 1μW which is more than 1000 times lower than the international safety standard of 1mW/cm^2 for continuous whole body exposure.

Because of the wide bandwidth and low frequency, the MIR signals will penetrate the human body and so can be used to monitor both heart and arterial movement. Non contact respiration monitoring is another application. Because the sensitive area can be gated, the system would be ideal as a monitor for individual patients in ICU, a terrorist behind a wall or as a cot alarm to monitor babies who might be susceptible to sudden infant death syndrome (SIDS). The following figure shows the experimental results of body detection through a wall.

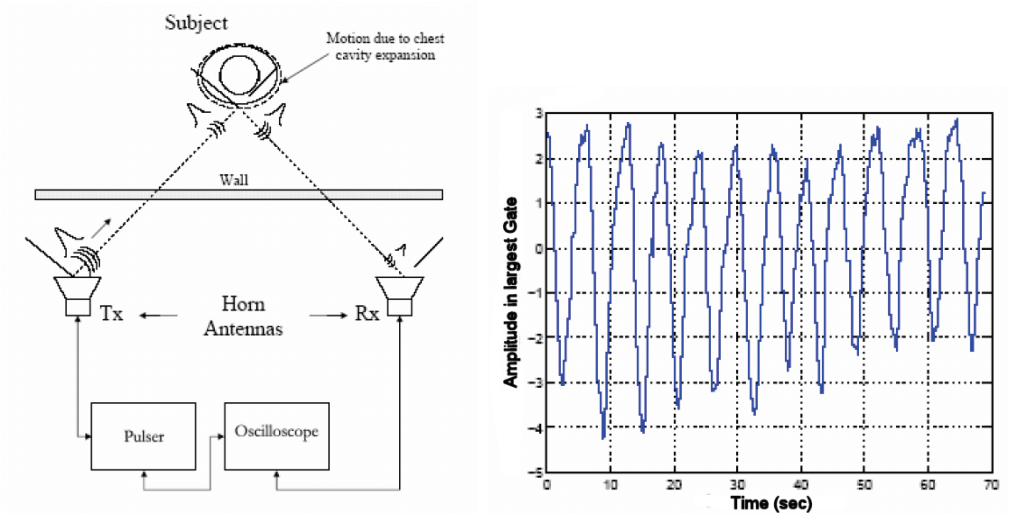


Figure 7.30: Using MIR to detect movement through a wall

Ground Penetrating Application

(The HERMES (High-Performance Electromagnetic Roadway Mapping and Evaluation System) Bridge Inspector is a radar-based sensing system mounted in a trailer.

HERMES uses 64 MIR modules mounted underneath a trailer pulled by a vehicle at traffic speeds. The sensors, assembled into an array about 2m wide, are spaced about 30mm apart. They send out UWB pulses with frequencies ranging from 1 to 5 gigahertz, penetrating concrete to a depth of up to 300mm. As the pulses propagate

through the bridge deck, the echoes are recorded by a computer inside the trailer and compiled into a three-dimensional map of the deck.

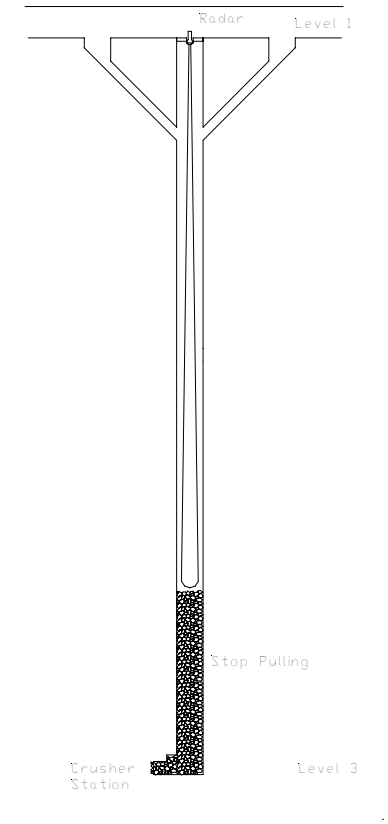


Figure 7.31: (a) HERMES trailer, (b) Interior showing the array of 64 modules and (c) an image showing where potential delamination may have occurred

Other Applications

Other applications include range meters, intrusion alarms, level detectors, automation, robotics, human speech analysis, weapons and novelty products.

7.11.Orepass Radar Development: Case Study



7.11.1. Requirement

- To measure the range from 10m to the bottom of a 300m deep 6m diameter ore pass so that an estimate (accurate to 1%) can be made of the amount of ore available.
- A typical pass configuration is shown in the diagram
- The pass will be filled with loose rock, which may be dry or wet, and there will be lots of dust.
- A grizzly (coarse grid) at the top of the pass ensures that rocks do not exceed 1m in diameter.
- The radar should be capable of operating while rock is being tipped into the pass
- The range measurement update rate should be sufficiently high to monitor the progress of the rock as it falls down the pass
- Blasting takes place within 50m of the radar and the concussion wave that travels through the development is intense.

Figure 7.32: Orepass schematic diagram

7.11.2. Selection of a Sensor

This was discussed previously. Dust attenuation makes the laser option unworkable and the long range eliminates ultrasonic techniques. Radar is the only viable option.

7.11.3. Range Resolution

The rock surface will not be regular, large rock diameters and the angle of repose of the rock surface will result in reflections occurring over at least 1.5m in range.

To obtain a measurement accuracy of 1% over a 300m deep pass requires a resolution of 3m or better.

We select a range resolution of 2m, which is quite well matched to the target size (to maximise the radar cross section) and is also less than the required measurement accuracy.

To obtain a range resolution of 2m, the transmitted pulse width τ and the range gate size ΔR must both be 2m.

7.11.4. Target Characteristics

The pile of rock may be wet or dry. It can be shown that the radar cross section, σ, is a function of the relative dielectric constant ε_r:

$$\sigma = k\left|\frac{\varepsilon_r - 1}{\varepsilon_r + 2}\right|^2 \tag{7.4}$$

For the rock the $\varepsilon_r = 2.25$ and for water it is 80[1]. The ratio of the RCS for wet and dry rock targets is $\sigma_{water}/\sigma_{rock} = 0.9282/0.0865 = 10.7$ (10.3dB).

The pile of rock can be described as a number of facets of various sizes and facing in different directions. Scattering from the various facets may add constructively or destructively and thus a large variation in the reflectivity (cross section per unit area) can be expected.

Without going into details regarding scattering from rough surfaces, we can glean from the literature that the mean reflectivity σ^o will be about –10dB, when the rock is dry.

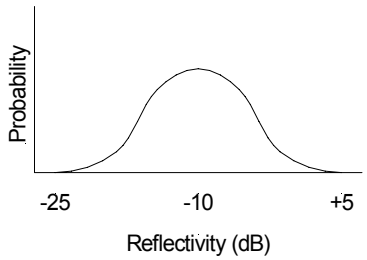


Figure 7.33: Rock reflectivity distribution

Because we can expect both deep fades and large specular returns, we will assume a log-normal distribution with the tails extending 15dB on either side of the mean as shown in the figure above.

7.11.5. Clutter Characteristics

The walls of the pass are made of the same material as the target; they are also very rough so we can assume the same variation in reflectivity.

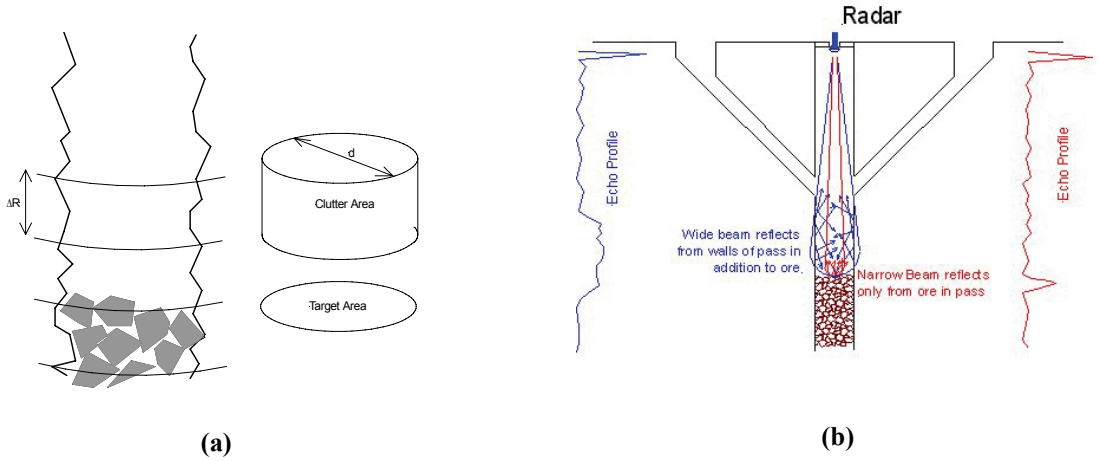
Because the grazing angle is much lower, we can assume a slight reduction in the mean reflectivity to –15dB.

7.11.6. Target Signal to Clutter Ratio (SCR)

For adequate detection probability, the target to clutter ratio requirements can be determined in a similar manner as the signal to noise ratio requirements. We assume that at least 13dB is required for adequate P_d and P_{fa}.

[1] This is not true at 94GHz where the dielectric constant of water is much lower

The maximum mean target cross section is the product of the mean reflectivity and the beam footprint $\sigma = \sigma^o A$. This occurs when the beam fills the pass.



(a) **(b)**

Figure 7.34: Diagrams showing (a) target and clutter areas and (b) beamwidth effect on echo

To simplify the calculations we convert everything to dB. The target area in dB is just $10\log_{10}(A)=10\log_{10}(\pi d^2/4) = 14.5\text{dBm}^2$.

The mean target RCS, $\sigma_{tar} = 14.5\text{-}10 = 4.5\text{dBm}^2$.

The clutter area within the same gate as the target echo is a cylinder of the pass with diameter d and height equal to the gate size ΔR.

The clutter area is $10\log_{10}(\pi.d.\Delta R) = 10\log10(37.7) = 15.8\text{dBm}^2$.

The mean clutter RCS, $\sigma_{clut} = 15.8\text{-}15 = 0.8\text{dBm}^2$

The target to clutter (SCR) ratio is 4.5-0.8 = 3.7dB, which is much too low for a good probability of detection. It is not possible to use integration to improve the effective SCR because the target returns are correlated in the same way as the signal returns.

The logical alternative is to ensure that the beamwidth is sufficiently narrow that no reflections are returned from the walls of the pass.

7.11.7. Antenna Beamwidth

At a range of 300m antenna footprint must not exceed 6m

$\theta_{3dB} = 6/300 = 0.02$ rad (1.15°). For a slight safety margin, make the beamwidth 1°.

7.11.8. Antenna Size and Radar Frequency

The beamwidth in degrees and the antenna diameter (for a circular aperture) are related by the following empirical formula:

$$d = \frac{70\lambda}{\theta_{3dB}} \tag{7.5}$$

If we consider the size of the antenna that will be required as a function of the operational frequency, we can select an appropriate frequency.

- The smaller the antenna the easier it is to mount and align the radar.
- Components costs are proportional to frequency
- Propagation losses increase proportional to frequency

Table 7.7: Antenna diameter as a function of operational frequency

f (GHz)	λ (m)	d (m)	Comment
10	0.03	2.1	Much too large
35	0.0086	0.6	Too large
77	0.0039	0.27	ok
94	0.0032	0.22	ok

It can be seen from the table that a frequency of 77 or 94GHz would be satisfactory.

7.11.9. Radar Configuration

The proposed radar configuration is shown below:

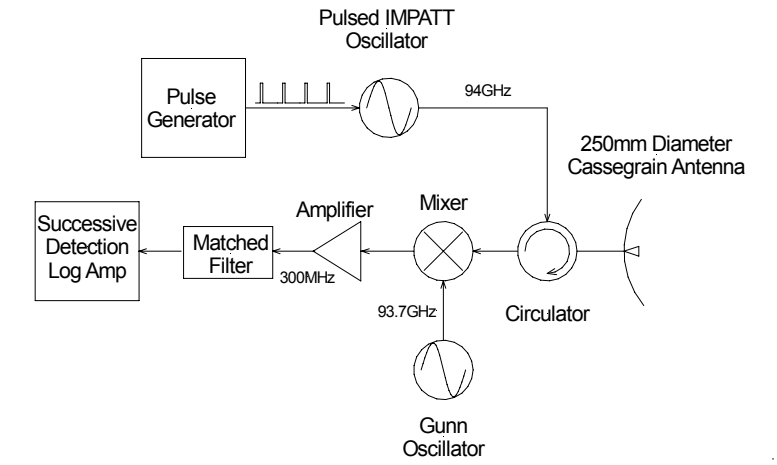
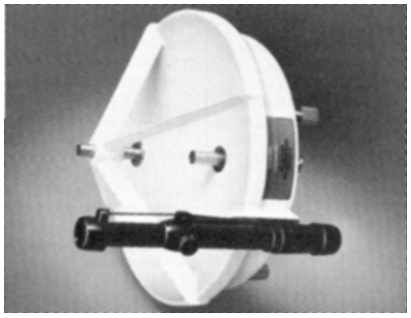


Figure 7.35: Pulsed radar schematic diagram

7.11.10. Component Selection

Antenna Options

Antennas are available with diameters of 200, 250 and 300mm. We select a 250mm diameter antenna for operation at 94GHz



Select a 250mm diameter Cassegrain antenna from Millitech or a 250mm horn lens from Flann Microwave.

At 94GHz the characteristics of the two antennas are similar

Gain = 46dB, $\theta_E = 0.8°$ $\phi_H = 0.9°$

Cassegrain antenna sidelobes will be marginally higher than those of the horn lens.

Figure 7.36: Cassegrain antenna

We can confirm these specifications by calculation. For an aperture efficiency ρ_A=0.7 (typical for a Cassegrain antenna)

$$G = \frac{4\pi\rho_A A}{\lambda^2} = 42432 \text{ (46.2dB)} \tag{7.5}$$

$$\theta = \phi = \frac{70\lambda}{d} = 0.89° \tag{7.6}$$

Radar Transmitter

Pulsed time of flight with an uncompressed pulse width of 2m

$$\tau = \frac{2\Delta R}{c} = 13.3\text{ns} \tag{7.7}$$

The lowest cost option will be a pulsed radar based on a non-coherent solid state Gunn or IMPATT diode based transmitter.

The off-the-shelf options from Millitech are as follows:

- Pulsed Gunn τ = 20ns to 1000μs with a maximum duty cycle of 50% and P_t = 0.1W (20dBm). Typical chirp 100MHz
- Pulsed IMPATT τ = 50ns or 100ns with a PRF between 10 and 75kHz and P_t = 12W (40.8dBm). Typical chirp 100MHz

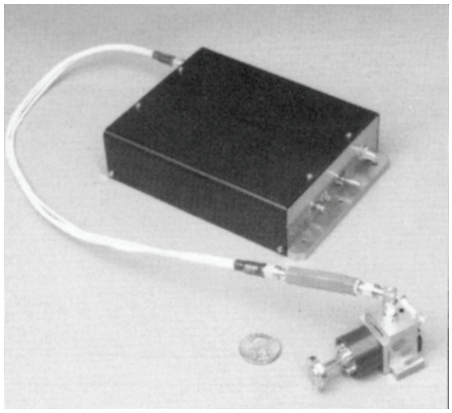


Figure 7.37: Pulsed IMPATT transmitter

Neither transmitter meets the 13.3ns pulse width requirement. However, we select the Gunn option as being the closest at 20ns (3m), which is still equal to the specified 1% without using interpolation methods to improve the measurement resolution.

Receiver Options

The receiver configuration could be one of the following:

- RF amp – Mixer – IF Amp – Matched Filter (*G* = 20dB DSB *NF* = 6dB)
- Mixer – IF Amp – Matched Filter (*L* = 8dB DSB *NF* = 7dB)

Amplifiers at 94GHz are still extremely expensive ($15k each), so the small noise figure advantage is not justified.

We will use the 2nd option

Local Oscillator

Not much choice. A mechanically tuned Gunn oscillator with an output power Pout = 40mW (16dBm) is adequate.

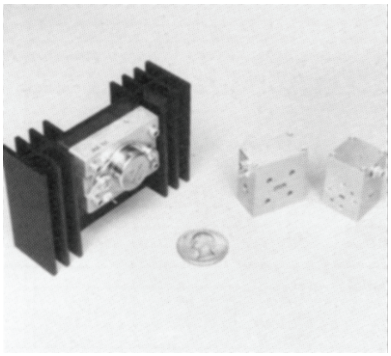


Figure 7.38: Gunn local oscillator

Duplexer

Options include the following:

- 3dB Directional Coupler, 20dB directivity, 1.6dB Tx insertion loss and 4.6dB Rx insertion loss
- Junction Circulator, 20dB isolation, 0.8dB insertion loss for both Tx and Rx paths.

From both insertion loss and isolation (directivity) the circulator is either superior or equal to the coupler. The coupler can handle higher powers, but the circulator is good to 5W peak that is fine for our application.

The circulator is also smaller and lighter than the coupler.

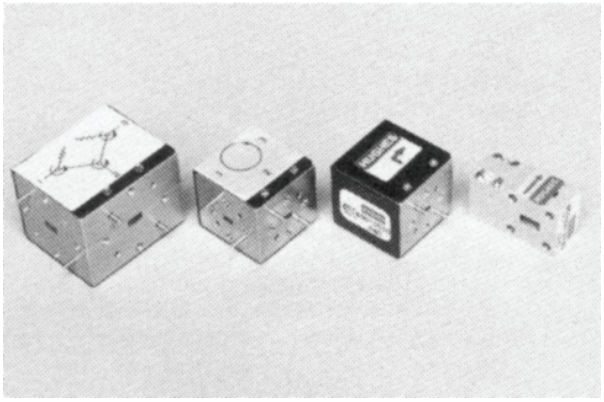


Figure 7.39: Circulators

Matched Filter

Assuming a rectangular transmit pulse and 2 cascaded single tuned stages, he optimum $\beta.\tau$ will be 0.613 with a loss in SNR of 0.56dB. For τ = 20ns, the optimum bandwidth β = 30.65MHz.

Because the transmitter chirps about 100MHz during the pulse period, using a filter with a bandwidth of only 30MHz would result in a significant loss of received power $10\log_{10}(30/100)$ = 5dB.

It is very difficult to make a matched filter for the uncontrolled transmitter chirp as it is extremely non-linear and is a function of a number of factors that are difficult to control.

We will use a compromise filter with a bandwidth of 50MHz that will have a loss of about 3dB compared to a matched filter.

The IF Frequency

The IF frequency is selected according to the following:

- Amplifier components easy to obtain and low cost
- The matched filter with a bandwidth of 50MHz is easy to construct
- Detectors are available at that frequency

A typical amplifier would have the following specifications

- Band 200-400MHz
- Gain 30dB
- Noise Figure 1.5dB

The Transmit and Local Oscillator Frequencies

For the selected IF centre frequency of 300MHz, the transmitter is tuned to operate at 94GHz and the LO at 93.7GHz.

We do not have an image filter, so the Transmitter could just as well operate at 93.4GHz.

Dynamic Range Requirements

The system dynamic range requirements are as follows:

- Target RCS variation 30dB due to physical characteristics
- Target RCS variation 10dB due to wet/dry surface
- Because the area illuminated and hence the RCS is proportional to R^2, the range dependent change in signal level S_{rec} as predicted using the radar range equation is a function of R^{-2}.
- Dynamic Range = $20\log_{10}(R_{max}/R_{min})$ = 30dB

The total echo dynamic range is 30+10+30 = 70dB

Detector Options

The following detector options are considered

- Envelope Detector with an STC controlled variable gain amplifier to minimise the dynamic range requirements of the rest of the system.
- Successive detection Log Amplifier (SDLA) with an instantaneous dynamic range of greater than 70dB and no STC requirements.

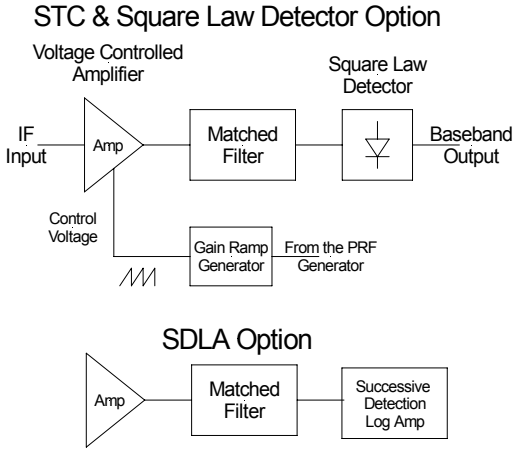


Figure 7.40: Detector options

Because of the uncertainties in the overall design (RCS levels etc), the SDLA is selected because its performance is more robust than the detector. It is also easier to interface to the post-detection electronics.

A Pascal SDLA has a DC voltage output proportional to the input power.

The specifications are as follows:

- Dynamic Range >70dB
- Tangential Sensitivity –75dBm
- Pulse rise time 3ns
- Pulse Decay time 6ns
- Transfer Function 25mV/dB
- Output level 2V for a 0dBm input signal

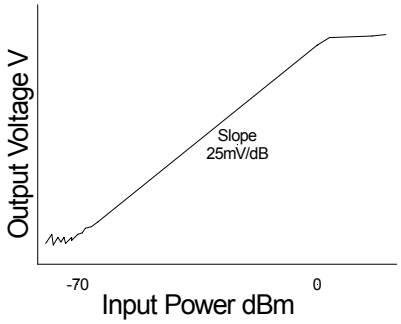


Figure 7.41: SDLA transfer function

7.11.11. Signal to Noise Ratio

Transmitted power $P_{tx} = P_{osc} - L_{line} - L_{circ}$ = 20-0.4-0.8=18.8dBm

SSB Noise Figure. If we use the formula which includes the mixer loss L_m = 8dB and an IF amplifier with a noise figure of 1.5dB as well as line losses $L_{rec} = L_{line} + L_{circ}$ = 0.4+0.8 = 1.2dB

$NF_{rec} = L_{rec} + L_m + NF_{IF}$ = 1.2+8+1.5 = 10.7dB

Matched Filter Loss L_{match} = 3dB is added to the noise figure making the total noise figure NF_{tot} =13.7dB.

7.11.12. Output Signal to Noise Ratio

The received power is calculated using the radar range equation which is re written in dB terms:

$$P_r = P_t + 2G + 10\log_{10}\frac{\lambda^2}{(4\pi)^3} + \sigma - 40\log_{10} R \quad \text{dBm} \tag{7.8}$$

At the maximum operational range of 300m, and using the mean RCS of 4.5dBm^2, the received power is:

P_r = 18.8 + 2x46 – 82.9 + 4.5 – 99 = -66.6dBm

The noise power in dBm for a bandwidth of 50MHz

$P_n = 10\log_{10}(kT\beta) + NF_{tot}$ = -127+13.7+30 = -83.2dBm

The signal to noise ratio SNR = -66.6 –(-83.2) = 16.6dB

However, because of fluctuations in the target RCS, the minimum predicted single pulse SNR may be 15dB lower than this:

SNR_{min} = 16.6-15 = 1.6dB

7.11.13. Required IF Gain

We want the minimum signal into the SDLA to equal –70dBm so that we can make use of the full dynamic range of the device.

The actual signal power after down conversion for the minimum predicted RCS at the longest range would be:

$P_{if} = P_r - L_{rec} - L_m$ –15 = -66.6-1.2-8 -15 = -90.8dBm

A minimum IF gain of 21dB would be required.

7.11.14. Detection Probability and Pulses Integrated

Assuming that we need a detection probability P_d = 0.95 and a very low false alarm probability $P_{fa} = 10^{-12}$, then we require an effective SNR of 16.3dB

To achieve a post detection integration gain of 16.3-1.6 = 14.7dB we need to integrate N pulses. Where $N = 10^{(14.7/8)}$ = 68 pulses.

Note that this is not altogether true as the formula was derived for a square law detector and we are using a SDLA. To compensate, we will integrate an additional 60 pulses (N = 128)

7.11.15. Measurement Update Rate

For a maximum unambiguous range of 300m, we can operate the radar at a maximum PRF of $c/2R_{max}$ = 500kHz. With 128 pulses integrated, the update rate for measurement output is reduced to 3.9kHz.

7.11.16. Monitoring Rock Falling Down the Pass

We assume that the rock that enters the pass accelerates due to gravity until it hits the bottom.

- There is no terminal velocity due to air resistance
- There is no terminal velocity due to friction from the walls of the pass

By the time the rock reaches 300m down it will be travelling at 76m/s. At an update rate of 3.9kHz, the rock will have moved all of 20mm between samples.

The Doppler shift will be $f_d = 2v/\lambda$ = 39kHz which is a very small fraction of the 50MHz IF bandwidth, so can be ignored.

7.12. Prototype Build and Test

A prototype pulsed radar unit was built as described

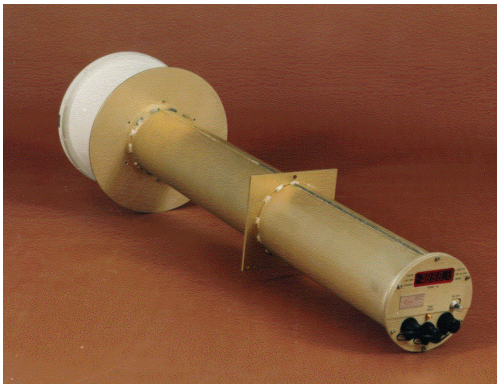


Figure 7.42: The prototype orepass radar

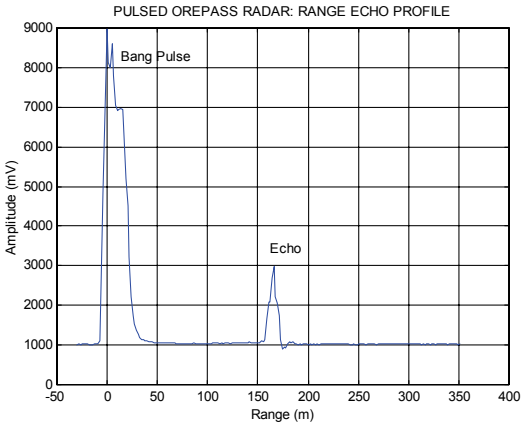


Figure 7.43: Orepass echo profile obtained using a pulsed W-band radar

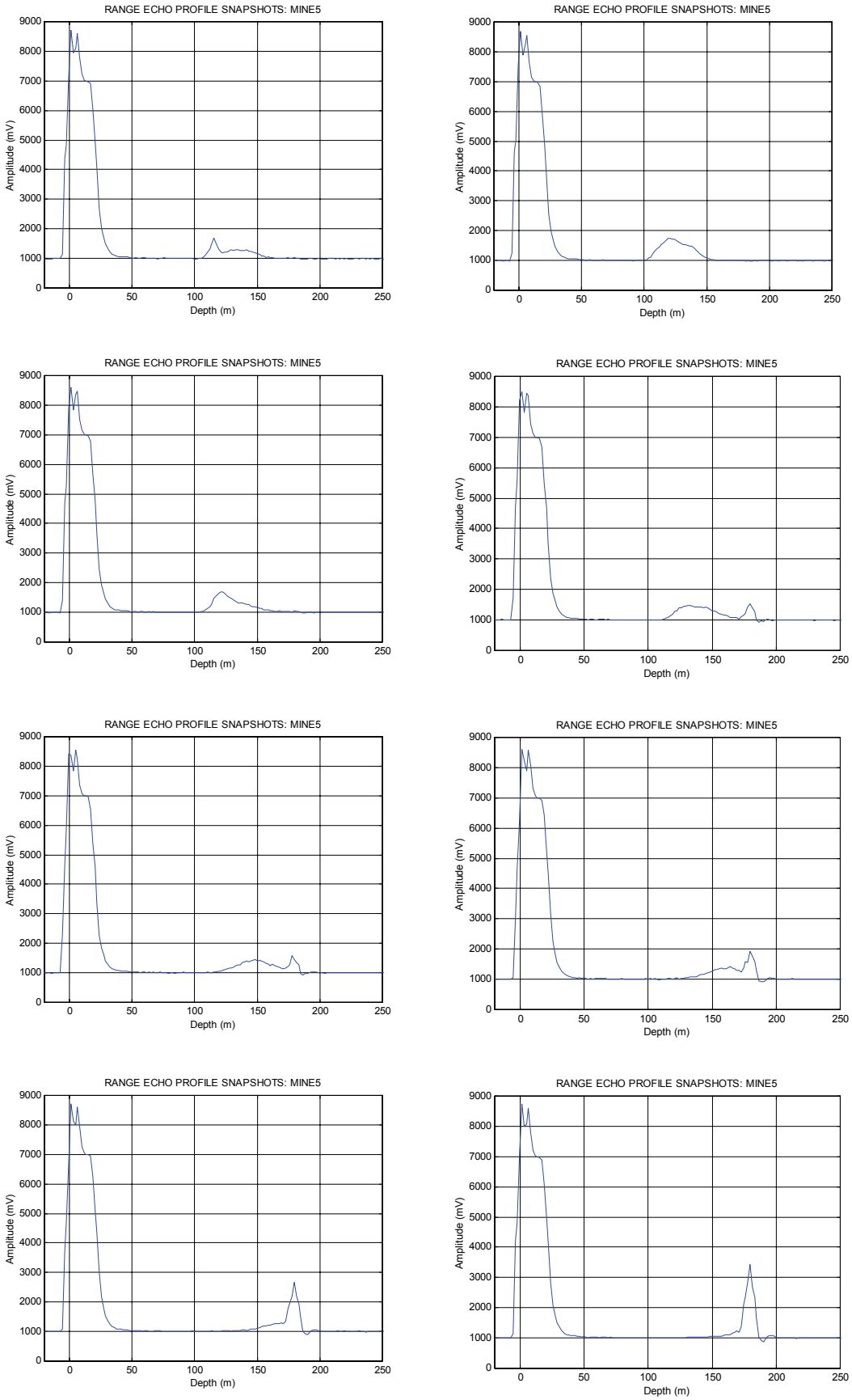


Figure 7.44: Pulsed radar snapshots of rock falling down a pass.

7.13.References

[1] http://www.sensormag.com/articles/1000/55/main.shtml, 30/11/2000.
[2] http://www.control.com/control_com/950622665/index.html, 30/11/2000.
[3] Noncontact level measurement – an alternative to intrusive technologies, *SA Instrumentation and Control,* May 1998.
[4] R. Grover, Radar level measurement – an overview, *SA Instrumentation and Control,* May 1998.
[5] J. Foszcz, Selecting Liquid Level Detectors for Tanks, http://www.manufacturing.net/magazine/planteng/3500/articles/98/023528.html, 30/11/2000.
[6] W. Hendrick, Industrial Applications of Radar Technology for Continuous Level Measurement, *Proceedings of the 32nd Symposium. Instrumantation in the Pulp and Paper Industry.* Vol. 23, 1992.
[7] Radiation Based Level Gages. http://www.omega.com/literature/transactions/volume4/T9904-14-RAD.html, 30/11/2000.
[8] http://www.krohne.com, 16/08/2000
[9] Radar Level Measurement, *Krohne Brochure*, 10/1991.
[10] Microwave Level Measurement. Micropilot FMR 130*, Endress+Hauser Technical Brochure*, January 1999.
[11] RangeMaster Ultrasonic Level Transmitter. *Hawk Installation Instructions*, 1995.
[12] Acoustic Level Measurement: Principles and Application*. Hawk Publication*, 2000.
[13] Level Measurement Solutions Guide. Milltronics
[14] Milltronics IQ Radar, http://www.milltronics.com/level/iqradar.html, 22/02/2000.
[15] Simple Solution fir Noncontact Liquid Level Measurement. The Milltronics Probe. *SA Instrumentation and Control,* may 1998.
[16] The probe, http://www.milltronics.com/level/probe.html, 17/08/2000.
[17] ST, LR and ST-H Series Transducers, http://www.milltronics.com/level/probe.html, 22/02/2000.
[18] AiRanger XPL, Non Contacting Ultrasonic measurement, *Milltronics Brochure*
[19] An Acoustic transducer System, *European Patent Application,* 0 039 986
[20] SAAB TankRadar PRO, *SAAB Brocure*
[21] Radar Technology, http://www.saabtankcontrol.co.uk/technol.html, 30/11/2000.
[22] SAAB TankRadar L/2, *SAAB Brochure,* September 1994.
[23] Product Review 97/98, *VEGA Brochure.*
[24] Apex and Apex Sentry. *Rosemount Brochure*.
[25] Eyesafe Laser Rangefinder; LaserTape FG21, http://www.riegl.co.at, 26/02/2001.
[26] Industrial Distance Sensor: LD90-450, http://www.riegl.co.at, 26/02/2001.
[27] Principle of a Pulsed Laser Sensor, http://www.riegl.co.at, 26/02/2001.
[28] Product Range, *Laser m Technical Brochure*
[29] K.Tsipis, Cruise Missiles, *Scientific American,* February 1977.
[30] http://physics.iop.org/Physics/Electron/Exhibition/section5/magnetron.html, 21/02/2001.
[31] http://gallawa.com/microtech/mag_test.com, 21/02/2001.
[32] Hughes Millimeter-Wave Products for 1987/1988
[33] M.Skolnik, *Radar handbook,* McGraw-Hill, 1970.
[34] M.Skolnik, *Introduction to Radar Systems,* McGraw-Hill, 1980.
[35] Micropower Impulse Radar, http://lasers.illnl.gov/lasers/idp/mir/files/MIR_govt_info.html, 16/02/2001
[36] J Stalley, Interfacing with a Laser Rangefinder, Honours Thesis, AMME, University of Sydney, 2005
[37] Exploring the Ultrawideband, http://www.eurekalert.org/features/doe/2004-09/dlnl-etu091604
[38] P Probert-Smith, *Active Sensors for Local Planning in Mobile Robotics,* World Scientific, 2001

.

Chapter 8.

Signal Propagation

8.1. The Sensing Environment

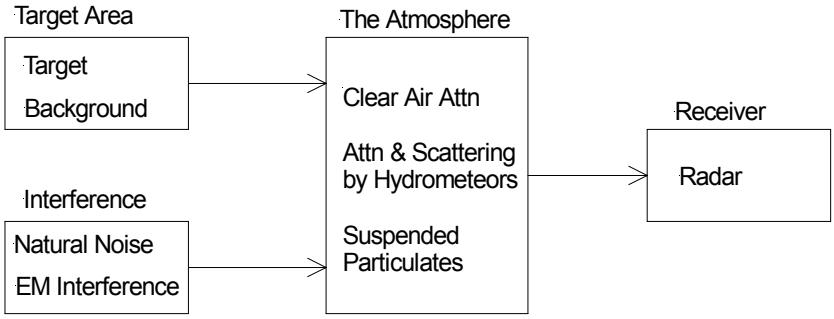


Figure 8.1: The sensing environment

8.1.1. Target Area

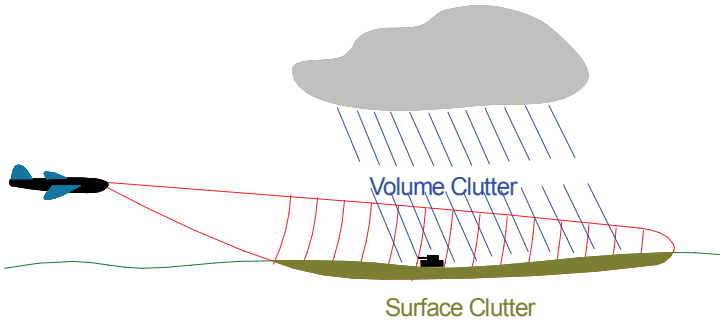


Figure 8.2: Schematic diagram illustrating the surface and volume clutter returns that compete with the target echo within each range gate

The background returns in the same range cell that compete with the target return are referred to as Clutter. These returns may be from the surface surrounding the target (ground), or from the volume of space around it (rainfall) as shown in the figure above.

8.1.2. Interference

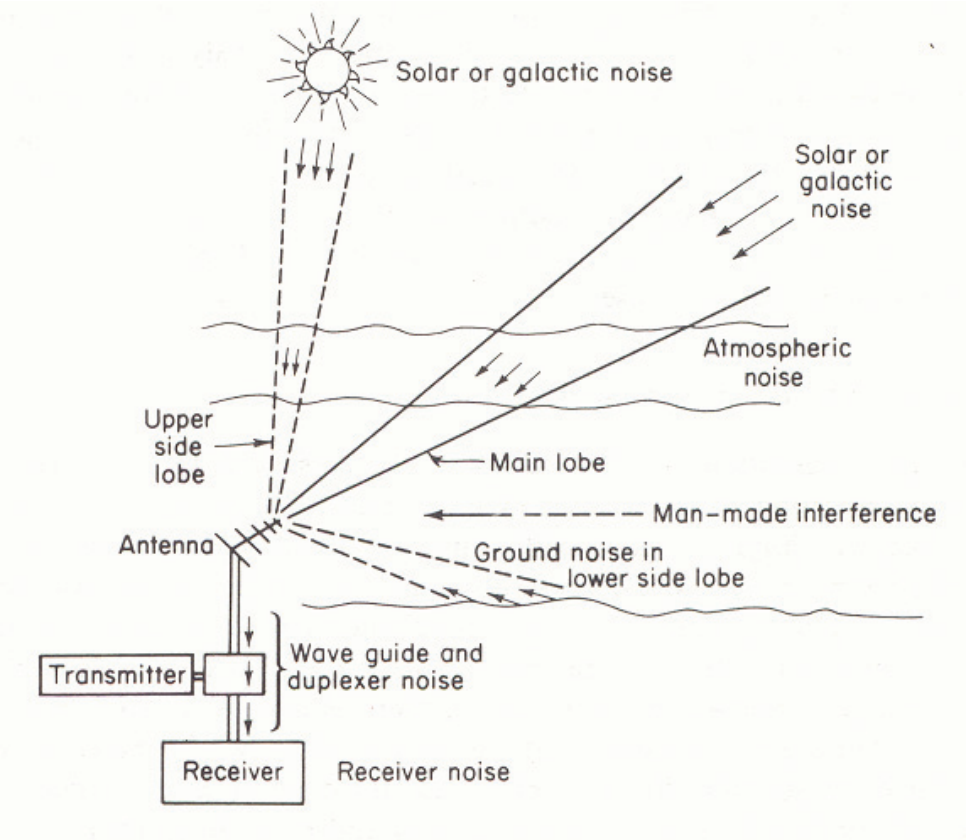


Figure 8.3: Sources of noise

Interference may come from natural sources such as the sun and local galaxies, it may be man made, for example car ignition systems, fluorescent lights, other sensors etc or it may be generated in an attempt to deceive the sensor, in which case it is known as jamming

8.1.3. The Atmosphere & Hydrosphere

Clear air attenuation is mostly caused by molecular absorption and scattering by O_2 and H_2O, attenuation and scattering caused by hydrometeors including rain, hail, clouds and fog.

Attenuation and scattering are also caused by suspended paticulates including dust and smoke. Some smokes for IR systems and Chaff for microwave systems are released to deceive or blind the sensor.

Attenuation caused by friction etc for underwater sonar systems

8.1.4. Multipath Effects

Multipath propagation effects are common with both sonar and radar systems. These are caused by multiple paths of different lengths between the sensor and the target that result in constructive or destructive interference at the receiver. These effects are most severe over flat smooth surfaces like water or agricultural lands for radar. In the case of sonar, both the water surface and the flat sea-bottom are culprits.

8.2. Propagation of Electromagnetic Waves

8.2.1. Signal Attenuation and the Range Equation

As an electromagnetic signal propagates through the atmosphere, molecular interactions with the wave absorb energy, and the signal amplitude decreases with range.

For the attenuation α per unit metre, the relationship between the received and transmitted power can be written as follows:

$$P_r = P_t e^{-2\alpha R}. \tag{8.1}$$

Using the change of base theorem

$$\log_a N = \frac{\log_b N}{\log_b a}, \tag{8.2}$$

it can be shown that $\log_{10} N = 0.43429 \log_e N$

This allows the attenuation to be expressed as $\alpha_{dB} = 4342.9\alpha$ in dB/km (one way), making the relationship:

$$10\log_{10}(P_r) = 10\log_{10}(P_t) - 2\alpha_{dB}R_{km}, \tag{8.3}$$

where: P_r – Received power (w),
P_t – Transmitted power (w),
α_{dB} – Attenuation (dB/km),
R_{km} – Range (km).

The non-linear term makes the range equation difficult to solve in a closed form and numerical or graphical methods are normally resorted to.

Writing the range equation for a monostatic radar system in dB to include this attenuation term as follows

$$10\log_{10}(P_r) = 10\log_{10}(P_t) + 10\log_{10}\left(\frac{\lambda^2}{(4\pi)^3}\right) + 20\log_{10}(G) + 10\log_{10}(\sigma) - 10\log_{10}(L) - 40\log_{10}(R) - 2\alpha R_{km} \quad (8.4)$$

8.3. Attenuation of Electromagnetic Waves

8.3.1. Clear Weather Attenuation

Primary absorption by O_2 and H_2O

Because absorption is proportional to the absolute humidity, the graph shown in the figure below is adjusted to be representative of a water content of 7.5g/m^3 at 20°C.

Below 60GHz there are two main absorption bands, that from the 1.35cm (22.235GHz) water vapour line and its skirts, and a series of oxygen lines centred around 0.5cm (60GHz). These two peaks bracket the 35GHz window.

The atmosphere becomes progressively more opaque at higher frequencies, with windows at 94, 140, 220 and 360GHz between the O_2 and H_2O lines and is totally opaque between 1THz an 10THz (0.3mm and 30μm).

Windows appear between absorption lines of H_2O and CO_2 in the IR region starting at 10μm and attenuation is very low in the visible region around 500nm.

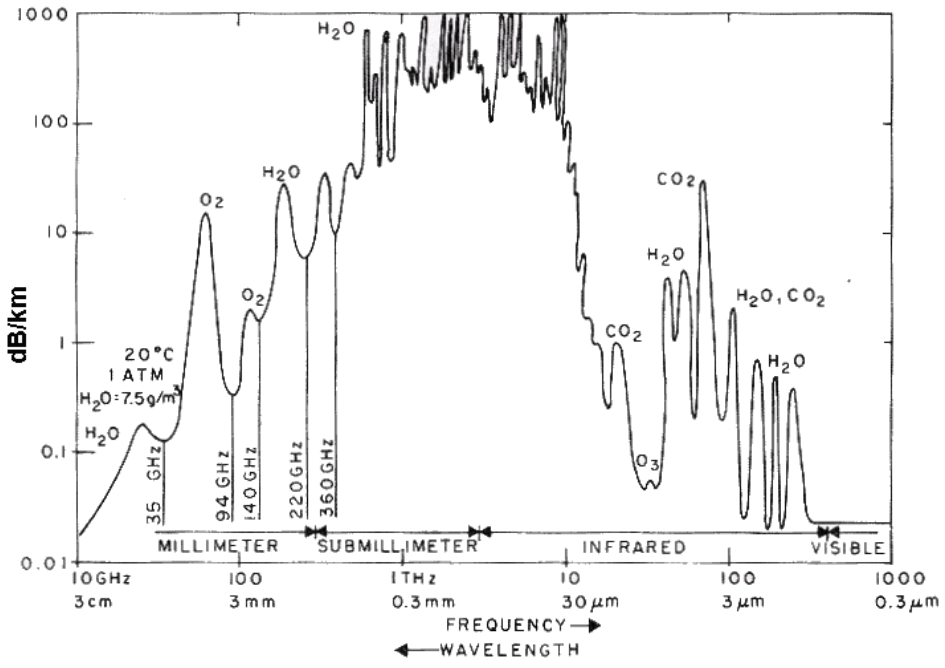


Figure 8.4: Clear weather attenuation as a function of frequency

8.3.2. Effect of Atmospheric Pressure (air density)

Because absorption is proportional to air density, this graph above is adjusted to be representative of the pressure at sea level.

The figure below shows the difference in the attenuation (for micro and millimetre waves) at sea level and 4km up from which it can be seen that the effects of resonances are more pronounced at high altitude.

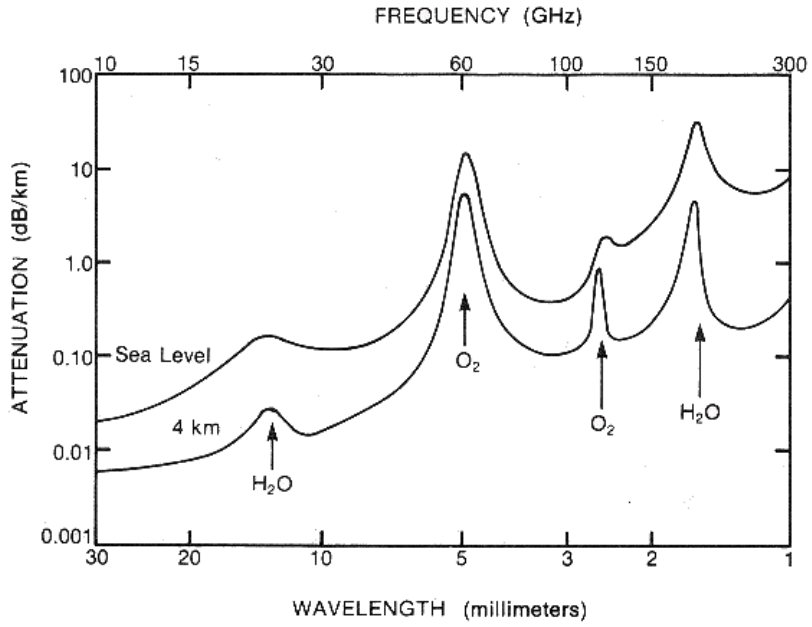


Figure 8.5: Average atmospheric absorption of millimetre waves for horizontal propagation at two different altitudes

8.3.3. Effect of Rain

The effect of light rain that has a uniform spatial distribution can be predicted with relative ease. However heavier rain normally comes in squalls with unpredictable intensity and drop size distribution

Mie (resonance) scattering theory is generally applied for frequencies up to the sub-millimetre wave band

$$\alpha_r = aR^b \tag{8.5}$$

where: α_r – Attenuation (dB/km),
R – Rain rate (mm/hr),
a,b – Coefficients dependent on frequency and temperature.

Table 8.1: Empirical coefficients used to determine attenuation

Coefficient	35GHz	94GHz
a	0.27	1.6
b	0.985	0.64

The Laws Parsons model is used to determine drop-size distributions if they are required with the mean raindrop size proportional to rainfall rate and the atmospheric conditions.

Table 8.2: Typical drop size for different conditions

Condition	Drop Size (μm)
Haze	0.01-3
Fog	0.01-100
Clouds	1-50
Drizzle	3-800
Moderate rain (4mm/hr)	3-1500
Heavy rain (16mm/hr)	3-3000

For typical conditions, the attenuation with frequency up to 300GHz is plotted here at a number of different rainfall rates. Note that the absorption peak shifts down in frequency as the rain rate increases because the mean drop size is proportional to rainfall rate.

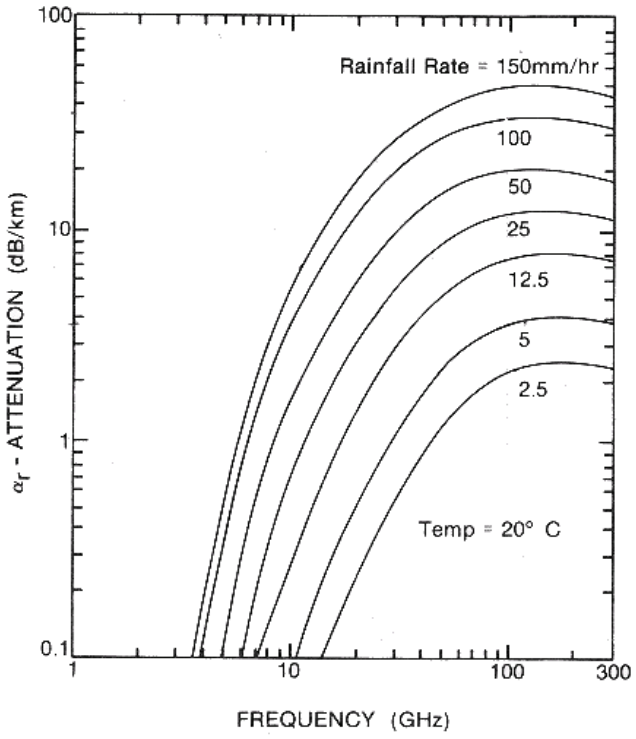


Figure 8.6: Theoretical rainfall attenuation as a function of frequency for different rainfall rates

At IR and visible wavelengths, the drop size is much larger than the wavelength, so the attenuation will be independent of drop diameter. At these frequencies attenuation is proportional to water content in g/m^3, and hence directly proportional to rain rate,

$$\alpha = cR, \tag{8.6}$$

where: c – Scaling coefficient,
R – Rain rate (mm/hr).

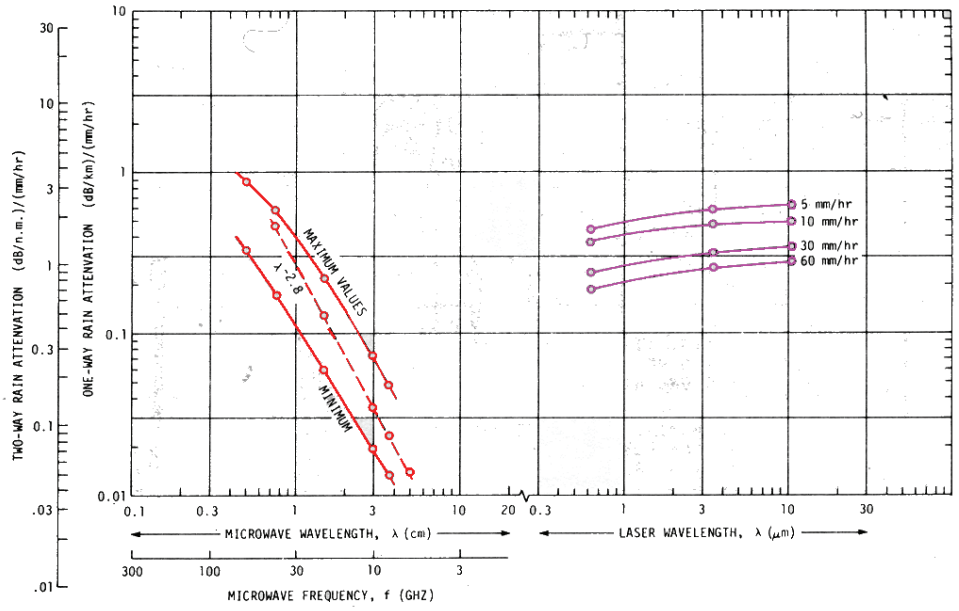


Figure 8.7: Specific attenuation (dB/km)/(mm/hr) in rain

8.3.4. Effect of Fog and Clouds

Clouds and fog are caused by the condensation of atmospheric water vapour into water droplets suspended in the air.

Fog is defined as a water droplet density that restricts the visibility to less than 1km where advection fog is caused by a horizontal movement of a warm air mass over cold water and radiation fog is caused by the cooling of air overnight (often occurs over rivers and swamps)

Clouds are similar in nature to fog with droplets less than 100μm in diameter.

Table 8.3: Fog characteristics

	Radiation Fog	Advection Fog
Average Drop Diameter	10μm	20μm
Typical Drop Size Range	5-35μm	7-65μm
Liquid water Content	0.11g/m^3	0.17g/m^3
Droplet Concentration	200/cm^3	40/cm^3
Visibility	100m	200m

An empirical relationship that relates visibility to liquid water content for advection fog is

$$M = 1660V^{-1.43}, \tag{8.7}$$

where M – Liquid water content (g/m^3),
V – Visibility (feet) (1foot=0.308m).

For a visibility of 125m (V = 400ft) M = 0.32g/m^3

Because the drop diameter of both fog and clouds is less than 100μm, the Rayleigh (low frequency) approximation can be used to evaluate the attenuation at microwave and millimetre wave frequencies

$$\alpha_{f,c} = k\rho_o \text{ dB/km}, \tag{8.8}$$

where: k – Attenuation coefficient (dB/km per g/m^3) at a specific frequency (see the figure below),
ρ_o – Water content (g/m^3).

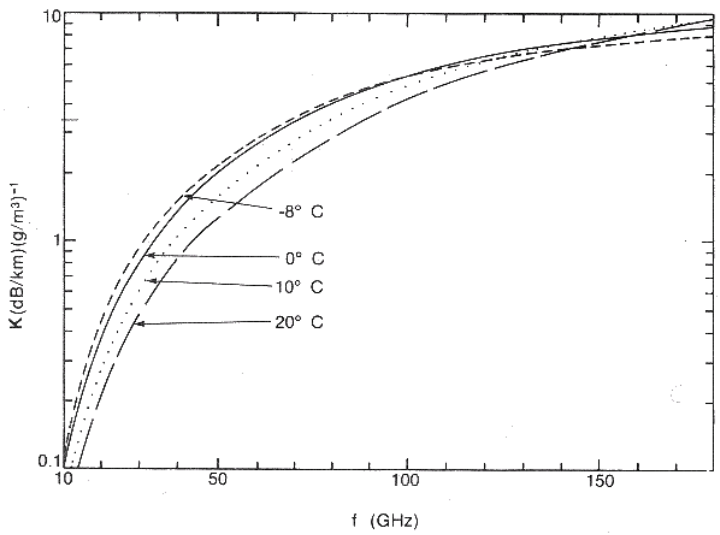


Figure 8.8: Specific attenuation through clouds and fog

Using the formula and taking k from the figure, for f = 100GHz, gives an attenuation of 4dB/km per g/m^3 at 20°C making the attenuation in fog with a visibility of 125m (400ft) is $\alpha = 4 \times 0.32 = 1.28$dB/km

For a moderate fog with liquid water content of 0.1g/m^3, this equates to an attenuation of 0.4dB/km. This ties in quite well with Figure 8.10 even though the predicted visibility does not!!

Table 8.4: Liquid water content of clouds

Cloud Type	Liquid Water Content (g/m^3)
Heavy Fog	0.15-0.4
Moderate Fog	0.01-0.1
Cumulus	0.5-1
Altostratus	0.35-0.45
Stratocumulus	0.3-0.6
Nimbostratus	0.6-0.7
Stratus	0.15-0.3

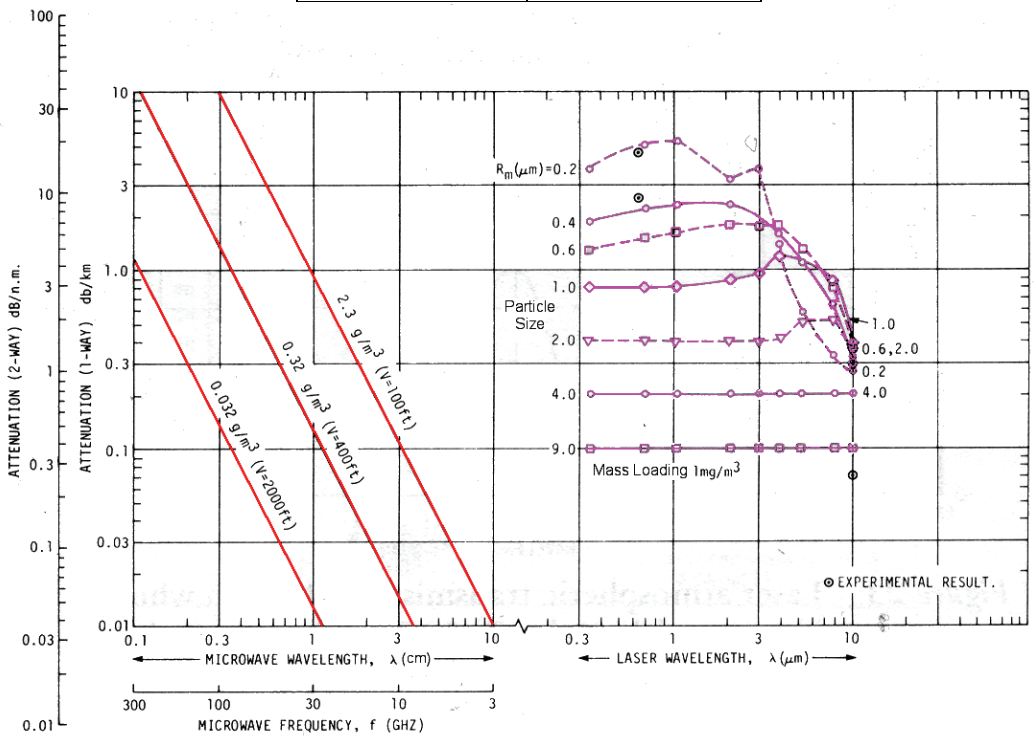


Figure 8.9: Comparison of radar and laser fog attenuation

As the frequency increases, the wavelength becomes an appreciable percentage of the circumference of the drop and resonance occurs (near IR region).

Because there is such a large variation in drop size and LWC for mist and fog, the actual attenuation that will occur in the IR region around this resonance is difficult to determine accurately. However, as a rule of thumb, the following formulae can be applied.

The figure above shows the laser attenuation (dB/km per mg/m^3) for various drop diameters R_m in μm

$$\alpha_{f,c} = k\rho_o \text{ dB/km}, \tag{8.9}$$

where: k – Attenuation coefficient (dB/km per mg/m^3) at a specific frequency and a specific R_m (See the figure below),
ρ_o – Water content (mg/m^3).

For radiation fog with a drop diameter D = 8μm (R_m = 4μm) and a LWC of 0.1g/m^3 (100mg/m^3) the attenuation is 0.2×100 = 20dB/km (one way).

For a heavy advection fog (visibility 50m), the LWC was calculated to be 1.2g/m^3 (1200mg/m^3) for the same drop diameter (8μm) the attenuation is 0.2×1200 = 240dB/km

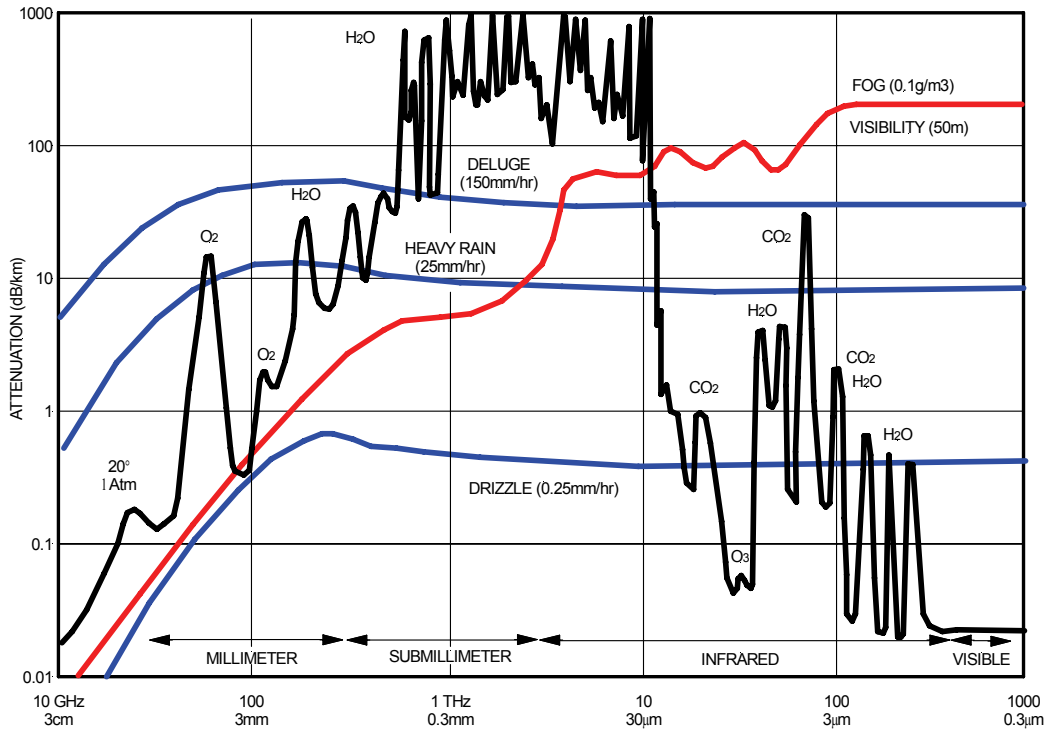


Figure 8.10: Summary of atmospheric attenuation for clear air, rain and fog

This figure summarises the effect of rain and fog as a function of frequency. Note the resonance effects due to fog particle size at wavelengths between 30 and 3μm.

Note the peak in the rain attenuation at wavelengths between 3mm and 1mm depending on the rainfall rate.

8.3.5. Other Solid Materials

Absorption of EM radiation by solids and liquids is determined by their relative dielectric constant ε_r and loss tangent, $\tan\delta$, of the material.

For most materials this is a function of frequency.

For normal incidence, the reflection coefficient is

$$|\Gamma| = \left|\frac{Z_L - Z_o}{Z_L + Z_o}\right| = \frac{\sqrt{\varepsilon_r} - 1}{\sqrt{\varepsilon_r} + 1}, \tag{8.10}$$

where: Z_o – Characteristic impedance of free space $\sqrt{\mu_o / \varepsilon_o}$,

Z_L – Impedance of the material $\sqrt{\mu_o / \varepsilon_o \varepsilon_r}$.

The attenuation in dB for propagation through an unbounded dielectric material is

$$\alpha_d = 27.3\sqrt{\varepsilon_r}\tan\delta\frac{d}{\lambda_o}, \tag{8.11}$$

where; α_d – One way attenuation (dB),
ε_r – Relative dielectric constant,
$\tan\delta$ - Loss tangent,
d – Distance (m),
λ_o – Wavelength (m).

Material characteristics are important for the design of radomes and lens antennas or to determine the propagation characteristics for ground penetrating radar (GPR).

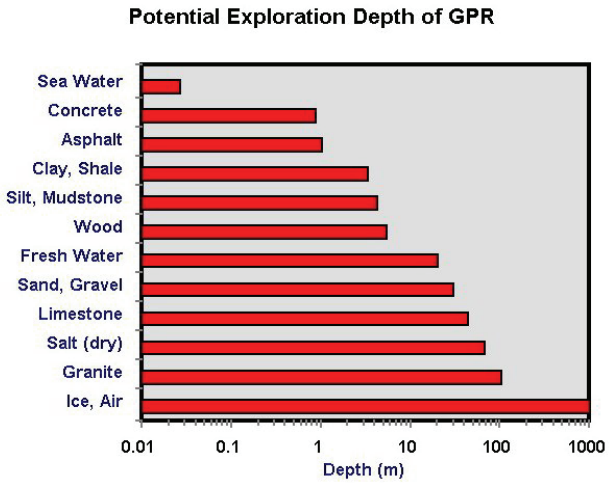


Figure 8.11: Potential GPR penetration depth

8.3.6. Attenuation Through Dust and Smoke

A good way to characterise the density of aerosols suspended in the atmosphere is according to the visibility of targets within the environment. This is still a fairly subjective measure, however, as the visibility is defined as the range at which a high contrast target can just be seen.

Using this definition, it has been found that in a dust storm the visibility can be related to the mass of dust per cubic meter of air (mass loading) by

$$M = \frac{C}{V^{\gamma}} \tag{8.12}$$

where M is the mass loading of dust in g/m^3, V is the visibility in metres and C and γ are constants that depend on the particle type and the meteorological conditions. Typical values for these constants are C = 37.3 and γ = 1.07. These are used to calculate the values in the table below.

Table 8.5: Relationship between visibility and mass loading

Visibility (m)	Mass Loading (g/m^3)
0.5	78
1	37
2	17.8
4	8.5
8	4
12.75	2.44**

** Average for sand storms in the Sudan

Ultimately it is not mass loading, but volume loading that determines the visibility, and (6) is often written as

$$v = \frac{C}{\rho V^{\gamma}}, \tag{8.13}$$

where v is the relative volume occupied by particles and ρ is the density of the dust material. For the constant C determined previously, ρ = 2.44g/cm^3 and thus a new constant C/ρ = 15.3 can be inserted into (8.13)

$$v = \frac{15.3}{V^{1.07}} \ \ (\text{cm}^3/\text{m}^3). \tag{8.14}$$

Attenuation of Laser Signals

It is possible to calculate the laser attenuation as a function of the visibility and the mass loading. However, measurements made at 0.55μm and 10.4μm are reproduced below and show that the measured transmission in the visible and IR wavelengths are similar. This

means that the most of the attenuation is caused by large particles operating in the optical region.

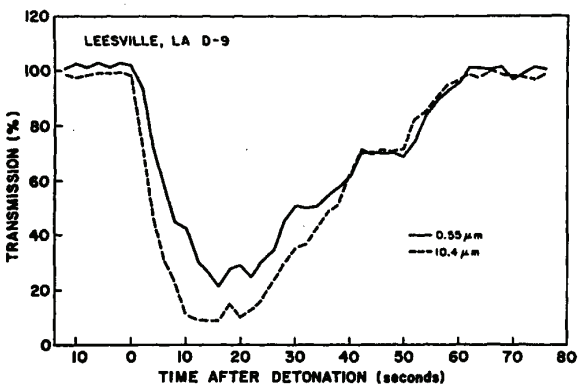


Figure 8.12: Measured transmission through explosion generated dust showing good correlation between visible and IR wavelengths

This relationship confirms the well known rule-of-thumb that states that *if you can see through it, a laser range finder can probably measure the range.*

Attenuation of Radar Signals

It can be shown that the attenuation of radar signals using the Rayleigh approximation is given by

$$\alpha = \frac{2.317\times10^{-3}.\varepsilon''}{\left[(\varepsilon'+2)^2+\varepsilon''^2\right]\lambda}.\frac{1}{V^{1.07}} \quad \text{(dB/km).} \qquad (8.15)$$

where ε' is the real component of the permittivity of the dust particle and ε'' is the imaginary component. These are also known as the dielectric constant and the loss factor. The permittivity of coal is ε_{coal} = 4.21-j0.156.at X-band. There are no measurements available at W-band.

The following graph shows the relationship between the visibility and the attenuation through coal dust

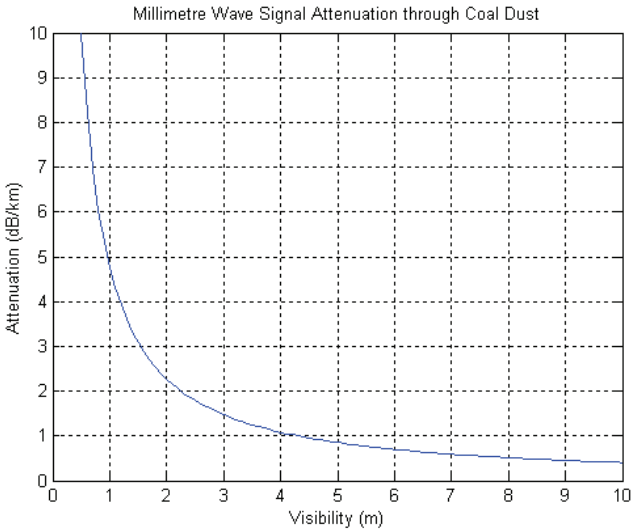


Figure 8.13: Relationship between visibility and attenuation for coal dust at 94GHz

It can be seen that in contrast to the poor transmission at visible and IR wavelengths, at a wavelength of 3.2mm (94GHz), attenuations are practically negligible for short range operation.

If the coal dust is replaced by water with a permittivity ε_w = 8.35-j15.45 at 89GHz with the particle/droplet distribution remaining unchanged, then the losses are much higher as shown.

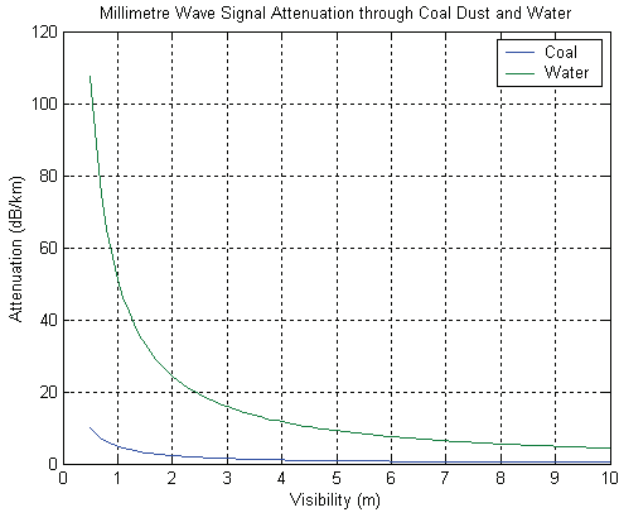


Figure 8.14: Relationship between visibility and attenuation for coal dust and for water droplets at 94GHz

These results can be compared to those obtained for rain earlier in this chapter. Though the losses appear large, the worst loss for a visibility of 5m is only about 10dB/km.

Table 8.6: Calculated attenuation through different dust types

Source	Material	Density (g/cm^3)	Dielectric const.	Diameter (cm)	Mass conc (g/m^3)	Attn (dB/km)	
						λ=4mm	λ=1-10μm
Dust: Winds	Quartz	2.6	2.25	1.E-03	10	2.20E-06	216.3
Stack: Stone Mill	Quartz	2.6	2.25	1.E-03	80	1.76E-05	1730.1
Stack: Steel Mill	Carbon	2	1000	1.E-03	10	3.28E-05	3230.6
Volcanic Dust	Quartz	2.6	2.25	2.E-04	5	8.80E-09	540.7
Bacteria	Water	1	80	2.E-04	5.E-05	2.45E-12	0.2
Pollen	Biological	0.5	1.5	1.E-03	60	1.62E-05	1591.8
Light fog	Water	1	80	1.E-03	1.E-03	6.13E-09	0.6
Heavy fog	Water	1	80	2.E-03	1	4.91E-05	301.7

8.4. Refraction of Electromagnetic Waves

Refraction effects are independent of wavelength at frequencies below 100GHz. The decrease in air density, and hence refractive index with height, results in a downward curvature of horizontally launched radar beams. Bending is equal to about ¼ of the curvature of the earth.

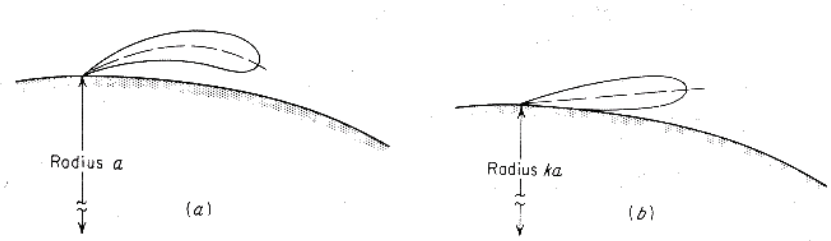


Figure 8.15: Bending of the beam due to refraction by the earth's atmosphere leads to an increase in the effective earth radius to produce an undeflected beam

To plot ray paths as straight lines, the effective earth radius is 4/3 of the actual value.

The radar horizon can be approximated by

$$d = \sqrt{2kah}\,, \tag{8.16}$$

where: d – radar horizon (m),
k – 4/3,
a – actual radius (6378×10^3m),
h – radar height (m).

This can be simplified to the following for the both horizon range and radar height in km

$$d(km) = 130\sqrt{h(km)}\,. \tag{8.17}$$

8.5. Acoustics and Vibration

Alternate compression and expansion of the medium (solid, liquid or gas) with certain frequencies are called sound waves.

Table 8.7: Common units of pressure

Units	Atmospheric Pressure
Atmospheres (atm)	1
Bar	1.01325
Pascal (Pa)	1.01325×10^5
Pounds/in^2 (psi)	14.6960
Inches Hg	29.9213
Mm of Hg (torr)	760
Inches of water	406.8
Dynes/cm^2	1.01325×10^6

8.5.1. Sound Propagation in Gases

The speed of sound in an ideal gas can be written in terms of the equations of state ($PV = nRT$)

$$c = \sqrt{\frac{RT\gamma}{M}} = \sqrt{\frac{P\gamma}{\rho_o}} = \sqrt{\frac{K}{\rho_o}} \quad \text{cm/s,} \tag{8.18}$$

where: c – Velocity of sound in the gas (cm/s),
R – Universal gas constant (8.314×10^7 ergs/mole/K),
T – Temperature (K),
M – Molecular Weight of the gas (g/mole),
γ - Ratio of specific heats (adiabatic exponent) C_p/C_v for the gas
P – Pressure (1atm = 1.013×10^6 dyne/cm^2),
ρ_o - Density (g/cm^3),
K – Bulk modulus ($K=\Delta P/\Delta V$).

The adiabatic exponent γ can be estimated as follows:

- 1.66 for monatomic gases (He, Ne, Ar),
- 1.40 for diatomic gases (H_2, O_2, N_2),
- 1.33 for triatomic and more complex gases (NH_3, CH_4, C_7H_8),
- 1.286 for very long molecules.

The table two pages on gives a more accurate measure of the adiabatic exponent for various gases.

At atmospheric pressure, the volume of one mole of gas is 22414 cm^3.

Effect of Molecular Weight

The speed of sound in any gas mixture can be calculated by using the molecular weight to determine the density as follows:

- Air is made up of O_2, N_2, and CO_2 in proportions (21:78:1)
- The molecular weights are 16+16 = 32, 14+14=28, 12+16+16 = 44

$$M = \frac{21\times 32}{100} + \frac{78\times 28}{100} + \frac{1\times 44}{100} = 29\,,$$

$$\rho_o = \frac{29}{22414} = 0.0013 \qquad \text{g/cm}^3,$$

$$c = \sqrt{\frac{1.01325\times 10^6 \times 1.404}{0.0013}} = 3.3147\times 10^4 \quad \text{cm/s.}$$

Effect of Temperature and Pressure

The density of the gas at any temperature and pressure is related to that at STP for the pressure in mm of mercury and the temperature in Kelvin.

$$\rho = \rho_o\left(\frac{P}{760}\right)\left(\frac{273.16}{T}\right). \tag{8.19}$$

For rough calculations of the speed of sound (ignoring changes in C_v) the changes in density are assumed to be the only variable affecting the velocity.

Table 8.8: Velocity of Sound in Gases

Gas	Formula	Speed (m/s) at 0°C
Air		331.45
Ammonia	NH_2	415
Argon	A	307.8
Carbon monoxide	CO	337.1
Carbon dioxide	CO_2	258.0
Carbon disulfide	CS_2	189
Chlorine	Cl_2	205.3
Ethylene	C_2H_4	314
Helium	He	970
Hydrogen	H_2	1269.5
Methane	CH_4	432
Neon	Ne	435
Nitric oxide	NO	325
Nitrogen	N_2	337
Nitrous oxide	N_2O	261.8
Oxygen	O_2	317.2
Steam (100°C)	H_2O	404.8

This table lists the constants that are required to calculate the speed of sound of various gases at different temperatures and pressures.

Table 8.9: Specific Heat at Constant Pressure C_p and the Ratio γ of C_p to the Specific Heat at Constant Volume C_v [*]

Gas	Temp °C	C_p (cal/g deg)	γ (C_p/C_v)
Air	-120 (10atm)	0.2719	1.415 (-118°C 1atm)
	-50 (10atm)	0.2440	1.408 (-78°C 1atm)
	0	0.2398	1.403
	50 (10atm)	0.2480	1.403 (17°C 1atm)
	100	0.2404	1.401
	400	0.2430	1.393
	1000	0.2570	1.365
	1400	0.2699	1.341
Ammonia	15	0.5232	1.310
Argon	15	0.1253	1.668
Carbon dioxide	15	0.1989	1.304
Carbon monoxide	15	0.2478	1.404
Chlorine	15	0.1149	1.355
Ethylene	15	0.3592	1.255
Helium	-180	1.25	1.660
Hydrogen	15	3.389	1.410
Hydrogen sulphide	15	0.2533	1.32
Methane	15	0.5284	1.31
Neon			1.64
Nitric oxide	15	0.2329	1.400
Nitrogen	15	0.2477	1.404
Nitrous oxide	15	0.2004	1.303
Oxygen	15	0.2178	1.401
Propane			1.13
Steam	100	0.4820	1.324
Sulphur dioxide	15	0.1516	1.29

8.5.2. Sound Propagation in Water

As with gases, the speed is equal to the bulk modulus of elasticity divided by the density of the liquid.

$$c = \sqrt{\frac{B}{\rho}} \text{ m/s.} \tag{8.20}$$

Because the modulus of elasticity and the density are functions of the pressure (depth), temperature and salinity, an empirical equation derived by Wilson (1960) is usually used to calculate the speed.

$$c = 1449.2 + 4.62T - 0.055T^2 + 1.4(S - 35) + 0.017z \quad \text{m/s}, \tag{8.21}$$

where: T – Temperature (°C),
S – Salinity (parts per thousand),
z – Depth (m).

Table 8.10: Speed of sound in fresh water

Temp (°C)	Speed p=1atm (m/s)		Pressure (psi)	Speed T=30°C (m/s)
0	1,403.0		14.7	1,509.7
10	1,447.8		2,000	1,532.9
20	1,482.9		4,000	1,556.2
30	1,509.7		6,000	1,579.4
40	1,529.3		8,000	1,602.4
50	1,542.9		10,000	1,625.3
60	1,551.3		12,000	1,648.1
70	1,555.1		14,000	1,670.9
80	1,554.7			
90	1,550.5			
100	1,542.5			

14.7lbf/in^2 = 101kN/m^2 = 1.01bar
1000lbf/in^2 = 6.89MN/m^2 = 68.9bar

1 dyne = 10μN
1 kgf = 9.807N
1 Pa = 1 N/m^2
10m depth = 1bar increase in pressure

Note that the speed of sound at the sea surface (35ppt salt) at 0°C is 1449.2m/s as contrasted with 1403m/s for fresh water, a fact that might be useful in determining the salinity of a water sample.

Table 8.11: Speed of sound in sea water at p = 1atm

Temp °C	Speed (m/s)		
	S = 3.3%	S = 3.5%	S = 3.7%
-3	1431.9	1435.0	1437.6
0	1446.3	1449.4	1451.9
5	1468.2	1471.2	1473.5
10	1487.6	1490.4	1492.7
15	1504.6	1507.4	1509.5
20	1519.6	1522.2	1524.2
25	1531.9	1535.1	1536.9
30	1543.8	1546.2	1547.9

Table 8.12: Speed of Sound in Sea Water at T = 20°C

Pressure kg/cm^2	Speed (m/s)		
	S = 3.3%	S = 3.5%	S = 3.7%
1.033	1519.6	1522.2	1524.1
100	1535.4	1537.1	1540.1
200	1551.5	1554.2	1556.3
300	1567.7	1571.5	1572.5
400	1584.0	1586.8	1588.9
500	1602.0	1603.1	1605.2
600	1616.8	1619.5	1621.6
700	1633.1	1635.8	1637.9
800	1649.4	1652.1	1654.2
900	1665.5	1668.2	1670.3

8.5.3. Sound Propagation in Solids

In an isotropic solid for which the wave-front is a large number of wavelengths, longitudinal and shear waves can exist which have velocities

$$c_{long} = \sqrt{\frac{\lambda + 2\mu}{\rho_o}} \quad \text{m/s.} \tag{8.22}$$

$$c_{shear} = \sqrt{\frac{\mu}{\rho_o}} \quad \text{m/s.} \tag{8.23}$$

where: μ,λ - Lame Elastic Moduli for the material (N/m^2)

For a rod, whose diameter is a small fraction of the wavelength, extensional and torsion waves can propagate

$$c_{tor} = \sqrt{\frac{\mu}{\rho_o}} \quad \text{m/s.} \tag{8.24}$$

$$c_{ext} = \sqrt{\frac{Y_o}{\rho_o}} \quad \text{m/s.} \tag{8.25}$$

where: $Y_o = \mu \dfrac{3\lambda + 2\mu}{\lambda + \mu}$ N/m^2.

Table 8.13: Elastic constants, wave velocities and characteristic impedances of various materials

Material	Y_o N/m^2	μ N/m^2	λ N/m^2	Poisson's Ratio, σ	C_{long} m/s	c_{shear} m/s	c_{ext} m/s	Z_I kg/s m^2	Z_s kg/s m^2
Aluminium	70×10^9	25×10^9	61×10^9	0.355	6420	3040	5000	17.3×10^6	8.2×10^6
Brass	104×10^9	38×10^9	113	0.374	4700	2110	3480	40.6×10^6	18.3×10^6
Copper	125×10^9	46×10^9	131	0.37	5010	2270	3750	44.6×10^6	20.2×10^6
Iron	152×10^9	59.9×10^9	69.2	0.27	4994	2809	4480	37.8×10^6	21.35×10^6
Lead	16×10^9	5.4×10^9	33	0.43	1960	690	1210	22.4×10^6	7.85×10^6
Nickel	214×10^9	80×10^9	164	0.336	6040	3000	4900	53.5×10^6	26.6×10^6
Steel	216×10^9	82.9×10^9	100.2	0.276	5941	3251	5250	46.5×10^6	25.4×10^6
Titanium	116×10^9	44×10^9	77.9	0.32	6070	3125	5090	27.3×10^6	14.1×10^6
Pyrex glass	62×10^9	25×10^9	23	0.24	5640	3280	5170	13.1×10^6	7.6×10^6
Lucite	4×10^9	1.43×10^9	5.62	0.4	2680	1100	1840	3.16×10^6	1.3×10^6
Nylon 6-6	3.55×10^9	1.22×10^9	5.11	0.4	2620	1070	1800	2.86×10^6	1.18×10^6
Polyethylene	0.76×10^9	0.26×10^9	2.88	0.458	1950	540	920	1.75×10^6	0.48×10^6

The characteristic impedances for longitudinal and shear wave propagation can be calculated using the following formulae

$$Z_{long} = \sqrt{\rho(\lambda + 2\mu)} \text{ kg/s m}^2 \tag{8.26}$$

$$Z_{shear} = \sqrt{\rho\mu} \text{ kg/s m}^2 \tag{8.27}$$

8.5.4. Characteristic Impedance (Z) and Sound Pressure

As sound propagates through a medium, each small volume element of that medium oscillates about its equilibrium position. For pure harmonic motion, the displacement is

$$y = y_m \cos\frac{2\pi}{\lambda}(x - vt). \tag{8.28}$$

It can be shown that the pressure exerted by the sound wave is

$$p = (k\rho_o v^2 y_m)\sin(kx - \omega.t), \tag{8.29}$$

where k - $2\pi/\lambda$ - The wave number,
ω - Angular frequency.

The amplitude of the wave is the sound pressure $p_m = k\rho_o v^2 y_m$ so a sound wave may be considered to be a pressure wave with a phase displacement of 90° to that of the displacement wave.

The pressure at any point changes sinusoidally with time around a mean value. The RMS (effective) value of this fluctuating component is known as the acoustic pressure, *P*. The acoustic pressure can be measured using an appropriately calibrated microphone or hydrophone.

The characteristic impedance is defined as the ratio of this acoustic pressure and the RMS volume velocity is ξ (this is NOT the wave velocity)

$$Z = \frac{P}{\xi}. \tag{8.30}$$

This is a complex quantity. For an idealised medium (with no loss) Z is real and reduces to a proportionality factor that equates intensity to the square of the sound pressure.

$$Z = \rho.c, \tag{8.31}$$

where: ρ - Density (kg/m^3),
c – Speed (m/s).

For air ρ = 1.3kg/m^2 and c = 332m/s, $Z_{air} \approx 400$ acoustic ohms

8.5.5. Sound Intensity (I)

Intensity of a sound is defined as the power transferred per unit area, and can be expressed through the acoustic impedance (similar to ohms law)

$$I = P\xi = \frac{P^2}{Z}, \tag{8.32}$$

where: I – Sound Intensity (sound power density) (W/m^2),
P – Acoustic Pressure (N/m^2 or Pascal),
ξ - Volume velocity (m^3/s).

It is common to specify the sound not as an intensity, but as a ratio β in dB relative to the minimum audible threshold, $I_o = 10^{-12}$ W/m^2

$$\beta = 10\log_{10}\frac{I}{I_o}. \tag{8.33}$$

Pressure levels can also be expressed in dB as follows

$$\prod = 20\log_{10}\frac{P}{P_o}, \tag{8.34}$$

where $P_o = 2\times10^{-5}$ (N/m^2) = 2.9×10^{-9} (psi) is the Minimum Audible Threshold (MAC)

$$I_o = \frac{P_o^2}{Z} = \frac{\left(2\times10^{-5}\right)^2}{400} = 10^{-12} W/m^2.$$

Table 8.14: Sound levels

Sound	Level (dB)
Stream flow, rustling leaves	15
Watch ticking, soft whisper	20-30
Quiet street noises	40
Normal conversation	45-60
Normal city or freeway traffic	70
Vacuum cleaner	75
Hair dryer	80
Motorcycle, electric shaver	85
Lawn mower, heavy equipment	90
Garbage truck	100
Screaming baby	115
Race car, loud thunder, rock band	120-130
Jet airplane's takeoff from 120 feet	120
Jet airplane's takeoff from 120 feet	120
Pain threshold	130
Rocket launch from 150 feet	180

8.5.6. Attenuation of Acoustic Waves in Air

For a plane acoustic wave, a small attenuation will occur because of the viscosity. The attenuation factor is $e^{-\alpha\eta x}$ for the pressure (or particle velocity).

$$\alpha_\eta = \frac{2\eta\omega^2}{3\rho c^3}, \tag{8.35}$$

where: η - coefficient of viscosity for the gas (dyne-s/cm^2),
ω - Angular frequency (rad/s),
ρ - Density (g/cm^3),
c – Speed of sound (cm/s).

A plane acoustic wave will be attenuated slightly due to thermal conductivity effects.

$$\alpha_t = \frac{k(\gamma-1)\omega^2}{2\gamma\rho C_v c^3}, \tag{8.36}$$

where: k – Thermal conductivity (cal/cm-s-deg),
γ - Ratio of specific heats (defined previously),
C_v – Specific heat at constant volume (cal/g-deg).

In addition to the dispersion of sound due to wind, turbulence in the atmosphere and temperature gradients, two properties combine to attenuate the sound. The first is caused

by molecular absorption due to collisions, and the second due to viscosity and heat conduction described previously.

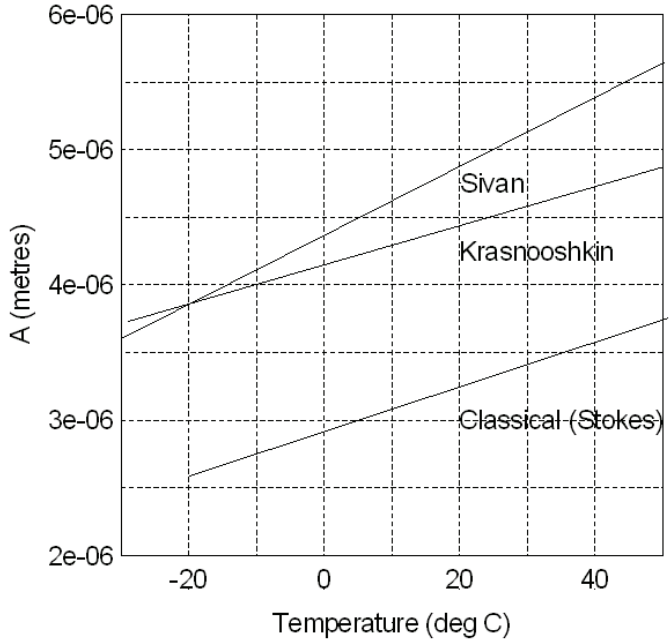


Figure 8.16: Plot of A as a function of temperature $A=\alpha_c\lambda^2/0.0437$, where λ is the wavelength in m and α_c is the attenuation constant in dB/m for a free travelling plane wave. The upper line (Silvan) is closest to the measured values than either of the other two

The combined attenuation is

$$\alpha_c = \alpha_\eta + \alpha_T = 0.0437\frac{A}{\lambda^2}\ \text{dB/m}, \tag{8.37}$$

where λ - Wavelength (m),
A – taken from the curves above.

Total attenuation is very sensitive to humidity and temperature, and can be calculated using the following empirical relationship derived by Knudsen and Harris from measured data (accurate for relative humidity >30%).

$$\alpha_A = \left(\frac{f}{1000}\right)^{3/2}\frac{0.283}{20+\phi_t}\ \text{dB/m}, \tag{8.38}$$

where: f – frequency (Hz),
ϕ_t - $\phi_{20}(1+0.067\Delta t)$,
ϕ_{20} – Relative humidity % at 20°C,
Δt – Denotes the temp difference from 20°C.

The following graphs show the attenuation of sound in air as a function of frequency with relative humidity as a parameter

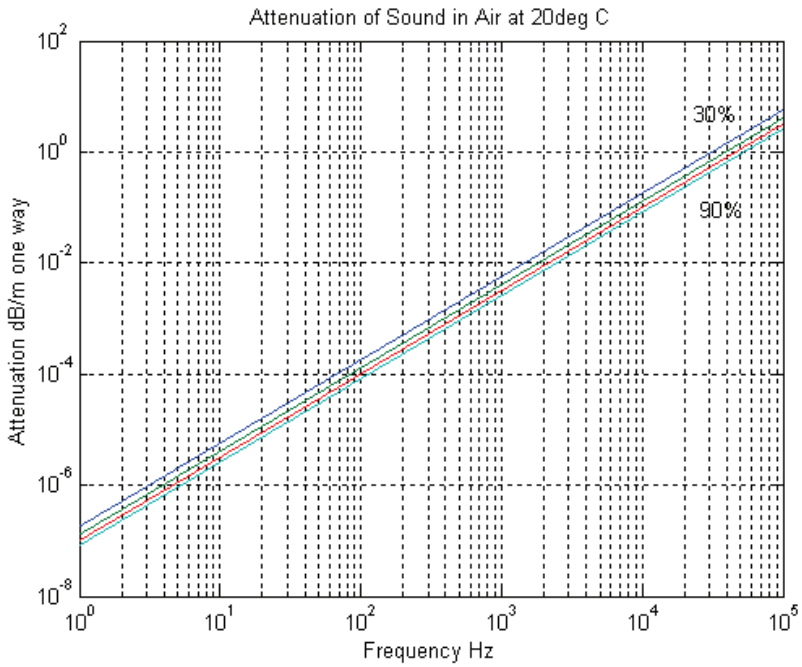


Figure 8.17: Sound attenuation as a function of frequency in air

It can be seen that the effects of humidity are rather small compared to the effects of frequency.

8.6. Attenuation of Acoustic Waves in Water

The absorption of sound as it propagates through pure water has a strong frequency-dependence shown by the following empirical formula,

$$\alpha_{dB} = \frac{0.036 f^2}{f^2 + 3600} + 3.2 \times 10^{-7} f^2 \quad \text{dB/m,} \tag{8.39}$$

where: α_{dB} - Attenuation (dB/m),
f – Frequency (kHz).

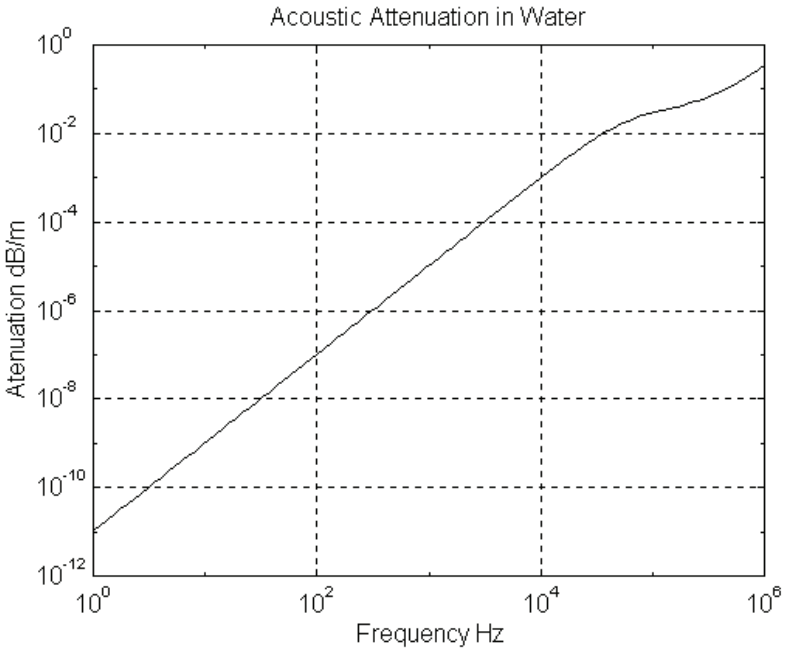


Figure 8.18: Sound attenuation in water

Because of the strong frequency dependence, tables are often constructed to exclude this characteristic. For the tables shown below the attenuation coefficient applies to the following

$$A = A_o e^{-\alpha x} \quad . \tag{8.40}$$

Applying the change of base theorem,

$$20\log_{10}\left(\frac{A}{A_o}\right) = 20\log_{10}\left(e^{-\alpha x}\right) = -8.686\alpha x \, . \tag{8.41}$$

The attenuation coefficients are related as,

$$\alpha_{dB} = 8.686\alpha \ \text{dB/m} \, . \tag{8.42}$$

Table 8.15: Attenuation of sound in water

Temp (°C)	Attenuation (1atm) $10^{15}\alpha/f^2$ (s^2/m)
0	56.9
5	44.1
10	36.1
15	29.6
20	25.3
30	19.1
40	14.6
50	12.0
60	10.2
70	8.7
80	7.9
Pressure (atm)	Attenuation (30°C) $10^{15}\alpha/f^2$ (s^2/m)
0	18.5
500	15.4
1000	12.7
1500	11.1
2000	9.9

Transmission Loss Anomaly (A): Other losses occur due to scattering off suspended particles in the water. These and the losses due to refractive effects are lumped together and called the transmission loss anomaly

8.7. Reflection and Refraction of Acoustic Waves

Reflection and refraction occur where changes in the impedance of the medium occur.

8.7.1. Waves normal to the Interface

If a wave with intensity I_o is propagating through medium-1 with a characteristic impedance of Z_1 through a boundary normal to the direction of travel into medium-2 with characteristic impedance Z_2 then the transmitted and reflected sound intensities are

$$I_{ref} = I_o\left(\frac{|Z_1 - Z_2|}{|Z_1 + Z_2|}\right), \tag{8.43}$$

$$I_{trans} = I_o\left(\frac{|2Z_2|}{|Z_1 + Z_2|}\right), \tag{8.44}$$

where: I_{ref} – Reflected sound intensity,
I_{trans} – Transmitted sound intensity,
I_o – Incident sound intensity.

It can be seen that the reflected intensity increases as a function of the difference in characteristic impedance between the two media.

8.7.2. Waves at an angle to the Interface

When the incident wave strikes the interface at an angle θ_I to the normal, a reflected wave moves away from the interface at an angle $\theta_r = \theta_I$ while the transmitted wave is refracted at an angle θ_t to the normal in the second medium. The angle of refraction is given by Snell's law,

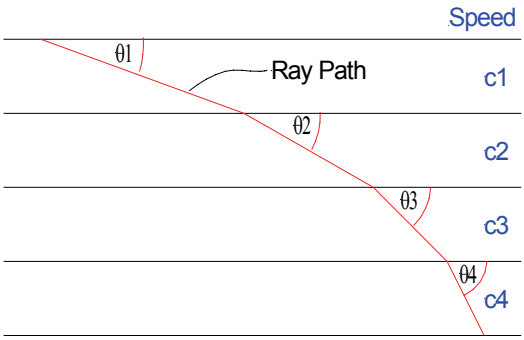
$$\theta_t = A\sin\left(\frac{c_2 \sin\theta_i}{c_1}\right), \tag{8.45}$$

with incident and reflected intensities,

$$I_{ref} = I_o\left(\frac{|Z_1\cos\theta_t - Z_2\cos\theta_i|}{|Z_1\cos\theta_t + Z_2\cos\theta_i|}\right), \tag{8.46}$$

$$I_{trans} = I_o\left(\frac{|2Z_2\cos\theta_i|}{|Z_1 + Z_2|}\right). \tag{8.47}$$

The effects of reflection and refraction are generally visualised using a technique called ray acoustics that is analogous to standard electromagnetic ray-tracing techniques.



If a wave is propagating through n horizontal layers in each of which, the speed is considered to be constant, then Snell's law can be rewritten

$$\frac{c_1}{\cos\phi_1} = \frac{c_2}{\cos\phi_2} = \frac{c_n}{\cos\phi_n}, \tag{8.47}$$

where: ϕ_n – Angle made with the horizontal at that point.

Note than ϕ_n is the complement of the angle usually associated with Snell's law. It is called the grazing angle.

In practise, the temperature does not change abruptly, but rather the gradient will increase or decrease at a measurable rate,

$$c(z) = c_0 + gz, \tag{8.48}$$

where: c_o – speed at the surface (or transducer depth),
g – Speed gradient dc/dz between the surface and depth z.

The net result is that ray traces appear to be curves rather than straight lines with a radius of curvature at any point

$$R = -\frac{c_0}{g} = \frac{c}{g\cos\phi}. \tag{8.49}$$

The actual path followed by sound waves in the sea under different conditions is discussed below.

8.7.3. Refraction Examples

Refraction and Refraction

Sound propagation through water can rely on both reflective and refractive effects as shown in the following figure.

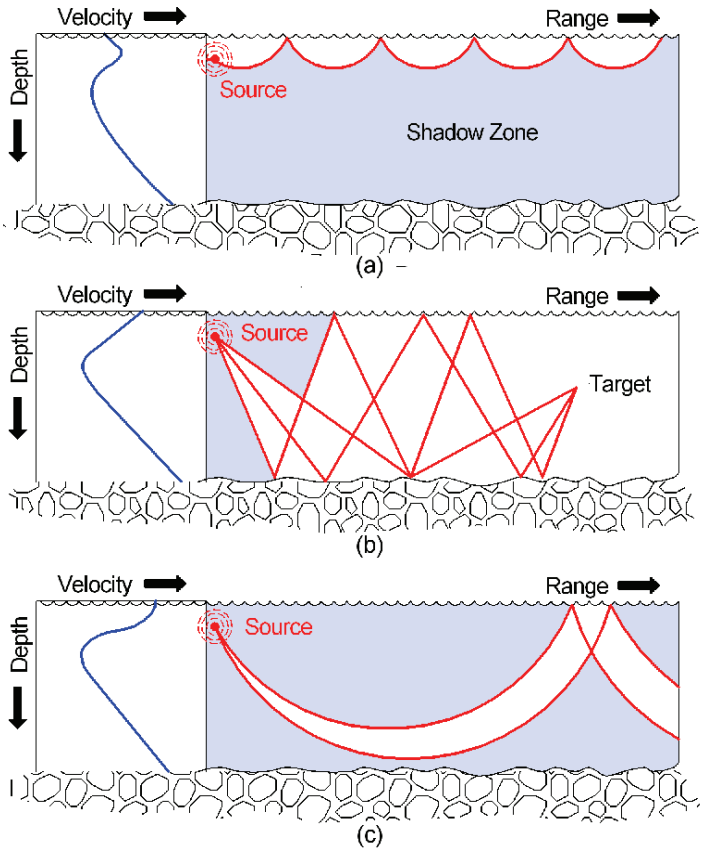


Figure 8.19: Sound propagation effects due to refraction and reflection (a) surface duct, (b) bottom bounce and (c) convergence zone

In general, however, many of the common effects are primarily due to refractive effects as reflection from the surface or the bottom is determined by the characteristics of those surfaces which must be reasonably flat to become effective reflectors of acoustic energy

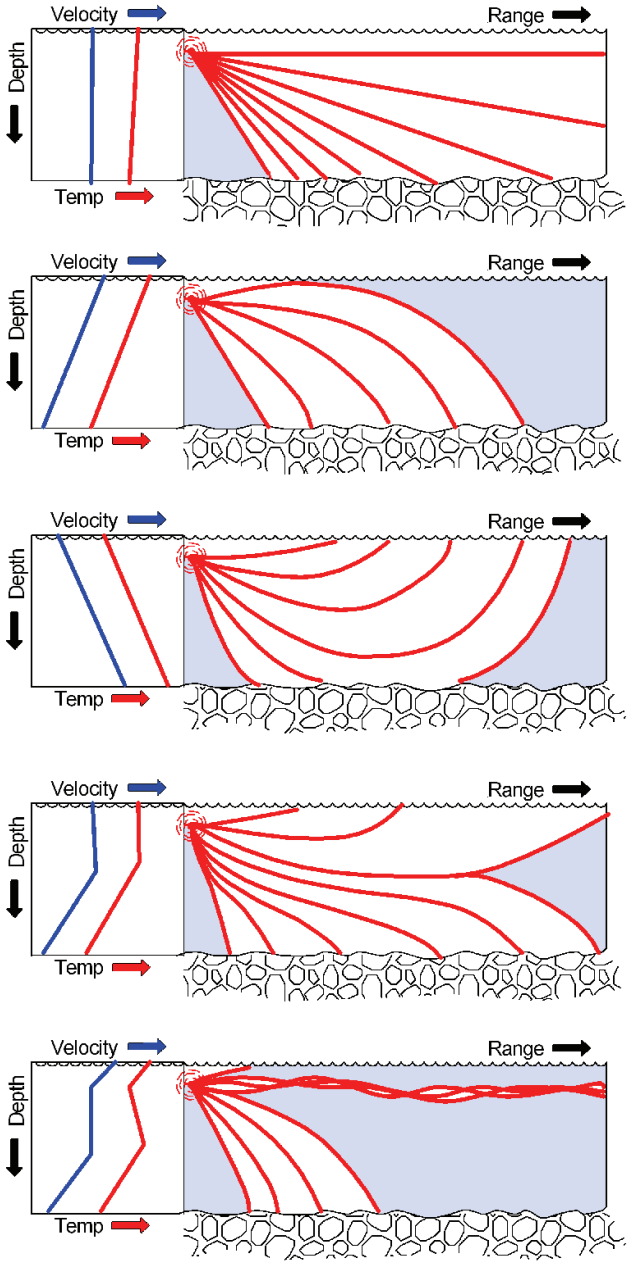


Figure 8.20: Sound propagation effects due to refraction (a) isovelocity - straight line transmission, (b) negative gradient - rays curve downward, (c) positive gradient - rays curve upward, (d) isovelocity over negative gradient - split beam pattern and (e) negative gradient over positive - sound channel

The SOFAR Channel

At depths of approximately 1,000 metres, pressure becomes the important factor where it combines with temperature and salinity to produce a zone of minimum sound speed. This

zone has been named the SOFAR (sound fixing and ranging) channel. If a sound is generated by a point source in the SOFAR zone, it becomes trapped by refraction as shown in the figure. Dispersed horizontally rather than in three directions, the sound is able to travel for great distances

8.8. Multipath Effects

8.8.1. Mechanism

Specular reflection of radar signals from the ground or sonar signals reflecting from the surface of the sea can have a number of effects on target detection probability and tracking accuracy.

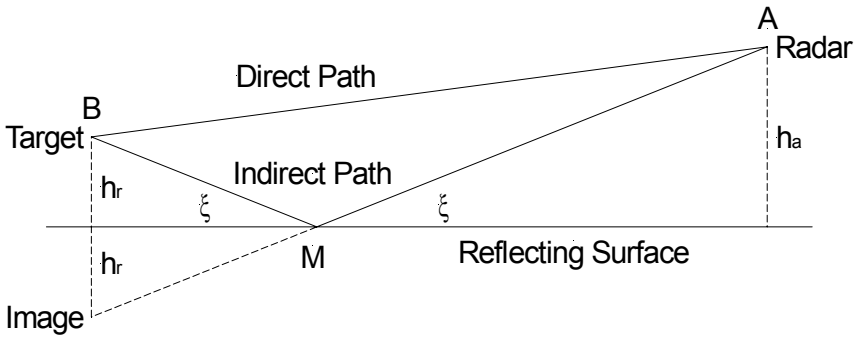


Figure 8.21: Multipath propagation over flat surface

The diagram shows that the energy radiating from the antenna reaches the target via two separate paths. One is the direct line-of-sight path from A to B and the other is the path reflected from the surface of the earth.

The echo re radiated by the target arrives back at the radar along the same two paths and the magnitude of the resultant echo signal will depend on the amplitudes and relative phase differences between the direct and the reflected path.

Modifications to the field strength (volts per metre) at the target caused by the presence of the surface are expressed by the ratio η.

The reflection coefficient form the surface may be considered as a complex quantity

$$\Gamma = \rho e^{j\varphi}, \tag{8.50}$$

where ρ - real part of the reflection coefficient,
φ - Phase shift on reflection.

In this example it is assumed that Γ = -1. The reflected wave suffers no change in amplitude, but its phase is shifted by 180°.

The difference between the reflected path AMB and the direct path AB for $R>>h_a$ is given by

$$\Delta = 2h_a \sin\xi \,. \tag{8.51}$$

For small ξ, sin(ξ) may be replaced by $(h_a+h_t)/R$ to produce

$$\Delta = \frac{2h_a(h_a + h_t)}{R} \,. \tag{8.52}$$

For $h_t >> h_a$

$$\Delta \approx \frac{2h_a h_t}{R} \,. \tag{8.53}$$

The total phase difference (including the 180°) between the direct and reflected signals as they combine at the target is therefore

$$\psi = \psi_d + \psi_r = \frac{2\pi}{\lambda}\frac{2h_a h_t}{R} + \pi \,. \tag{8.54}$$

The vector sum of the two signals with the same amplitude, but a phase difference ψ is $\left[2(1+\cos\psi)\right]^{1/2}$.

The ratio of the direct and the reflected powers at the target is the square of the amplitude

$$\eta^2 = 2\left(1 - \cos\frac{4\pi h_a h_t}{\lambda R}\right) = 4\sin^2\frac{2\pi h_a h_t}{\lambda R} \,. \tag{8.55}$$

By reciprocity, the path from the target back to the radar follows the same two paths, so the power ratio back at the radar is

$$\eta^4 = 16\sin^4\frac{2\pi h_a h_t}{\lambda R} \,. \tag{8.56}$$

The range equation describing the received echo power must be modified by the propagation factor η^4 (often also called F^4).

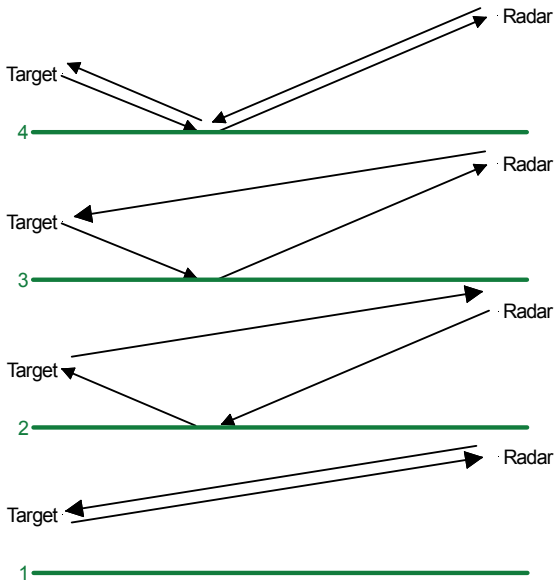


Figure 8.22: Possible signal paths between the radar and the target

8.8.2. Multipath Lobing

Since sine varies from 0 to 1, the factor η^4 varies between 0 and 16, hence, taking into account the 4th power relationship between the range and the echo signal, the radar range varies between 0 and 2 times the range of the same radar in free space.

The field strength is a maximum when the argument in the sine term is equal to π/2, 3π/2,…(2n+1) π/2.

$$\frac{4h_a h_t}{\lambda R} = 2n+1 \quad \text{maxima.} \tag{8.57}$$

$$\frac{4h_a h_t}{\lambda R} = n \quad \text{minima.} \tag{8.58}$$

The presence of a perfectly conducting flat ground plane causes the continuous elevation beam coverage to break up into a lobed structure. Thus if a target is in the lobe, it will be detected at twice the range that it would in free space, but if it is between lobes, then, theoretically, it will not be detected at all.

The first lobe occurs at an angle of $\lambda/4h_a$ (radians)

The curvature of the earth cannot be neglected when predicting radar performance at low elevation angles near the horizon and in this case, the lobing is not a severe as it is for the flat earth approximation because the beam is more divergent, so the minima are not so deep, nor the maxima so large.

Predicting range height coverage is generally managed by computer using packages such as VCCALC.

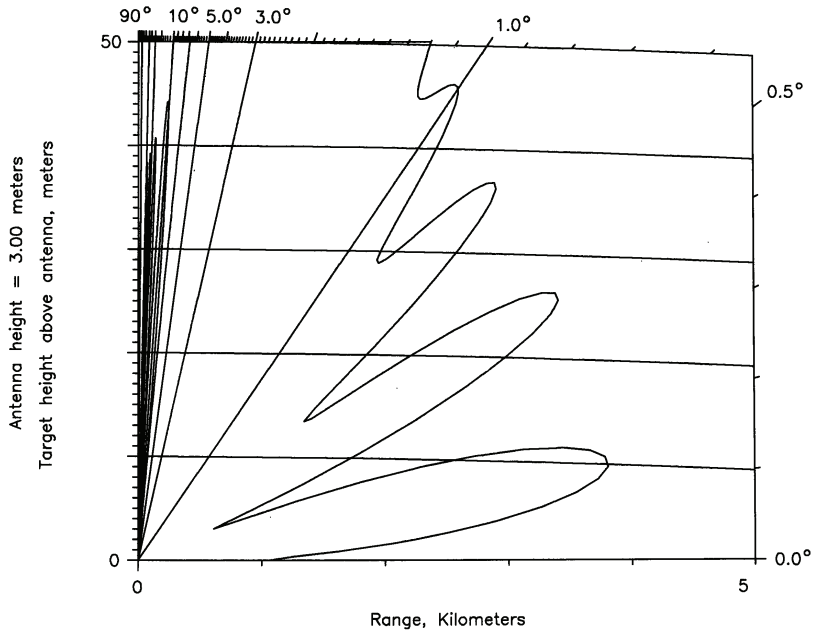


Figure 8.23: Effect of multipath interference on elevation beam pattern

Because the positions of the lobes is a function of both the geometry and the transmitted frequency, a frequency agile radar can be constructed that moves the lobes by altering its frequency. A range of frequencies is selected to obtain uniform coverage over the region of interest.

8.8.3. Multipath Fading

From the graph, it can be seen that an aircraft approaching the radar at constant height of 10,000ft would fly through a number of lobes once it appeared over the horizon at a range of 240km.

In reality, the surface of the earth is never totally flat, so in addition to a specular component, a diffuse component also exists. This diffuse component includes reflections from an elliptical “glistening” surface that extends from the ground below the radar to the ground below the target.

In addition to the possibility of losing the target because of signal fading, multipath has a major effect on the elevation tracking accuracy as discussed in the following section.

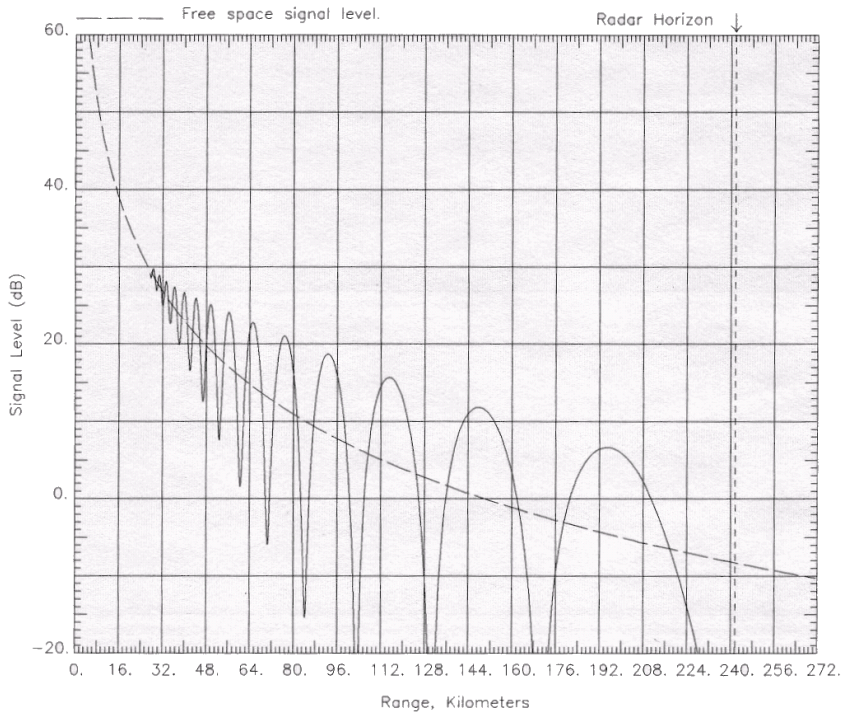


Figure 8.24: Multipath fading as a function of range compared to the free space signal

8.8.4. Multipath Tracking

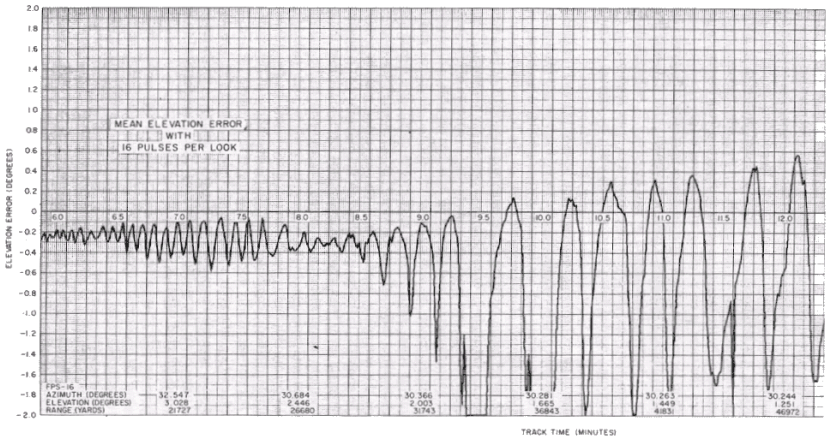


Figure 8.25: Elevation tracking errors caused by multipath interference on a phased array tracking radar with a beamwidth of 2.7°. The aircraft flew out at a constant altitude

The figure shown above, is of a target flying away from the radar at a constant height.

(a) At close range the target elevation angle is large and the antenna beam does not illuminate the surface, so the tracking is smooth.

(b) At intermediate range where the elevation angle varies from as much as 6 down to 0.8 beamwidths, the surface reflected signal enters the radar through the antenna near-in sidelobes. The surface reflected signal is small, so the radar makes small oscillations around some mean position.

(c) At greater ranges where the main beam illuminates the surface, the interference between direct and reflected signals can result in large errors in elevation angle. As with the glint error, these angular excursions can be many times the angular separation between the target and its image.

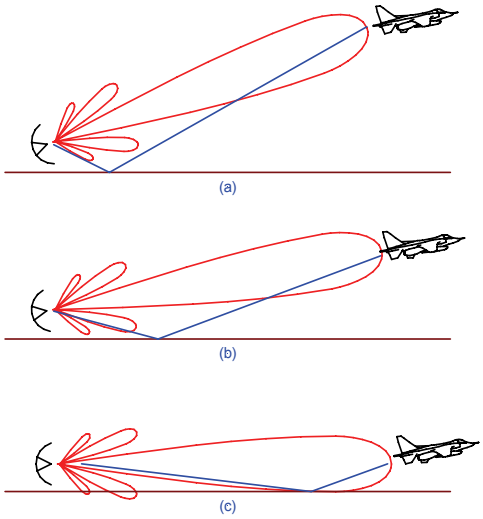


Figure 8.26: Illustration of the three multipath regions

Particularly over smooth water, the elevation tracking errors can generate such severe nodding in the antenna that it can cause the radar to break lock.

Fixes to the Tracking Problem

The surest method to avoid multipath induced tracking errors is to use an antenna with a sufficiently narrow beamwidth that it does not illuminate the surface. This requires, either a large antenna or a high operating frequency.

A second method is to fix the antenna elevation (once the target has reached the 0.8 beamwidths region), and to continue with closed loop tracking.

For very short pulses it is possible to separate the direct return and the delayed multipath return. However, this is often not practical, as the range resolution required to separate the signals is

$$\Delta R = \frac{2h_a h_t}{R}. \qquad (8.59)$$

For a radar height of 30m, a target height of 100m, and a range of 10km, the range resolution must be 0.6m which corresponds to a pulse width of 4ns.

Finally, frequency agility can also be effective in reducing multipath error, as a change in frequency changes the phase relationship between the direct and the reflected signals.

It turns out that the bandwidth required to achieve this fix is equivalent to that required for the short pulse option.

8.8.5. Effects on Imaging

For short range imaging applications which are typical for robotic guidance and collision avoidance, the primary effect of multipath is to displace targets from their correct bearing and elevation.

In the elevation plane the grazing angles are very shallow and specular reflection occurs. This results in targets appearing below the ground surface which could lead the robotic system to conclude that the gradient has changed, or more catastrophically, to allow it to drive into a dam under the misguided assumption that it has identified some trees (albeit upside down) in that direction.

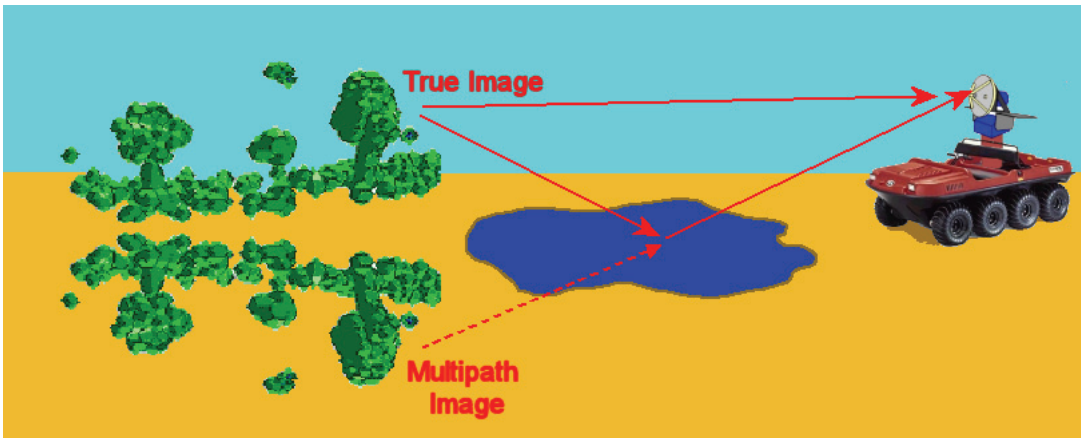


Figure 8.27: Effect of multipath on the interpretation of the ground level

Multipath in azimuth generally caused by the flat walls of buildings has the similar effect of displacing, and doubling up of the number of features visible to the radar. This can result in severe navigation problems.

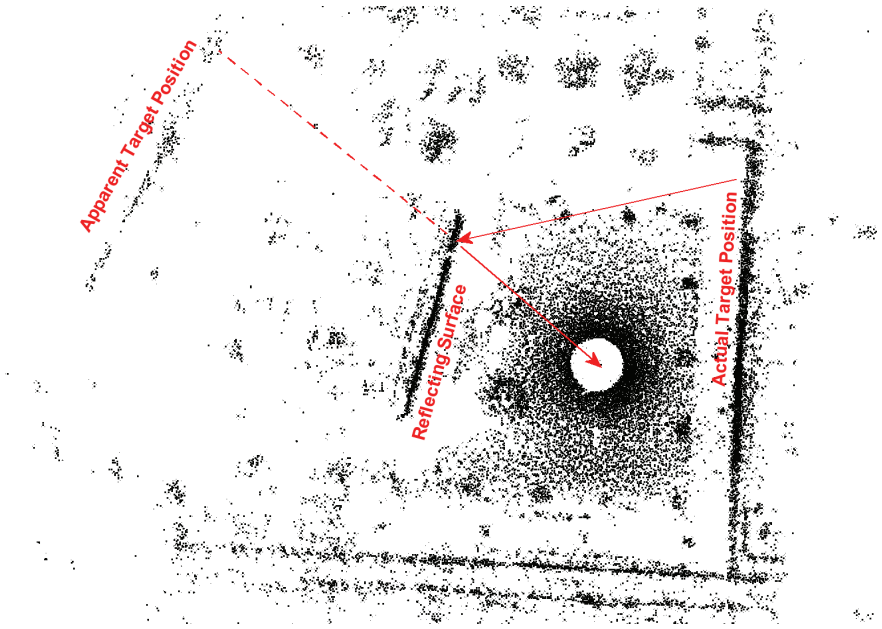


Figure 8.28: Effects of multipath on the apparent angular position of features

8.9. References

[1] E.Brookner, *Radar Technology,* Artech, 1980
[2] P.Bhartia & I.Bahl, *Millimeter Wave Engineering and Applications,* Wiley-Interscience, 1984.
[3] C.Currie (ed), *Principles and Applications of Millimeter Wave Radar,* Artech, 1987.
[4] Richard Cook (editor), *Acoustics*, National Bureau of Standards, ?
[5] L.Kinsler & A. Frey, *Fundamentals of Acoustics 2nd ed.,* John Wiley & Sons
[6] M.Skolnik, *Introduction to Radar Systems,* McGraw Hill, 1980.
[7] D.Barton, Modern Radar Systems Analysis, .Artech, 1988.
[8] J.Goldhirsh, "Attenuation and Backscatter from a Derived Two-Dimensional Duststorm Model", IEEE Transactions on Antennas and Propagation, Vol. 49, No. 12, December 2001

Chapter 9.

Target and Clutter Characteristics

9.1. Target Cross Section

The "cross section" qualitatively relates the amount of power that strikes the target to the amount of power that is reflected into the receiver. Assuming that the power density of a plane wave incident on the target is S_i W/m^2, and the amount of power scattered isotropically is P_r which is defined in terms of the cross section, σ, as follows

$$P_r = \sigma . S_i \,. \tag{9.1}$$

Then the power density S_r of the scattered wave at the receiving antenna is

$$S_r = \frac{P_r}{4\pi R^2} \,. \tag{9.2}$$

This allows the cross section to be defined in terms of the ratio of the power density at the receiver to that incident on the target

$$\sigma = 4\pi R^2 \frac{S_r}{S_i} \,. \tag{9.3}$$

And in order to ensure that the receiving antenna is in the far field and that the waves are planar

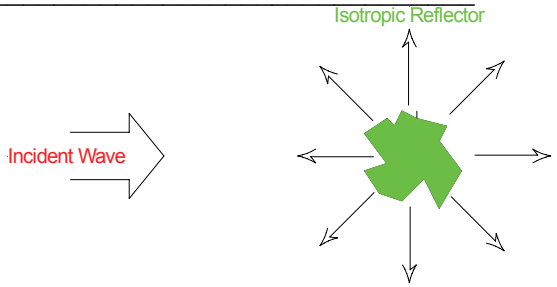
$$\sigma = \lim_{R \to \infty} 4\pi R^2 \left| \frac{S_r}{S_i} \right| . \tag{9.4}$$

In terms of the various fields and pressures that make up EM and acoustic waves, this equation can be expanded as follows,

$$\sigma = \lim_{R\to\infty} 4\pi R^2 \left|\frac{E_r^2}{E_i^2}\right|$$

$$\sigma = \lim_{R\to\infty} 4\pi R^2 \left|\frac{H_r^2}{H_i^2}\right| \qquad (9.5)$$

$$\sigma = \lim_{R\to\infty} 4\pi R^2 \left|\frac{p_r^2}{P_i^2}\right|$$



where E_r, E_I – Electric field magnitude at the receiver and incident on the target,
H – Magnetic field equivalent,
P – Acoustic Pressure equivalent.

If a target were to scatter power uniformly over all angles, its cross section would be equal to the area from which power was extracted from the incident wave. Since the sphere has the ability to scatter isotropically, it is convenient to interpret cross section in terms of the projected area of an equivalent sphere.

9.1.1. Cross Section and the Equivalent Sphere

A sphere with radius $a >> \lambda$ will intercept power contained in πa^2 of the incident wave. It will scatter power uniformly over 4π steradians of solid angle. The cross section of a sphere is therefore (by definition) equal to the projected area πa^2 even though the actual area that returns power to the receiver is a very small area where the surface lies parallel to the incident wavefront.

The cross section of a sphere is therefore equal to one quarter of its surface area, and by extension it can be shown that *the average cross section of any large object that consists of continuous curved surfaces will be equal to one quarter of its total surface area*.

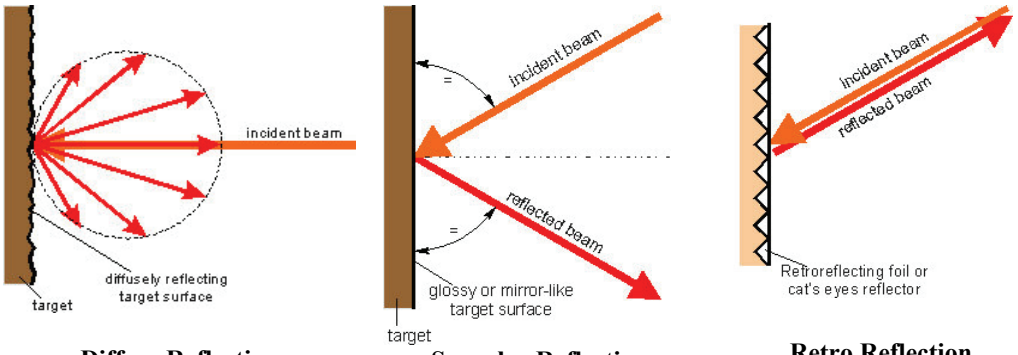
9.1.2. Cross Section of Real Targets

Unfortunately, very few real targets are spherical and the cross sections of complex targets are complicated functions of the viewing aspect, frequency and polarisation. Sophisticated modelling software can be used to obtain estimates of target cross sections, but the most accurate method to determining this is by measurement. However, even this method is fraught with difficulty, particularly if the target is large as it is often impractical to measure the cross section over all aspect angles in azimuth and elevation.



Target cross section is often related to the physical size of the item, but under certain circumstances it may be much larger or even much smaller. For example a corner reflector has an extremely large target cross section in relation to its size, whereas a B2-B stealth bomber has a very small cross section.

Because complex targets are made up from many scattering surfaces the cross section will also be made up of reflections from a large number of scatterers,. This means that even very small changes in the aspect angle of the target will result in relative phase changes between the scatterers and an altered cross section. The effective surface roughness of a target (as a function of λ) also plays an important roll in determining its cross section. There are three mechanisms that determine individually, or in combination, the target reflection characteristics.



Diffuse Reflection **Specular Reflection** **Retro Reflection**

Figure 9.1: Different modes of reflection depend on the surface characteristics of the target

There are a number of significant differences between the cross sections of targets measured using EM waves and those measured using acoustic waves, though the underlying theory of reflection is the same in the two cases

9.2. Radar Cross Sections (RCS)

Qualitatively, the RCS of an object is a measure of its "size" as seen at a particular radar wavelength and polarisation. RCS has units of m^2 and is often expressed in decibels relative to a square meter (dBm^2)

$$\sigma(dBm^2) = 10\log_{10}\left[\sigma(m^2)\right]. \tag{9.6}$$

To account for the polarisation dependency of RCS, the relationship between the transmitted and received electric fields must be considered in terms of their orthogonal linear polarisation components E_H and E_V. and the proportionality constants that relate them.

$$\begin{bmatrix} E_H^r \\ E_V^r \end{bmatrix} = \begin{bmatrix} a_{HH} & a_{VH} \\ a_{HV} & a_{VV} \end{bmatrix} \begin{bmatrix} E_H^t \\ E_V^t \end{bmatrix}. \tag{9.7}$$

This matrix of constants is referred to as the target scattering matrix and it can be used to justify the definition of a radar cross section scattering matrix with the same form.

$$\sigma = \begin{bmatrix} \sigma_{HH} & \sigma_{VH} \\ \sigma_{HV} & \sigma_{VV} \end{bmatrix}. \tag{9.8}$$

The relationship between the target and RCS scattering matrices is as follows (for each of the four terms)

$$a_{HH} = \sqrt{\sigma_{HH}}\, e^{j\phi_{HH}}\,, \tag{9.9}$$

where $\sqrt{\sigma_{ij}}$ - magnitude of each matrix component,

ϕ_{ij} - Phase of the associated element.

From hereon the target RCS will be considered to have a single value that corresponds to either σ_{HH} or σ_{VV} depending on the transmitted polarisation.

9.3. RCS of Simple Shapes

A number of simple reflectors are useful, particularly for use as calibration targets. The flat plate offers the highest RCS for a given size, but it is specular which makes it difficult to align. An alternative is the trihedral corner-reflector which also produces a high RCS which remains reasonably constant over a wide angle. This is the preferred RCS reference.

The sphere is also often used as a reference for moving targets as its RCS is invariant with observed angle. Unfortunately the RCS is very small for its size and also large conductive perfect spheres are difficult to manufacture.

9.3.1. Flat Plate

The flat plate has the largest peak RCS for its size of any target, It is specular and so the RCS falls off sharply with changes in incidence angle from 0°. The length, *a*, of the plate determines the width of the specular lobe and the interval between peaks as defined by

$$\sigma(\theta) = \sigma_o \left[\frac{\sin\left(\frac{2\pi}{\lambda} a \sin\theta\right)}{\frac{2\pi}{\lambda} a \sin\theta} \right]^2 \cos^2\theta \tag{9.10}$$

$$\sigma_o = \frac{4\pi a^2 b^2}{\lambda^2} \tag{9.11}$$

where $\sigma(\theta)$ – RCS as a function of angle (m^2),

a,b – Sides of the plate (m),

θ - Plate rotation around M axis (rad).

At incidence angles close to 90°, complex edge diffraction effects become dominant and small irregularities in the surface destroy the symmetry of the data in the following figure.

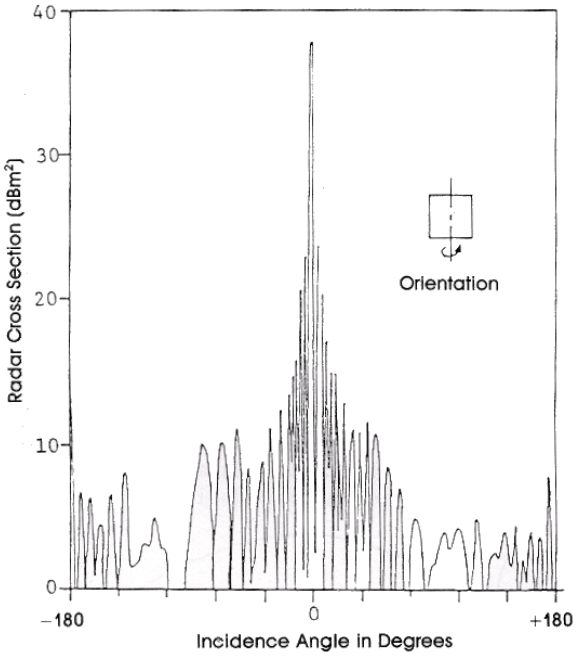


Figure 9.2: Measured RCS of a flat square plate as a function of orientation

9.3.2. The Sphere

A perfectly conducting sphere is the simplest shape whose RCS can be determined exactly. With its 3D symmetry its RCS is aspect independent, so is often used as a calibration target if the size is selected appropriate to the operating frequency of the measurement system.

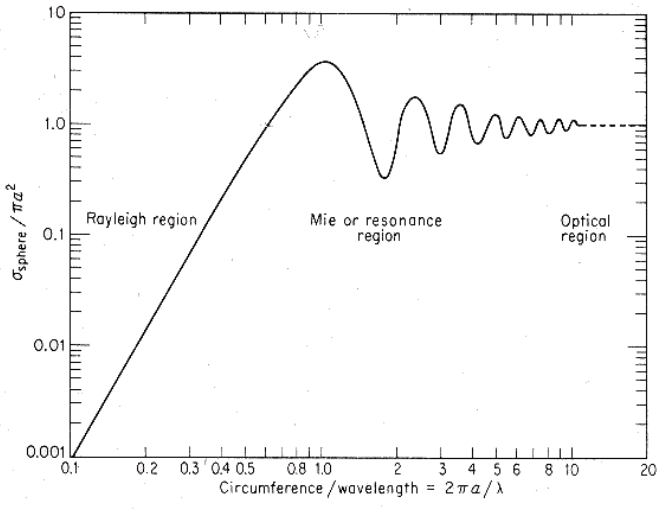


Figure 9.3: RCS of a sphere as a function of the circumference normalised by the wavelength of the incident radiation.

Note that there are three regions that are apparent from this figure:

- **Rayleigh Region ($2\pi a/\lambda < 1$):** The RCS is inversely proportional to the 4th power of the wavelength as indicated by the slope of the curve in that region.
- **Mie Region ($1 < 2\pi a/\lambda < 10$):** In this resonance region, a creeping wave travels around the sphere and back towards the receiver where it interferes constructively or destructively with the specular backscatter to produce a cyclical variation in the RCS
- **Optical Region ($2\pi a/\lambda > 10$):** The RCS of the sphere approaches its geometric projected area πa^2.

9.3.3. Trihedral Reflector

Trihedral corner reflectors have large non-specular radar cross sections and so make ideal as a test targets as alignment is not critical. Their behaviour for linear polarisations is good, but they cannot be used for circularly polarised measurements as the triple bounce reverses the sense of the polarisation. For a symmetrical reflector having each vertex with length a, the peak radar cross section is determined by

$$\sigma = \frac{4\pi a^4}{3\lambda^2}. \tag{9.12}$$

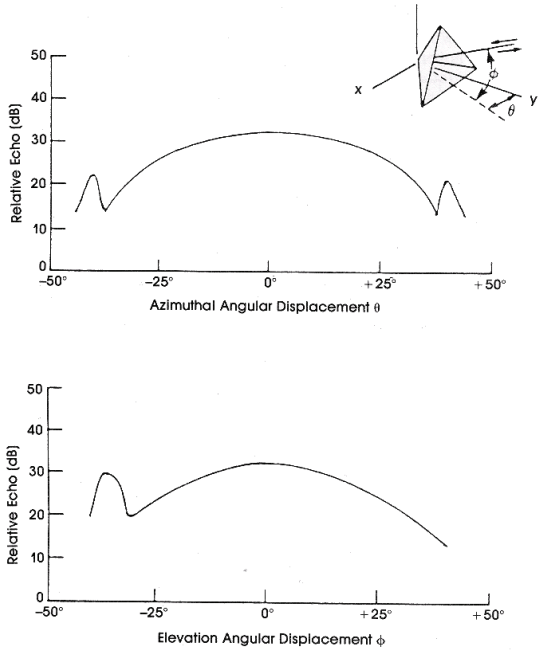
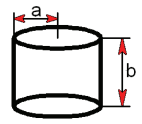
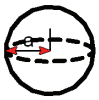
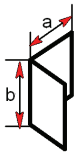
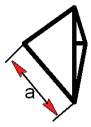
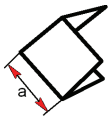
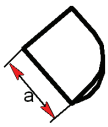
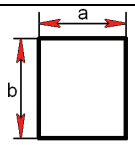
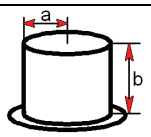
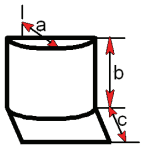


Figure 9.4: RCS of a trihedral reflector as a function of orientation

From the figure it can be seen that azimuth and elevation misalignments of up to about 10° can be tolerated without a significant reduction in RCS

9.3.4. Other simple reflectors

Table 9.1: Maximum RCS for typical calibration targets

Target		Maximum Cross Section	Advantages	Disadvantages
	Cylinder	$\sigma = \frac{2\pi ab^2}{\lambda}$	Nonspecular along the radial axis	Low RCS for size, specular along axis
	Sphere	$\sigma = \pi a^2$	Nonspecular	Lowest RCS for size, radiates isotropically
	Diplane	$\sigma = \frac{8\pi a^2 b^2}{\lambda^2}$	Large RCS for size, nonspecular along one axis. Useful for testing polarisation	Specular along one axis
	Triangular Trihedral	$\sigma = \frac{4\pi a^4}{3\lambda^2}$	Nonspecular	Cannot be used for cross polarised measurements
	Square Trihedral	$\sigma = \frac{12\pi a^4}{\lambda^2}$	Large RCS for size, nonspecular	Cannot be used for cross polarised measurements
	Circular Trihedral	$\sigma = \frac{0.507\pi^3 a^4}{\lambda^2}$	Large RCS for size, nonspecular	Cannot be used for cross polarised measurements
	Flat Rectangular Plate	$\sigma = \frac{4\pi a^2 b^2}{\lambda^2}$	Largest RCS for size	Specular along both axes, difficult to align
	Top Hat	$\sigma = \frac{2\pi ab^2}{\lambda \cos^3 \phi}$ ϕ is the elevation angle to the cylinder, ϕ=0° is perpendicular to the cylinder c>b to be effective for 90° rotation of polarisation ϕ = 45°	Low RCS for size	Difficult to align rotated seam
	Bruderhedral		Large RCS, easier to align for rotation than Top Hat	Moderately specular along one axis

9.4. RCS of Complex Targets

9.4.1. Aircraft

The RCS of full size aircraft can be measured in an outdoor range by placing the aircraft on a pedestal mounted on a turntable. To reduce the mass of such targets they are often gutted. However this can introduce inaccuracies in the results as jet engines and cockpit instrumentation often add a significant contribution to the overall RCS.

The following figure shows the now classic radar cross section of a B-26 bomber plotted on polar axes as a function of azimuth angle. Note that the RCS exceeds 35dBm^2 (3100m^2) from certain aspect angles. In contrast the RCS of the B2 stealth bomber is about –40dBm^2.

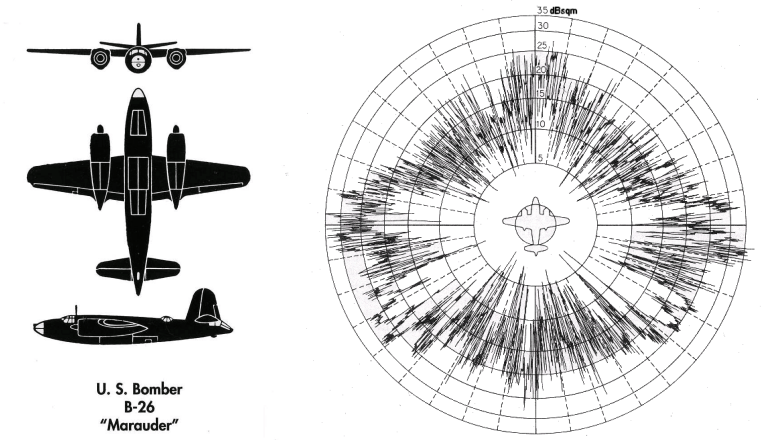


Figure 9.5: Diagram and radar cross section of a B-26 bomber as a function of aspect angle

A more modern aircraft, the C29 cargo plane shows a more dramatic variation in the RCS with strong peaks at +/-90°, generated by the aircraft fuselage. From the tail-on aspect, this aircraft also shows a reasonably large RCS that may be generated by the engine outlets.

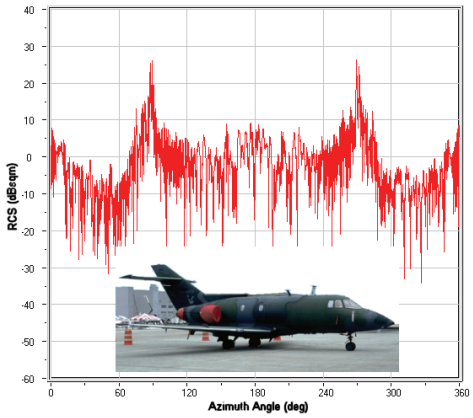
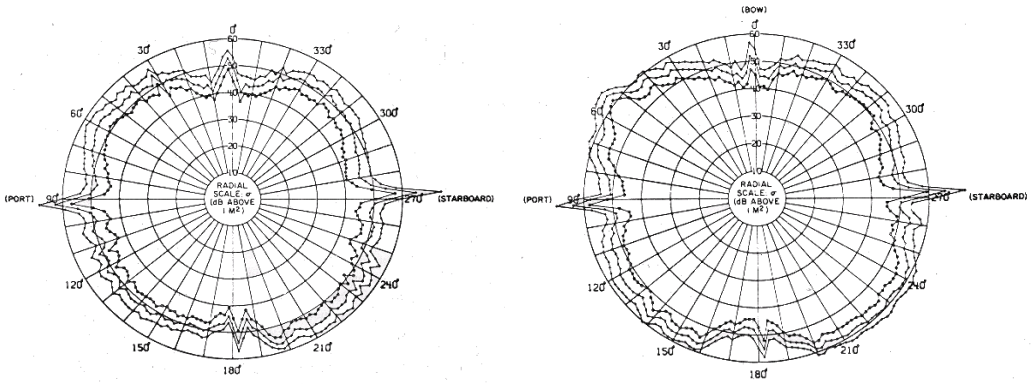


Figure 9.6: RCS of a C29 cargo plane with aspect angle

9.4.2. Ships



2.8GHz **9.225GHz**

Figure 9.7: RCS of a naval auxiliary ship with aspect angle

Measuring the RCS of a ship involves sailing in a circle while measuring the return from a fixed point and then compensating for variations in range as it is not practical to mount ships on turntables.

The median RCS of a typical ship at low grazing angles (excluding the specular broadside return) is related to its size by the empirical formula

$$\sigma = 52 f^{\frac{1}{2}} D^{\frac{3}{2}} \quad \text{m}^2, \tag{9.13}$$

where: f – frequency (MHz),
D – displacement (kilotons).

At higher grazing angles, the median RCS is equal to the displacement tonnage expressed in m^2.

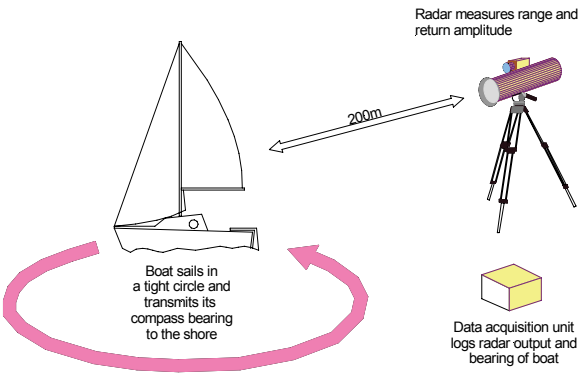


Figure 9.8: Polar RCS measurement setup

The empirical relationship described in (9.13) is not considered to be accurate at high frequencies due to the quadratic increase in the RCS of simple targets (flat plates and corners) with increasing frequency. We have measured the RCS of a number of small boats at 94GHz from a shore-mounted radar that tracked the craft in angles as it sailed slowly in a tight circle at an appropriate range.

The radar measured and logged the returns from the target, while the boat angle measured by an electronic compass on board, was communicated back to the shore where it was also logged. The data was then used to produce a polar RCS measurement such as the one shown in the following figure.

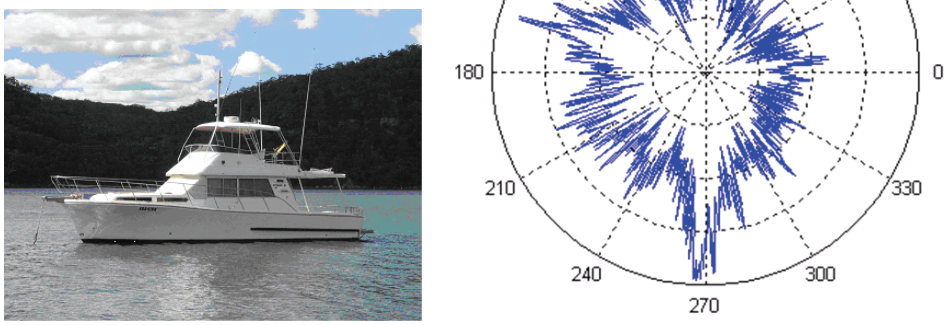


Figure 9.9: RCS of a Steber42 flybridge cruiser at 94GHz at a 2.5° grazing angle

It is interesting to note that the median RCS of this boat is about 10dBm^2, and the RCS calculated using (9.13) for a 13 ton displacement and a frequency of 94GHz comes out at 13dBm^2

9.4.3. Ground Vehicles

Measurement techniques for ground vehicles generally involve measuring the RCS through 360° in azimuth by rotating the vehicle on a turntable. These are normally made from a reasonably short range so that it is possible to produce plots at various grazing angles by adjusting the height of the measurement radar.



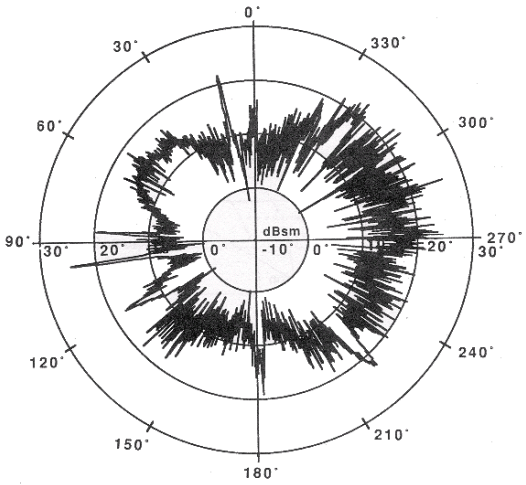
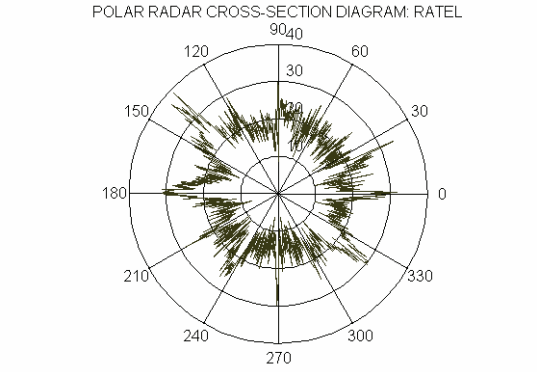
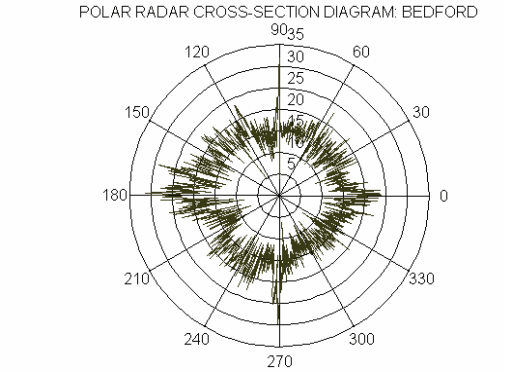


Figure 9.10: RCS of a Toyota utility vehicle at 35GHz from a 0° grazing angle

For the measurements made by us shown below, a turntable was not available, so the vehicle was driven slowly in a tight circle while the radar measurement system was aimed manually.







Armoured Personnel Carrier



Bedford Truck

Figure 9.11: RCS of larger military vehicles at 94GHz

The RCS of military vehicles is often larger than that of military aircraft because the latter are generally more rounded for aerodynamic reasons,. Ground vehicles are often made up of flat armour plates and lots of brackets, antennas etc.

9.5. RCS of Living Creatures

9.5.1. Human Beings

The RCS of a human being is a function of frequency as shown in the following table. Variations are with aspect angle and polarisation.

Table 9.2: RCS of a human being at different frequencies

Frequency (MHz)	RCS (m^2)
410	0.033 – 2.33
1120	0.098 – 0.997
2890	0.140 – 1.05
4800	0.368 – 1.88
7375	0.495 – 1.22

Some of our measurements of the human torso with a spot size of 1.3m made at 94GHz are reproduced below. They show a lower value for the RCS than those predicted from the table.

The measurements include a reference $1m^2$ reflector in the beam about a metre off the ground that is eclipsed as the human subject walks into the beam directly in front of it.



Figure 9.12: Measuring RCS of human beings

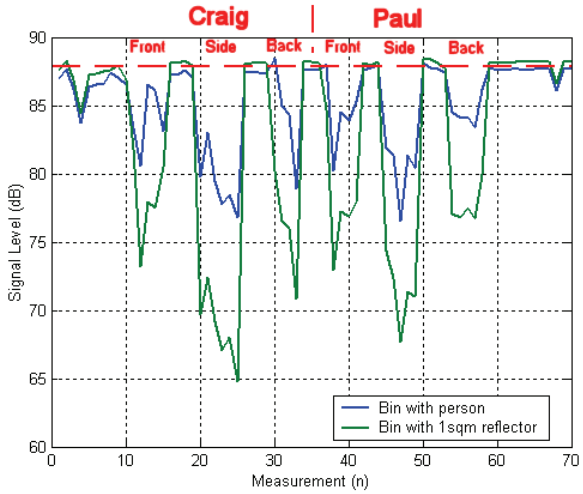


Figure 9.13: Human RCS at 94GHz is determined by having a person block the path from the radar to the 0dBsqm corner reflector. The RCS is determined by the decrease in the signal level

It can be seen from these measurements that the RCS of the two human beings varies between 3dB and 8dB lower than the 1sqm (0dBsqm) reflector at 94GHz.

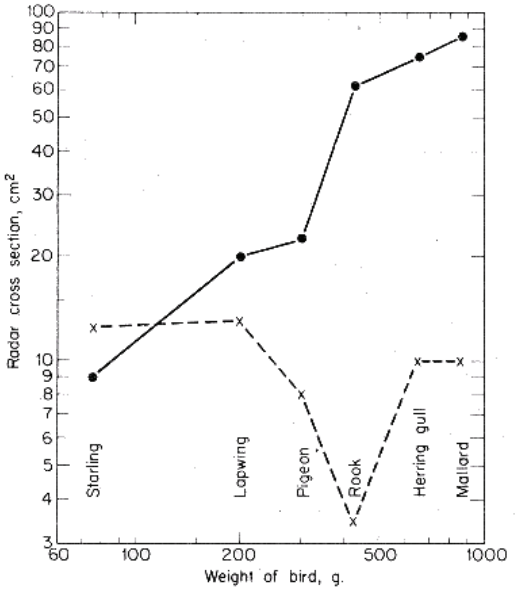
9.5.2. Birds

Radar returns are often returned from areas that appear clear, these are called ghosts or angles. They are often returns from flocks of birds or swarms of insects. Because birds can fly at up to 50knots their returns are not rejected by Doppler or MTI (Moving Target Indicator) processing.

Table 9.3: RCS of birds as a function of frequency

Bird Type	Frequency	Mean RCS (cm^2)	Median RCS (cm^2)
Grackle	**X**	16	6.9
	S	25	12
	UHF	0.57	0.45
Sparrow	**X**	1.6	0.8
	S	14	11
	UHF	0.02	0.02
Pigeon	**X**	15	6.4
	S	80	32
	UHF	11	8.0

Note that resonance effects play a large role in the measured RCS of the birds shown here. This is verified in the following graph that relates bird mass to RCS.



Radar cross section of birds (with closed wings) at *S* band. Vertical polarization. Solid circles apply to average values $\pm 20°$ around broadside, x's apply to the average of the $\pm 20°$ sector about the head and the $\pm 20°$ sector about the tail. (*After Houghton and Smith.*[93])

Figure 9.14: Bird RCS at 3GHz

Fluctuations in the RCS of a single bird in flight have been measured to have a log-normal distribution and an empirical formula that relates the wing beat frequency, f (Hz) of birds to their length, l (mm) is $f.l^{0.827} = 572$.

9.5.3. Insects

Appreciable echoes are only obtained from insects if their body length exceeds $\lambda/3$. Insects viewed broadside have RCS values between 10 and 1000 times larger than when viewed head on.

At X-band the RCS of a variety of insects showed a variation from 0.02 to 9.6cm^2 with longitudinal polarisation and between 0.01 and 0.95cm^2 for transverse.

A bee would have a broadside RCS of about 1cm^2 at X-band. This would not increase significantly up to W-band as the bee moved from the Mie to the Optical region

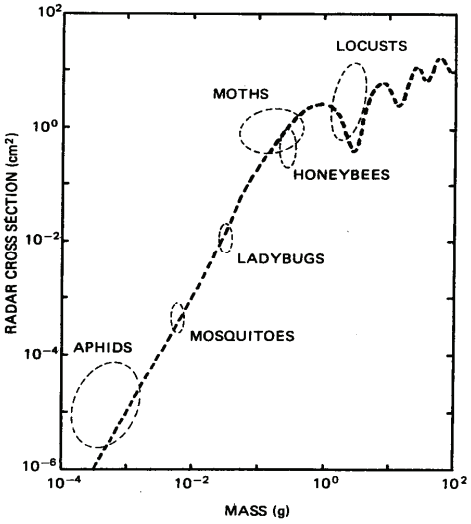


Figure 9.15: RCS of insects at 9.4GHz.

Line shows RCS of a water droplet

9.6. Fluctuations in RCS

9.6.1. Temporal Fluctuations

If either the radar or the target is moving, variations in the aspect angle result in fluctuations in the RCS. These fluctuations have a major effect on the probability of detection of a target as discussed in Chapter 10.

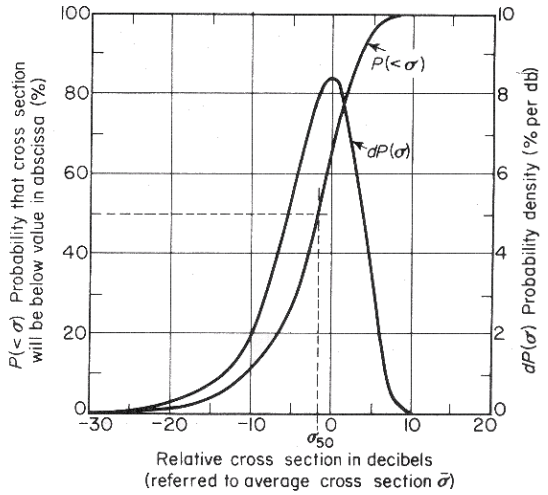


Figure 9.16: Log plot of the Rayleigh distribution

The RCS of a target, such as the Toyota Utility Vehicle described above, can be displayed as a probability density function based on the full 360° azimuth coverage, or a smaller sector if required. Though these are in effect spatial variations in RCS, if the vehicle is moving and showing a changing aspect to the radar, then they map into temporal variations.

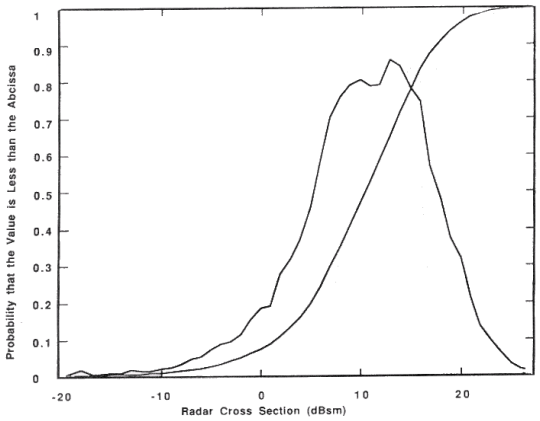


Figure 9.17: PDF and cumulative probability distributions of the RCS of a Toyota at 35GHz

A single value for RCS is often used to characterise a target. The mean or median are the most common values selected for this purpose.

9.6.2. Spatial Distribution of Cross Section

The various reflecting bodies across a target interfere constructively and destructively as it moves. This results in a physical displacement of the effective target position that varies with time. The angular position of the target measured by a radar system is determined by the direction-of-arrival of the phase front of the reflected radar signal from all scatterers, and so it is possible that the angular error can extend beyond the physical boundaries of the target. These variations in direction of arrival are known as glint.

The graph below shows a sample of the angular tracking error measured for an aircraft flying towards the radar.

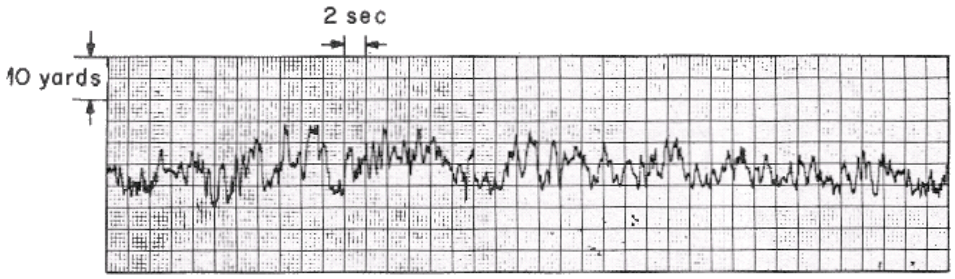


Figure 9.18: Displacement of the tracked centroid of an aircraft echo with time illustrates a phenomenon known as glint

Assuming that the aircraft path is undisturbed, and can be approximated by the mean value of this signal, then if the extent of the tracking error is compared to the physical extent of the aircraft, it is found that the PDF extends past the wingtips as shown below.

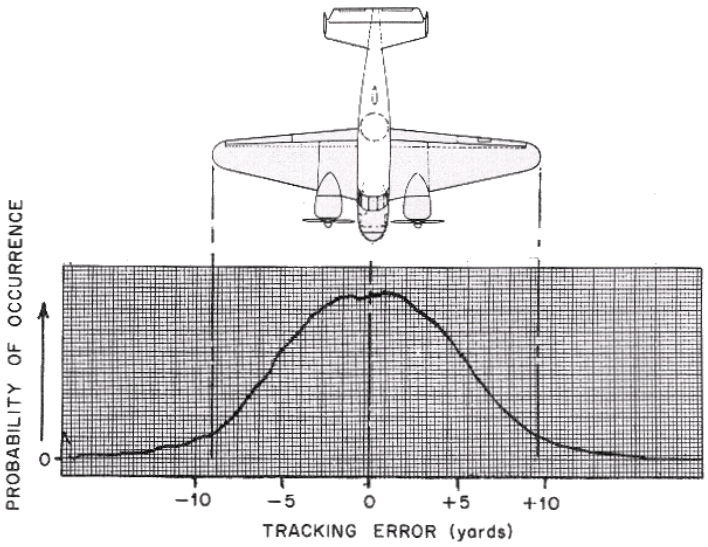


Figure 9.19: Probability distribution function of the angle tracking error

Measured glint spectra show that a significant portion of the return occurs at frequencies below 3Hz. This is within the bandwidth of most tracking filters, and so glint can cause serious tracking errors if the appropriate precautions are not taken (see Chapter 13).

9.7. Radar Stealth

Stealth in this context is the ability of an aircraft to evade radar, however that is only one of several factors that must be considered to keep aircraft undetected. The others can be almost as important and include the following:

- Sound
- Sight
- Heat
- Leaking electronic signals

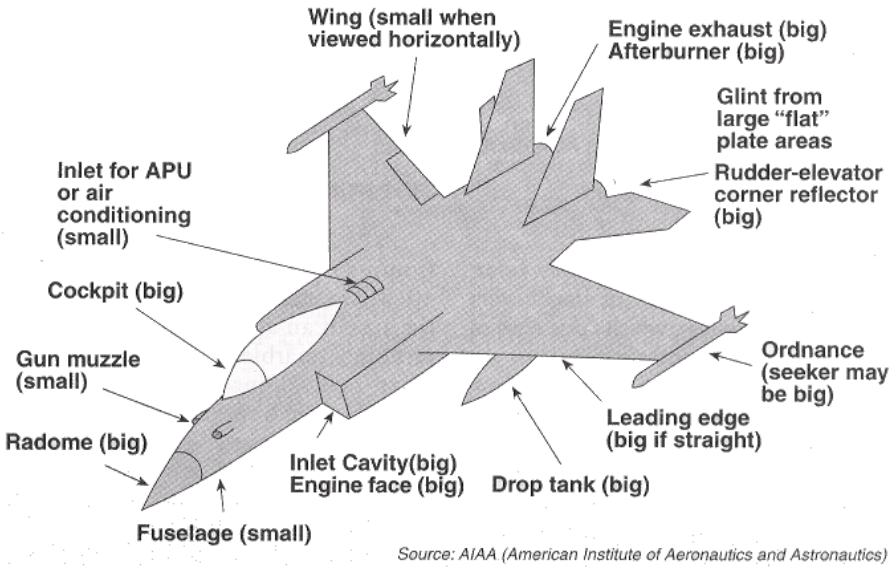


Figure 9.20: Major contributors to radar cross section of a jet fighter

The basics of stealth have been known since the 1950's and the computational power required to design stealth aircraft has been available since the 1970's. These requirements dictate the shape of modern stealthy aircraft are summarised in the figure below.

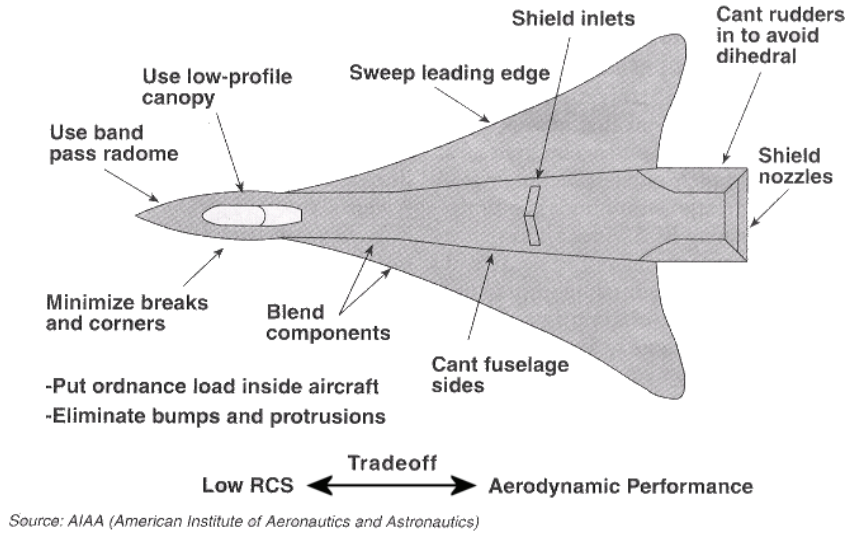


Figure 9.21: Designing aircraft for stealth

The first stealth aircraft relied on faceting to reflect power away from its source. This is effective against monostatic systems, but it is not effective against bistatic, or multistatic systems where the transmitter and receiver are spatially separated.

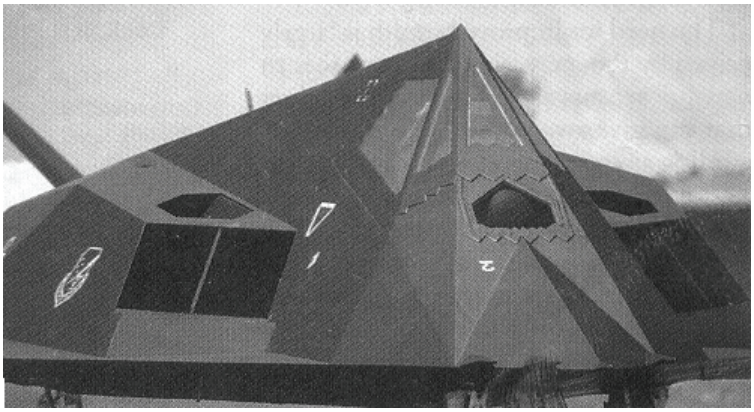


Figure 9.22: F117 faceted stealth design

In addition to designing the aircraft shape to minimise returns, the material from which the aircraft is also important. Multi-layered composite structures are constructed to match the impedance to free space (ensuring that very little energy is reflected), and then progressively absorbing it using materials that are loaded with resistive carbon. This is known as radar absorbing material (RAM).

Windows are covered with conductive materials to minimise penetration into the cockpit, and subsequent reflection from instruments etc. that would be difficult to make stealthy. Air intakes are covered with mesh, or follow convoluted paths designed to ensure that the radar signal is reflected many times and mostly absorbed before it exits.

Research is being conducted in the USA and Russia to create a plasma that can cover the aircraft. Plasma makes an excellent RAM as is not reflective but offers high attenuation. Under certain circumstances, a 10mm thick plasma could reduce the radar reflectivity of the underlying surface by 20dB.

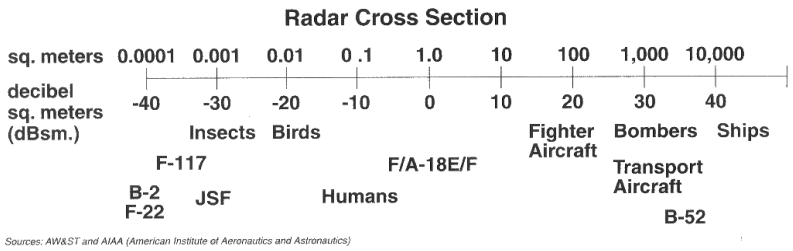


Figure 9.23: Comparison of the radar cross sections of various objects

Hand in hand with stealth development is that of electronic countermeasures. Broadband noise-like sources increase the noise floor and can reduce the effectiveness of a radar. These and other electronic jammers form a major part of the defensive strategy of modern military aircraft.

Stealth aircraft cannot operate in total electronic silence, so communications and radar systems on stealth aircraft must be low probability of intercept (LPI) and the radars must have extremely low sidelobes to reduce the probability of detection still further.

Anti stealth technology includes the following techniques:

- Radar with wavelengths longer than the aircraft.
- Bistatic and multi static radar configurations. These can use dedicated transmitters or existing FM or mobile phone broadcasts.
- Wide-band radar as it is difficult to make good wide-band RAM.
- Wake and exhaust detection and tracking, as neither of these can be completely eliminated.
- Wingtip vortex detection as vortices generate turbulence that changes the refractive index of the air, and so reflects radar signals.

9.7.1. Stealth Foiled by Cell Phone Network

America's multi-billion-dollar stealth bombers could be rendered obsolete by a British invention that uses existing mobile telephone masts to detect and track aircraft that were previously invisible to radar.

US stealth fighters and bombers such as the F117, B1 and B2 played key roles in the Gulf and Kosovo wars as they are almost impossible to detect using conventional radar.

However, the ease with which the mobile telephone mast system – developed at a laboratory in Hampshire – can be used to detect the aircraft has greatly concerned the military.

Peter Lloyd, the head of projects at Roke Manor Research, said: "I cannot comment in detail because it is a classified matter, but let's say the US military is very interested."

Stealth aircraft, each of which costs at least $A3.6 billion, are shaped to confuse radar. A special paint absorbs radio waves, reducing the radar signature to the equivalent of a gull in flight

The Roke Manor scientists discovered that telephone calls sent between mobile phone masts detected the precise position of stealth aircraft with ease. "We use just the normal phone calls that are flying about in the ether," Lloyd said. "The front of the stealth plane cannot be detected by conventional radar, but its bottom surface reflects very well."

Mobile telephone calls bouncing between base stations produce a screen of radiation. When the aircraft fly through this screen they disrupt the phase pattern of the signals. The Roke Manor system uses receivers, shaped like television aerials, to detect distortions in the signals.

A network of aerials large enough to cover a battlefield can be packed in a Land Rover.

Using a laptop connected to the receiver network, soldiers on the ground can calculate the position of stealth aircraft with an accuracy of 10 metres with the aid of the GPS satellite navigation system.

"It's remarkable that a stealth system that cost £60 billion [$158 billion] to develop is beaten by £100,000 mobile phone technology," Mr Lloyd said. "It's almost impossible to disable a mobile phone network without bombing an entire country, whereas radar installations are often knocked out of action with a single bomb or missile".

9.8. Stealth and Frequency Allocation

According to Barry Fox (New Scientist 31 August 2002), satellite based TV communications bandwidth is almost exhausted in the X-band and frequency has been allocated in the Ku-Band (12-18GHz). This band falls directly within the band transmitted by the radar of the B2 stealth bomber, and so not only would the high powered radar advertise the presence of the aircraft, but it could even damage sensitive receivers on the ground.

It will cost at least $1 billion to upgrade the radar

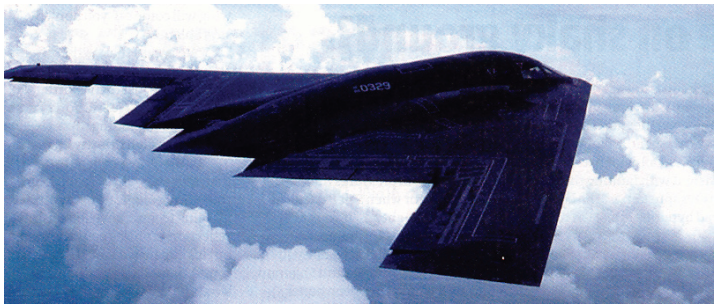


Figure 9.24: B2 stealth bomber

9.9. Complex Targets at IR and Visible Frequencies

The primary differences between radar and laser target characteristics are as follows:

- Laser targets are generally larger than the beam footprint
- Targets which appear smooth at microwave frequencies will appear rough at higher frequencies
- A diffuse surface at microwave may appear as a collection of specular scatterers to a laser

For a laser, the scattering cross section is

$$\sigma = \rho G A, \tag{9.14}$$

where: σ - Cross Section (m^2),
ρ - Reflectivity of the surface,
G – Gain of target,
A – Projected physical area (m^2).

The gain is given by

$$G = \frac{4\pi A_c}{\lambda^2} = \frac{4\pi}{\Omega}, \tag{9.15}$$

where: A_c – Area of coherence - λ^2/Ω,
Ω - Solid angle spread of the reflected light around the direction between the laser and the target.

On a scattering surface of physical area A, there are M independent coherence areas such that $M = A/A_c$. For coherent scattering $M = 1$ or $M >> 1$. For intermediate values of M, interference between coherence areas must be considered.

For $M = 1$, a specular scatterer

$$\sigma = \frac{4\pi\rho A^2}{\lambda^2}. \tag{9.16}$$

For $M >> 1$, the diffuse surfaces from which, according to Lamberts Law, the scattered power density decreases proportional to $\cos(\theta)$, where θ is the angle from the surface normal. In this case $\Omega = \pi$ and $\sigma = 4\pi\rho A\cos(\theta)$.

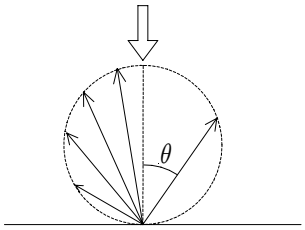


Figure 9.25: Lambertian scatterer

Instead of using cross section to define the characteristics of a laser target, distributed targets are often characterised by their reflection coefficient ρ (as discussed in Chapter 6) and physical cross section.

The reflection coefficient is a function of frequency, so the tables reproduced earlier for microwave and 10μm infrared will not be the same as those for 0.9μm infrared shown in the table below.

Table 9.4: IR reflectivity of various materials at 0.9μm

Diffusely Reflecting Material	**Reflectivity (%)**
White paper	Up to 100
Cut clean dry pine	94
Snow	80-90
Beer foam	88
White masonry	85
Limestone, clay	Up to 75
Newspaper with print	69
Tissue paper 2-ply	60
Deciduous trees	Typ 60
Coniferous trees	Typ 30
Carbonate sand (dry)	57
Carbonate sand (wet)	41
Beach sand and bare desert	Typ 50
Rough wood pallet (clean)	25
Smooth concrete	24
Asphalt with pebbles	17
Lava	8
Black neoprene	5
Black rubber tyre wall	2
Specular Reflecting Material	
Reflecting foil 3M2000X	1250
Opaque white plastic[1]	110
Opaque black plastic[1]	17
Clear plastic[1]	50

1 Measured with the beam perpendicular to the surface to achieve maximum reflection

For a diffuse scatterer, the reflection coefficient cannot exceed 100% but for a specular scatterer, the reflection coefficient can be many times this value. Manufacturers of laser range finders generally specify their performance for a target with 80% diffuse reflectivity.

9.10.Complex Acoustic Targets

9.10.1. Target Composition

As with electromagnetic radiation, in the acoustic domain, all materials will partially reflect, partially absorb and partially transmit the incident acoustic pulse. This is quantified by the coefficient of reflectivity K_r of the target

$$K_r = \frac{I_r}{I_i} = \left(\frac{Z_a - Z_o}{Z_a + Z_o} \right)^2 , \tag{9.17}$$

where: I_r – Reflected intensity of the sound,
I_I – Incident intensity of the sound,
Z_a – Acoustic impedance of air,
Z_o – Acoustic impedance of object (target).

As the acoustic impedance Z_o is related to the propagation velocity, hard dense targets tend to reflect well (as their propagation velocity is high), while soft light targets tend to transmit or absorb. Z_o is not only a function of the target material, but also its texture.

Lower frequencies are often absorbed by porous targets and so air, it is generally best to operate at the highest frequency that will propagate effectively.

9.10.2. Target Properties

While material properties are important at a microscopic level, the acoustic pulse interacts with a relatively large area of the target, so the strength and quality of an echo from the target will depend on its geometry. For short range applications where the insonified footprint is smaller than the target extent, shape is important because many targets are specular in nature (walls etc.)

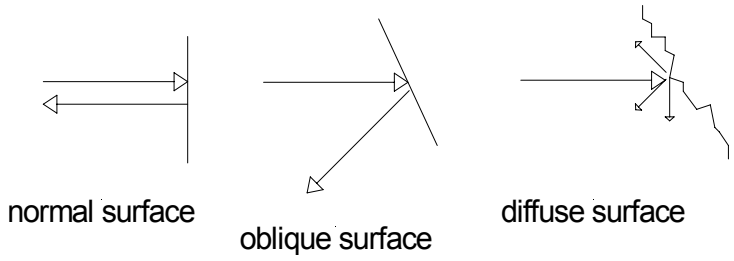
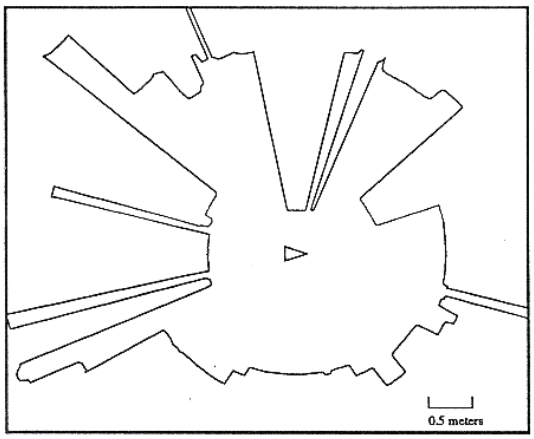


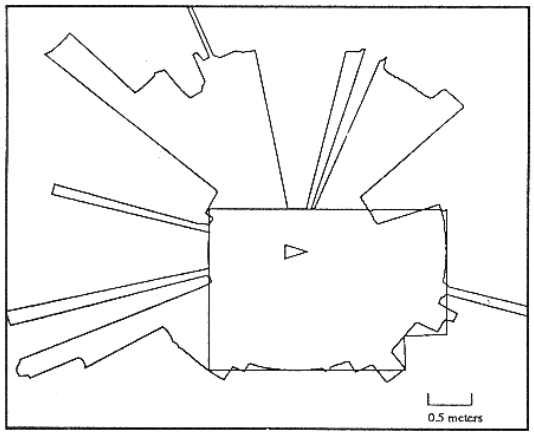
Figure 9.26: Rough and smooth scatterers

As the frequency increases, the effective surface roughness, as a function of the acoustic wavelength, increases and targets become less specular. As a rule of thumb, the amplitude of the signal reflected from a target will increase with its size until it is approximately 10λ across.

One of the major issues with short range ultrasound sonar applications, particularly in structured environments like the interiors of buildings, is that the walls are specular and the corners are retro reflectors. These characteristics, in conjunction with a significant multipath problem, results in extremely confusing returns. It can be seen in the figure below that most of the returns are either from corners, the walls at normal incidence or as a result of multiple bounce echoes from corners via the walls at oblique angles.



A typical sonar scan



The same scan, superimposed on a hand-measured room model

Figure 9.27: Effect of specular reflection on sonar scans for a smooth walled room

From these results it is obvious that the use of this technology for indoor navigation is fraught with difficulty if the observation is made from a single position. However, by

fusing the results of multiple scans made from different positions within the room, a better match between the measured and true interior is possible.

9.10.3. Particulate Targets

Many acoustic systems are used for industrial applications such as level measurement in bins or silos. The target in these instances are often particulate in nature and it has two characteristics that are important as regards the echo strength:

- Small scale granularity
- Large scale angle of repose and undulation.

Granularity

Granular particles scatter the reflected wave in all directions which is essential for an echo return if the material is lying at an angle to the normal. If, however, the particle size is comparable to $\lambda/2$, then significant cancellations can occur.

As a rule of thumb, the acoustic wavelength should be chosen to exceed the grain size by a factor of four.

Angle of Repose and Undulations

If the material surface lies at an angle to the incident acoustic wave, the echo can be reflected away from the transducer towards the walls of the vessel. This can result in the echo return following a zig-zag path and an incorrect range reading.

With an undulating surface in which the period of the undulations is shorter than beam footprint, the echo can be directed to follow multiple paths back to the transducer which can spread the pulse and result in a lower probability of detection.

9.11.Clutter

One complicating factor in the study of clutter is that it means different things in different situations. For example, to an engineer developing a missile to detect and track a tank, the return from vegetation and other natural objects would be considered to be “clutter”. However, a remote sensing scientist would consider the return from natural vegetation as the primary target. Clutter is thus defined as the return from a physical object or a group of objects that is undesired for a specific application.

Clutter may be divided into sources distributed over a surface (land or sea), within a volume (weather or chaff) or concentrated at discrete points (structures, birds or vehicles).

9.12.Surface Clutter

The magnitude of the signal reflected from the surface back to the receiver is a function of the material, roughness and angle. There are three primary scattering types into which clutter is generally classified. These are specular, retro and diffuse as shown in the figures below.

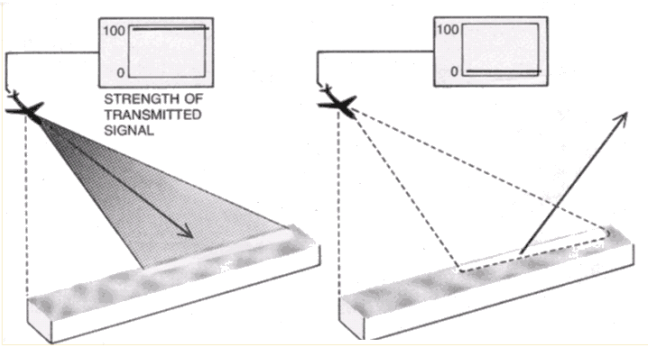


Figure 9.28: Specular clutter in which most of the signal is reflected away from the radar because the surface is smooth compared to the transmitted wavelenght

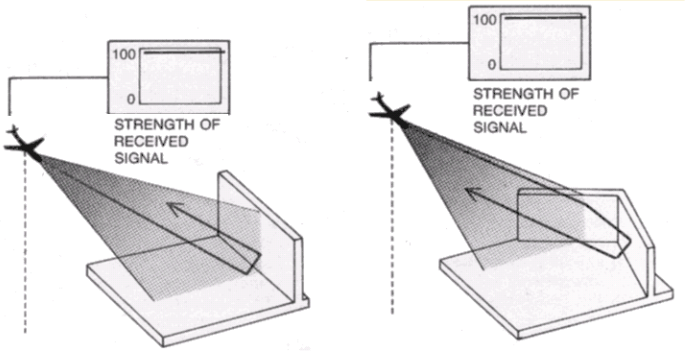


Figure 9.29: Retro clutter in which a large portion of the signal is returned to the radar due to the configuration of multiple reflecting surfaces

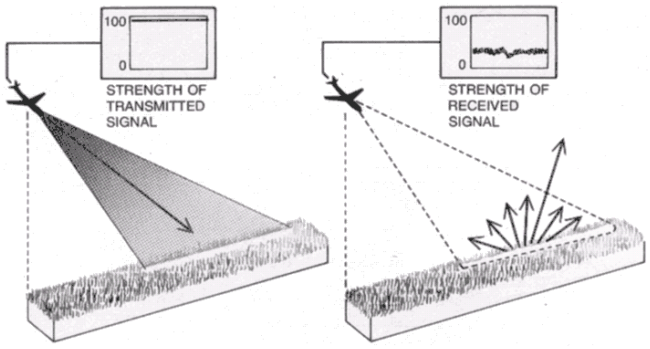


Figure 9.30: Diffuse clutter in which the signal is reflected in all directions with the result that a small fluctuating proportion is reflected back to the radar

9.12.1. Ground Clutter

Because of the statistical nature of clutter, the mean reflectivity is most often quoted. A convenient mathematical way to describe this mean value for surface clutter is the *constant* γ model in which the surface reflectivity is modelled as

$$\sigma^o = \gamma \sin\psi \,, \tag{9.18}$$

where σ^o – Reflectivity (cross section per unit area m^2/m^2),
ψ - Grazing angle at the surface (rad),
γ - Parameter describing the scattering effectiveness.

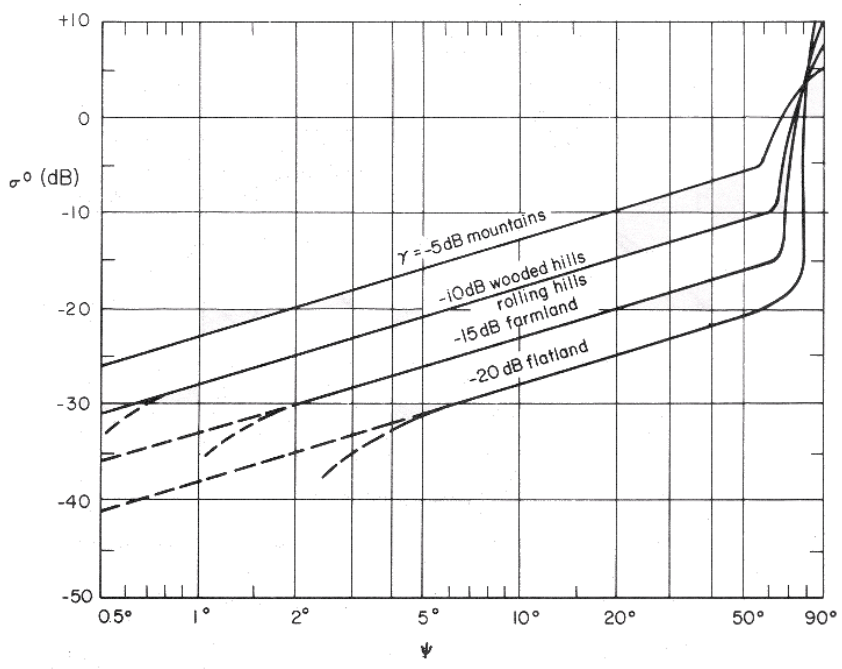


Figure 9.31: Effect of grazing angle on clutter reflectivity for different clutter types

It can be seen from the figure that, at low grazing angles, the measurements fall below the model because of propagation-factor effects. At high grazing angles the measured reflectivity rises above the value predicted by the model because of quasi specular reflections from surface facets.

For different surface types, the following are typical:

- Values for γ between –10 and –15dB are widely applicable to land covered by crops, bushes and trees.
- Desert, grassland and marsh are more likely to have γ near –20dB
- Urban or mountainous regions will have γ near –5dB

These values are almost independent of wavelength and polarisation, but they only apply to modelling of mean clutter reflectivity.

9.12.2. Measured Clutter Reflectivity at 95GHz

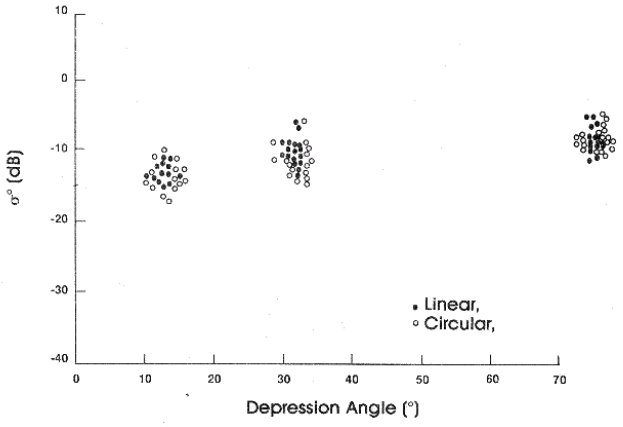


Figure 9.32: Reflectivity of grass and crops at 95GHz

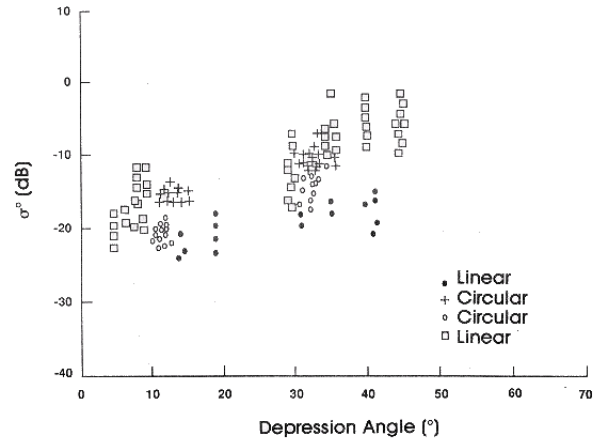


Figure 9.33: Reflectivity of deciduous trees at 95GHz

9.12.3. Sea Clutter

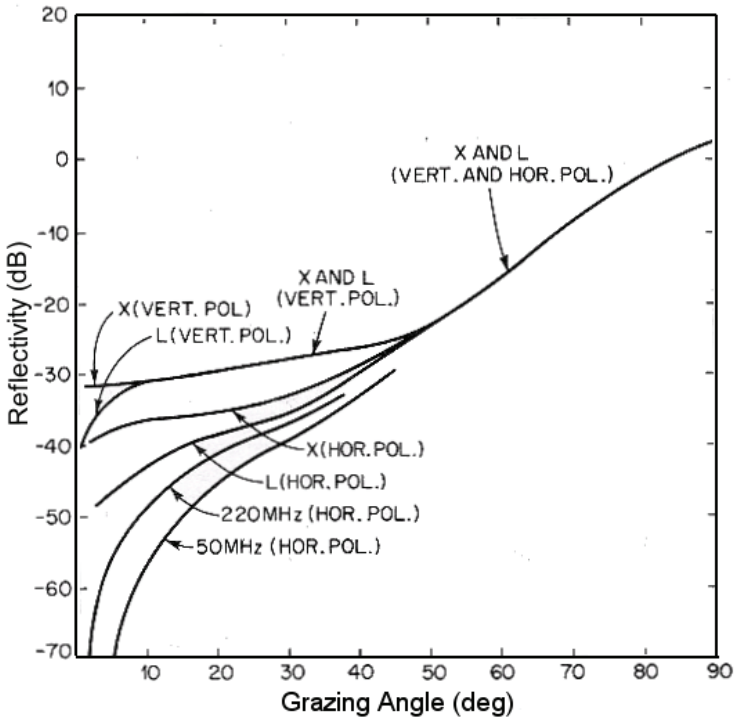


Figure 9.34: Effect of grazing angle on sea clutter reflectivity

When the model is applied to sea clutter, averaging over all wind directions, it is found that γ depends on the Beaufort wind scale K_B and the wavelength according to the following empirical relationship,

$$10\log\gamma = 6K_B - 10\log\lambda - 64\,. \tag{9.19}$$

At low grazing angles there is also a component that is a function of the polarisation

Table 9.5: The Beaufort Scale

Beaufort No	Description	Wind Speed (kts)
0	Calm	<1
1	Light air	1-3
2	Light breeze	4-6
3	Gentle breeze	7-10
4	Moderate Breeze	11-16
5	Fresh breeze	17-21
6	Strong breeze	22-27
7	Near gale	28-33
8	Gale	34-40
9	Strong gale	41-47
10	Storm	48-55
11	Violent Storm	56-63
12	Hurricane	>64

9.13.Calculating Surface Clutter Backscatter

The cross section of the ground clutter is the product of the reflectivity, σ^o, and the illuminated area, A.

$$\sigma = \sigma^o A \tag{9.20}$$

The accurate measurement of surface reflectivity is fraught with difficulty. To begin with, two possible geometries must be considered when calculating the measurement area, depending on whether the whole ground echo falls within one range cell or not as determined by the inequality expressed in the figure below.

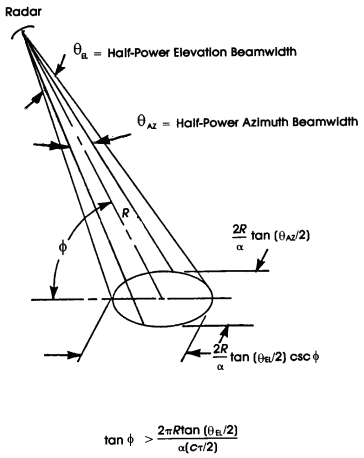


Figure 9.35: Definition of a beamwidth-limited resolution cell

In the beamwidth-limited example, it can be seen that the illuminated area is defined by the intersection of the conical beam with the surface of the target. Due to the geometry and the length of the transmitted pulse, it is assumed that all of the power reflected from this surface contributes to the RCS of the target and so the area can be approximated by an elliptical footprint.

$$A = \pi r_1 r_2 \tag{9.21}$$

$$A = \left(\frac{\pi R^2}{\alpha^2}\right)\left(\tan\frac{\theta_{AZ}}{2}\right)\left(\tan\frac{\theta_{EL}}{2}\right)\csc\phi \tag{9.22}$$

where A – Area illuminated on the surface (m2)
R – Slant range to the surface (m)
α - Beam shape factor =1.33 for a Gaussian shaped beam
θ_{AZ} – Azimuth 3dB two-way beamwidth (rad)
θ_{EL} – Elevation 3dB two-way beamwidth (rad)
ϕ - Grazing angle (rad)

This operational mode can be extended to encompass the measurement of the returns from walls and trees if the appropriate geometry is applied, and the tree surface is assumed to be impenetrable by the radar signal.

One alternative is to gate the radar signal in range, but once again, this is not satisfactory unless a well defined surface is being measured as shown in the range gate limited case below:

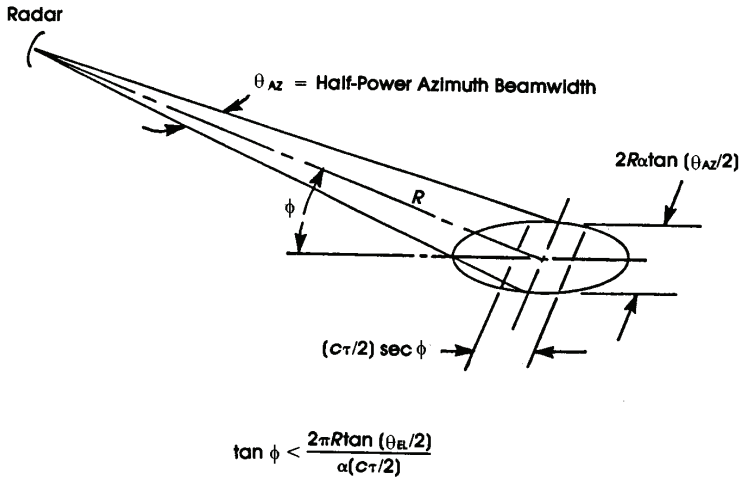


Figure 9.36: Definition of a pulse-width (range-gate) limited resolution cell

If the projected pulse width is sufficiently short compared to the length of the elliptical footprint defined by the elevation beam width, then the area can be approximated by a rectangle, and the formula for the area, A, is

$$A = \frac{Rc\tau}{\alpha}\left(\tan\frac{\theta_{AZ}}{2}\right)\sec\phi\,, \tag{9.23}$$

where c is the speed of light (3×10^8 m/s) and τ the transmitted pulsewidth (sec).

9.14.Calculating Volume Backscatter

An analogous normalisation parameter to σ^o exists for volume scatterers in the atmosphere such as dust particles or raindrops. It is called η.

η - RCS of illuminated volume / illuminated volume (m^2/m^3).

For the measurement of foliage or rain where some penetration occurs, then the volume clutter formulation is appropriate. In this formulation an analogous normalisation parameter to σ^o called η exists for volume scatterers. This volume reflectivity is defined as the RCS per unit illuminated volume (m^2/m^3).

As with the surface cases, the illuminated volume can be calculated from the geometry shown below, which is just the volume of an elliptical cylinder with diameters $R\theta_{AZ}$ and $R\theta_{EL}$ and with length $c\tau/2$.

$$V = \frac{\pi R^2 \theta_{AZ} \theta_{EL} c\tau}{8\alpha^2} \tag{9.24}$$

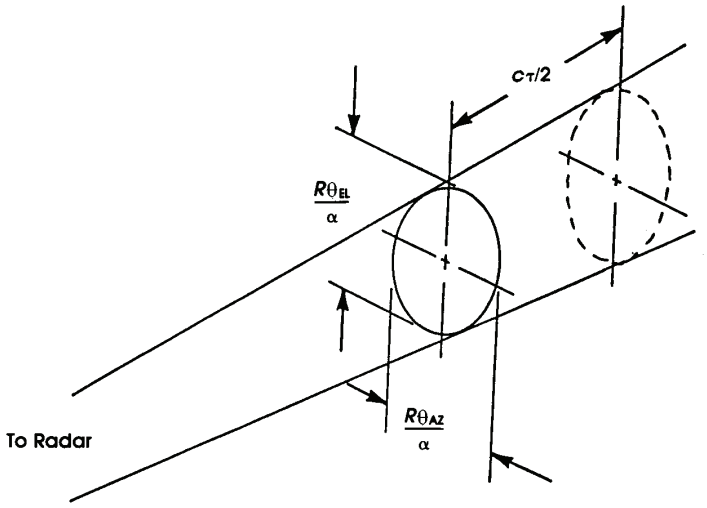


Figure 9.37: Definition of a volume-clutter resolution cell

The radar cross section is then the product of the volume reflectivity η and the volume V as follows:

$$\sigma = \eta V \tag{9.25}$$

This model assumes that there is minimal attenuation of the radar signal over the length of the cell. This may be a valid assumption at lower frequencies where foliage and rain penetration is good, but it is not in the millimetre-wave band where the measured two way attenuation in dry foliage exceeds 4dB/m and is even higher in dense green foliage. In this case a more complex formulation to take this attenuation into account is required.

$$\sigma = \eta V \exp(-c\tau\alpha) \tag{9.26}$$

This can be rewritten using the more common dB notation to take advantage of the fact that attenuation is most often given in that form.

$$\sigma_{dB} = 10\log_{10}(\eta V) - c\tau\alpha \tag{9.27}$$

9.14.1. Rain

The graphs below show the theoretical values for the reflectivity as a function of rainfall rate at different frequencies.

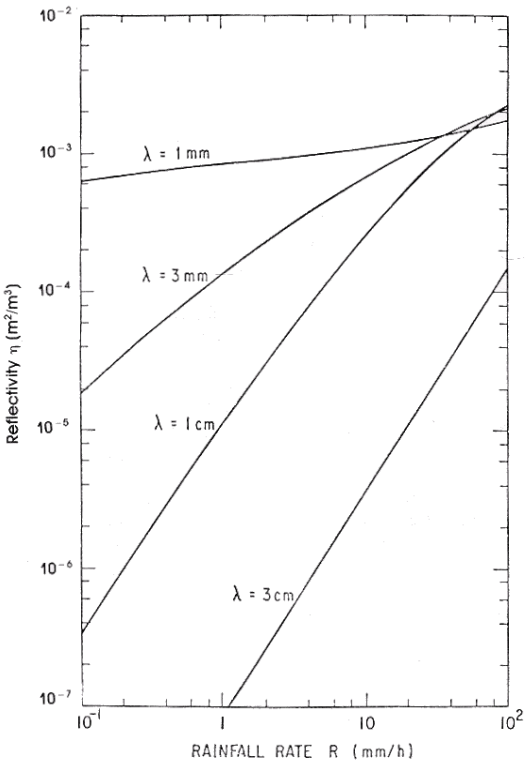


Figure 9.38: Theoretical raindrop reflectivity vs rainfall rate using Marshall Palmer drop size distribution

This data is determined using the relationship between the reflected and incident power on small spherical targets as discussed earlier in the section on the RCS of a sphere. Though a given rainfall rate does not imply a specific drop-size distribution, the trend that the drops get bigger as the rainfall rate increases, generally holds true.

In the Rayleigh region ($\pi D/\lambda < 1$), the RCS is given by the following formula

$$\sigma = 4\pi R^2 \frac{S_{refl}}{S_{inc}} = \pi^5 \left|K\right|^2 \frac{D^6}{\lambda^4}, \tag{9.28}$$

$$K = \frac{\varepsilon - 1}{\varepsilon + 2}, \tag{9.29}$$

where ε - Relative dielectric constant of the material,
D – Diameter of the scattering object,
λ - Wavelength.

When $\pi D/\lambda > 10$ the equation for RCS reduces to the geometric optics form

$$\sigma = \frac{\pi D^2}{4} . \tag{9.30}$$

These equations can be combined with the density of particles in the medium to determine the total reflectivity, η.

$$\eta = \sum_{i=1}^{N} \sigma_i . \tag{9.31}$$

9.14.2. Dust

The volume of dust that can be supported in the atmosphere is extremely small, and so the reflectivity can often be neglected for EM radiation with wavelengths of 3mm or more. However, under certain circumstances, if the dust density is very high (such as in rock crushers) or if the propagation path through dust is very long (in dust storms), then it can be useful to determine the reflectivity and the total attenuation.

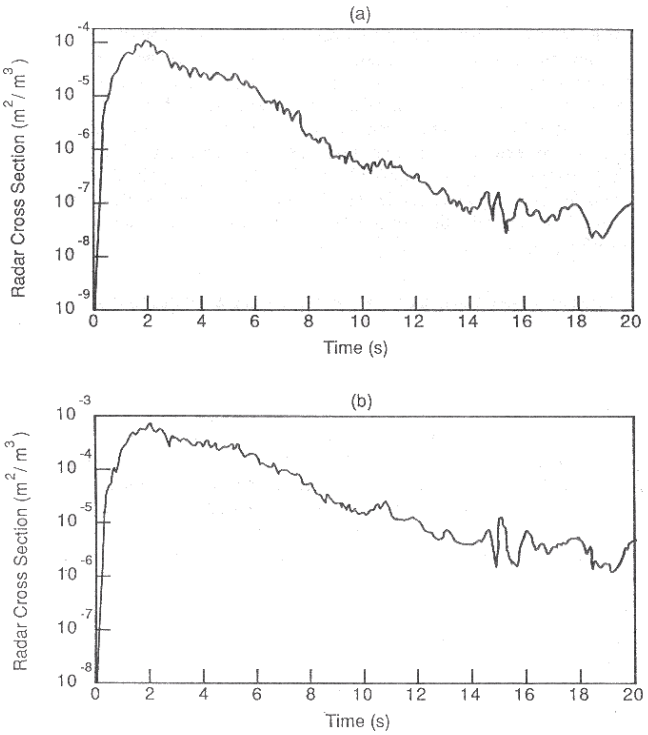


Figure 9.39: Backscatter from dust after explosion (a) at 10GHz and (b) 35GHz

9.15.Estimating the Backscatter of Radar Signals

The effectiveness of a ranging sensor, such as a radar or lidar, is dependent not only on the actual signal level returned from the target of interest, but also the relative level of this signal in comparison to other competing returns at the same range.

The most common sources of clutter are returns from dust or water droplets within the radar beam, or from large (high reflectivity) returns that enter through the sidelobes of the antenna.

Expanding on the equations described in the previous section, it can be shown that the reflectivity for a cloud of dust particles using the Rayleigh approximation is

$$\eta = \frac{\pi^5}{\lambda^4}\left|K_o\right|^2 Z\,, \tag{9.32}$$

where K_o is related to the relative dielectric constant of the particle and the Z factor is determined from the number and size of the particles

$$\left|K_0\right|^2 = \left|\frac{\varepsilon - 1}{\varepsilon + 2}\right|^2, \tag{9.33}$$

$$Z = \sum_i D_i^6 N_i \Delta D\,, \tag{9.34}$$

where $N_i(D)\Delta D$ is the number of dust particles whose diameters are between D_i and $D_i+\Delta D$, ε is the complex dielectric constant of the particle.

The equation for Z can be re-written in terms of the total number of airborne particles per unit volume N_T and P_i which is the probability that a dust particle has a diameter between D_i and $D_i+\Delta D$, per unit volume[2],

$$Z = N_T \sum_i D_i^6 P_i\,. \tag{9.35}$$

The total number of particles per unit volume can also be expressed in terms of the mass loading

$$M = \frac{4}{3}\pi\rho\sum_i N_i r_i^{\,3} = \frac{4}{3}\pi\rho N_T \sum_i P_i r_i^{\,3}\,. \tag{9.36}$$

[2] Also called the number fraction distribution

As discussed in Chapter 8, in a dust storm the visibility can be related to the mass loading of dust per cubic meter of air by

$$M = \frac{C}{V^{\gamma}} \tag{9.37}$$

where M is the mass loading of dust in g/m^3, V is the visibility in metres and C and γ are constants that depend on the particle type and the meteorological conditions. Typical values for these constants are $C = 37.3$ and $\gamma = 1.07$.

Using the mass loading formulation in the equation above and solving for N_T the following is obtained

$$N_T = \frac{3}{4\pi}.\frac{C}{\rho}.\frac{1}{V^{\gamma}}.\frac{1}{\sum_i P_i r_i^{\,3}} = \frac{2.25\times10^{-9}}{V^{1.07}\sum_i P_i r_i^{\,3}}. \tag{9.38}$$

Substituting

$$Z = \frac{2.25\times10^{-9}}{V^{1.07}}\frac{\sum_i P_i D_i^{\,6}}{\sum_i P_i r_i^{\,3}} \tag{9.39}$$

Estimates are made for the two summation terms based on typical size distributions determined for sandstorms

$\sum_i P_i r_i^{\,3} = 4\times10^{-14}$ (m^3),

$\sum_i P_i D_i^{\,6} = 2\times10^{-24}$ (m^6).

Substituting back to obtain a formula for the backscatter,

$$\eta = \frac{\pi^5}{\lambda^4}\left|K_o\right|^2.\frac{1.125\times10^{-19}}{V^{1.07}} \text{ (m}^2\text{/m}^3\text{).} \tag{9.40}$$

The following figure shows the reflectivity plotted for coal dust and for water spray with identical particle/droplet size distributions.

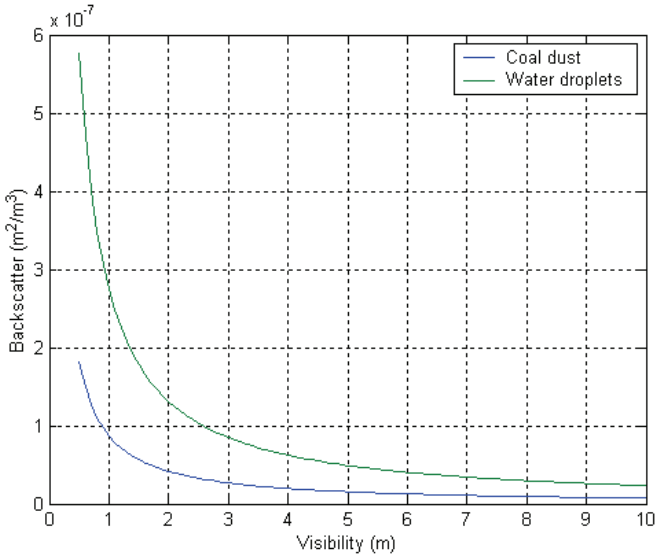


Figure 9.40: Backscatter from coal dust and water with identical particle size distributions as a function of the visibility at 94GHz

As the frequency is increased and the backscatter mechanism moves out of the Rayleigh region into the Mie region, total backscatter increases significantly and the medium becomes more opaque. This is the reason that clouds look white at optical frequencies whereas even heavy rain is more transparent even though the mass loading of rain often higher than that of clouds

Under most circumstances the optical visibility through the medium is a good indication of the performance in the near IR, so the rule-of-thumb that states "if you can see the target, a lidar sensor can measure the range" holds.

9.16.References

[1] M.Skolnik (ed), *Radar handbook,* McGraw Hill, 1970.
[2] N.Currie (ed), *Radar reflectivity Measurement, Techniques and Applications,* Artech, 1989.
[3] D.barton, *Modern Radar Systems Analysis,* Artech, 1988
[4] M.Skolnik, *Introduction to Radar Systems,* McGraw Hill, 1980
[5] H.Durrant-Whyte, *Sensors and Signals Notes.*
[6] Principle of a Pulsed Laser Sensor, http://www.riegl.co.at, 26/02/2001.
[7] D.Fulghum, Stealth Retains Value, But its Monopoly Wanes, Aviation Week and Space technology, Feb. 5, 2001.
[8] Stealth Foiled by Cell Phones, http://www.smh.com.au/news/0106/12/world/world2.html, June 13, 2001.
[9] P.Bhartia, I.Bahl, *Millimeter Wave Engineering and Applications,* John Wiley & Sons, 1984
[10] Principle of a Pulsed Laser Sensor, http://www.riegl.co.at, 26/02/2001.
[11] D.Barton, *Radar Systems Analysis*, Artech 1976.
[12] C.Currie (ed), *Principles and Applications of Millimeter Wave Radar,* Artech, 1987.
[13] H. Jensen et. al., Side-Looking Airborne Radar, Scientific American, October 1977.

Chapter 10.
Detection of Signals in Noise

10.1.Receiver Noise

Noise is the unwanted energy that interferes with the ability of the receiver to detect the wanted signal. It may enter the receiver through the antenna along with the desired signal or it may be generated within the receiver. In underwater sonar systems external acoustic noise is generated by waves and wind on the water surface, by biological agents (fish, prawns etc) and manmade sources such as engine noise. In radar and lidar sensors the external electromagnetic noise is generated by various natural mechanisms such as the sun and lightning amongst others. Manmade sources of electromagnetic noise are myriad, from car ignition systems and fluorescent lights through other broadcast signals.

As discussed earlier, noise within the sensor is generated by the thermal motion of the conduction electrons in the ohmic portions of the receiver input stages. This is known as *Thermal* or *Johnson* Noise.

Noise power P_N is expressed in terms of the temperature T_o of a matched resistor at the input of the receiver

$$P_N = kT_o\beta \quad \text{W}, \tag{10.1}$$

where: k – Boltzmann's Constant (1.38×10^{-23} J/K),
T_o – System Temperature (usually 290K),
β – Receiver Noise Bandwidth (Hz).

The noise power in practical receivers is always greater than that which can be accounted for by thermal noise alone.

The total noise at the output of the receiver, N, can be considered to be equal to the noise power output from an ideal receiver multiplied by a factor called the Noise Figure, NF

$$N = P_N F_N = kT_o\beta.NF \quad \text{W}. \tag{10.2}$$

10.2.Noise Probability Density Functions

Consider a typical radar front-end that consists of an antenna followed by a wide band amplifier, a mixer that down converts the signal to an intermediate frequency (IF) where it is further amplified and filtered (bandwidth β_{IF}). This is followed by an envelope detector and further filtering (bandwidth $\beta_V = \beta_{IF}/2$).

The noise entering the IF filter is assumed to be Gaussian (as it is thermal in nature) with a probability density function (PDF) given by

$$p(v) = \frac{1}{\sqrt{2\pi\psi_o}} \exp \frac{-v^2}{2\psi_o}, \tag{10.3}$$

where $p(v)dv$ - probability of finding the noise voltage v between v and v+dv,
ψ_o - variance of the noise voltage.

If Gaussian noise is passed through a narrow band filter (one whose bandwidth is small compared to the centre frequency), then the PDF of the post-detection envelope of the noise voltage output can be shown to be

$$p(R) = \frac{R}{\psi_o} \exp \frac{-R^2}{2\psi_o}, \tag{10.4}$$

where R is the amplitude of the envelope of the filter output. This has the form of the Rayleigh probability density function

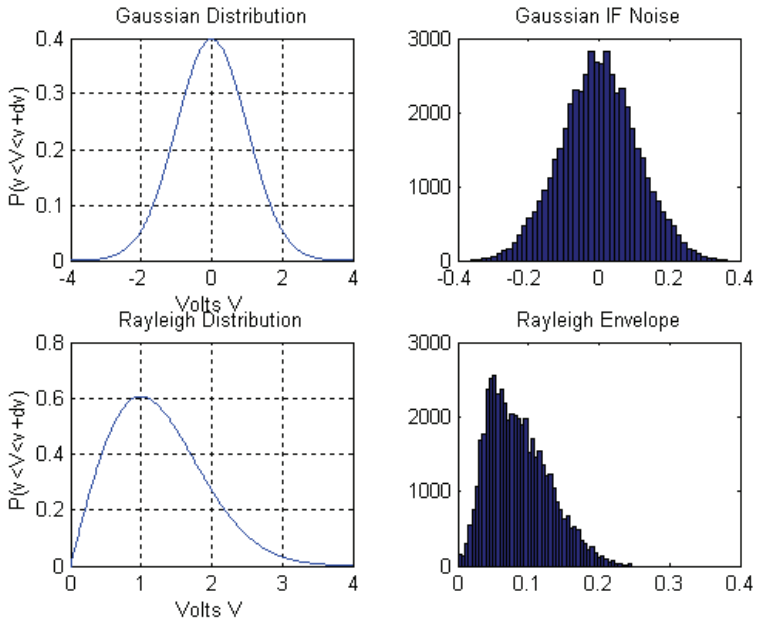


Figure 10.1: Amplitude distributions of thermal noise pre and post detection

10.3.Probability of False Alarm

A false alarm occurs whenever the noise voltage exceeds a defined threshold voltage, V_t, as illustrated in the figure below. The probability of this occurring is determined by integrating the PDF as shown

$$\Pr ob(V_t < R < \infty) = \int_{V_t}^{\infty} \frac{R}{\psi_o} \exp \frac{-R^2}{2\psi_o} dR = \exp \frac{-V_t^2}{2\psi_o} = P_{fa} . \tag{10.5}$$

It can be seen that the average time interval between crossings of the threshold, called the false alarm time, T_{fa}, can be written as

$$T_{fa} = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} T_k , \tag{10.6}$$

where T_k – Time between crossings of the threshold V_t by the noise envelope (when the slope of the crossing is positive).

The false alarm probability could also have been defined as the ratio of the time that the envelope is above the threshold to the total time as shown graphically in the figure below

$$P_{fa} = \frac{\sum_{k=1}^{N} t_k}{\sum_{k=1}^{N} T_k} = \frac{\langle t_k \rangle_{ave}}{\langle T_k \rangle_{ave}} = \frac{1}{T_{fa}\beta} , \tag{10.7}$$

where t_k and T_k are defined in the figure, and the average duration of a noise pulse is the reciprocal of the bandwidth β.

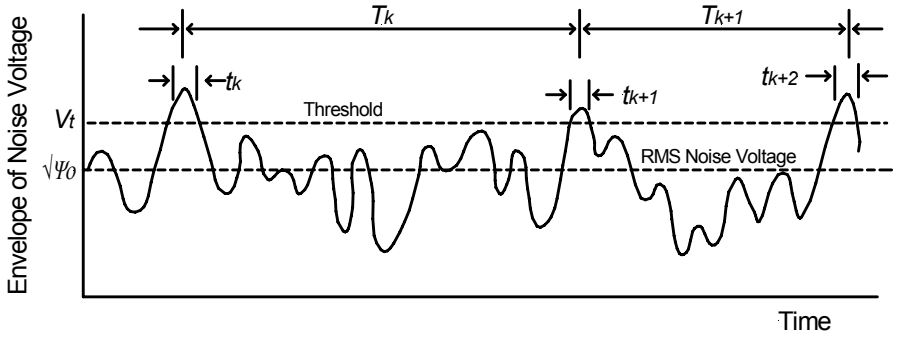


Figure 10.2: Receiver output voltage illustrating false alarms due to noise

For a bandwidth $\beta = \beta_{IF}$, the false alarm time is just

$$T_{fa} = \frac{1}{\beta_{IF}} \exp \frac{V_t^2}{2\psi_o} . \tag{10.8}$$

The false alarm times of practical radars must be very large (usually a couple of hours), so the probability of false alarm must be very small, typically $P_{fa} < 10^{-6}$.

10.4. Probability of Detection

Consider that a sine wave with amplitude, A, is present along with the noise at the input to the IF filter. The frequency of the sine wave is equal to the centre frequency of the IF filter. It is shown by Rice that the signal at the output of the envelope detector will have the following PDF (known as a Rician distribution)

$$p_s(R) = \frac{R}{\psi_o}\exp\left(-\frac{R^2 + A^2}{2\psi_o}\right) I_o\left(\frac{RA}{\psi_o}\right). \tag{10.9}$$

$I_o(Z)$ modified Bessel function of order zero and argument Z. It can be shown that for large Z, an asymptotic expansion for $I_o(Z)$ is

$$I_o(Z) \approx \frac{e^Z}{\sqrt{2\pi Z}}\left(1 + \frac{1}{8Z} + \ldots\right). \tag{10.10}$$

The probability that the signal will be detected is the same as the probability that the envelope, R, will exceed the threshold, V_t is

$$p_d = \int_{V_t}^{\infty} p_s(R)dR = \int_{V_t}^{\infty} \frac{R}{\psi_o}\exp\left(-\frac{R^2 + A^2}{2\psi_o}\right) I_o\left(\frac{RA}{\psi_o}\right)dR. \tag{10.11}$$

Unfortunately, this cannot be evaluated in a closed form and so numerical techniques or a series approximation must be used. However, this has already been done, and tables and a series of curves have been produced.

In terms of the PDFs for the noise and the signal plus noise voltages, the detection and false alarm process is shown graphically in the figure below. The lightly shaded area represents the P_{fa} and the dark shaded area (including the tail covered by the light shading) represents the P_d.

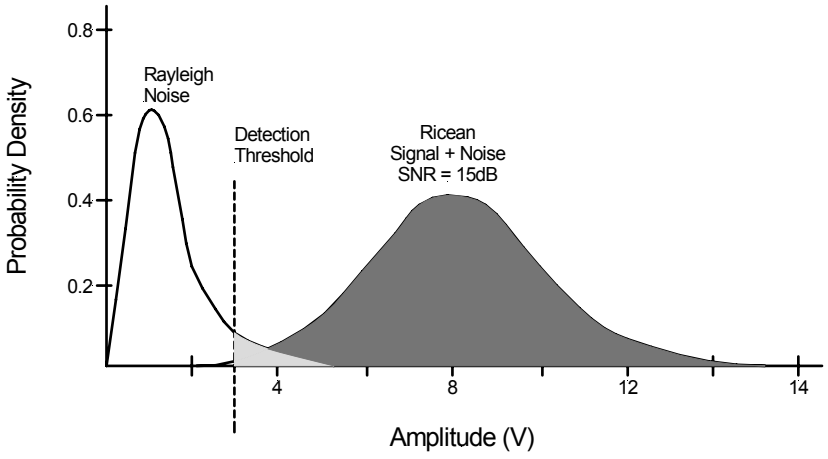


Figure 10.3: PDF's of noise and signal plus noise for $P_{fa} = 10^{-2}$

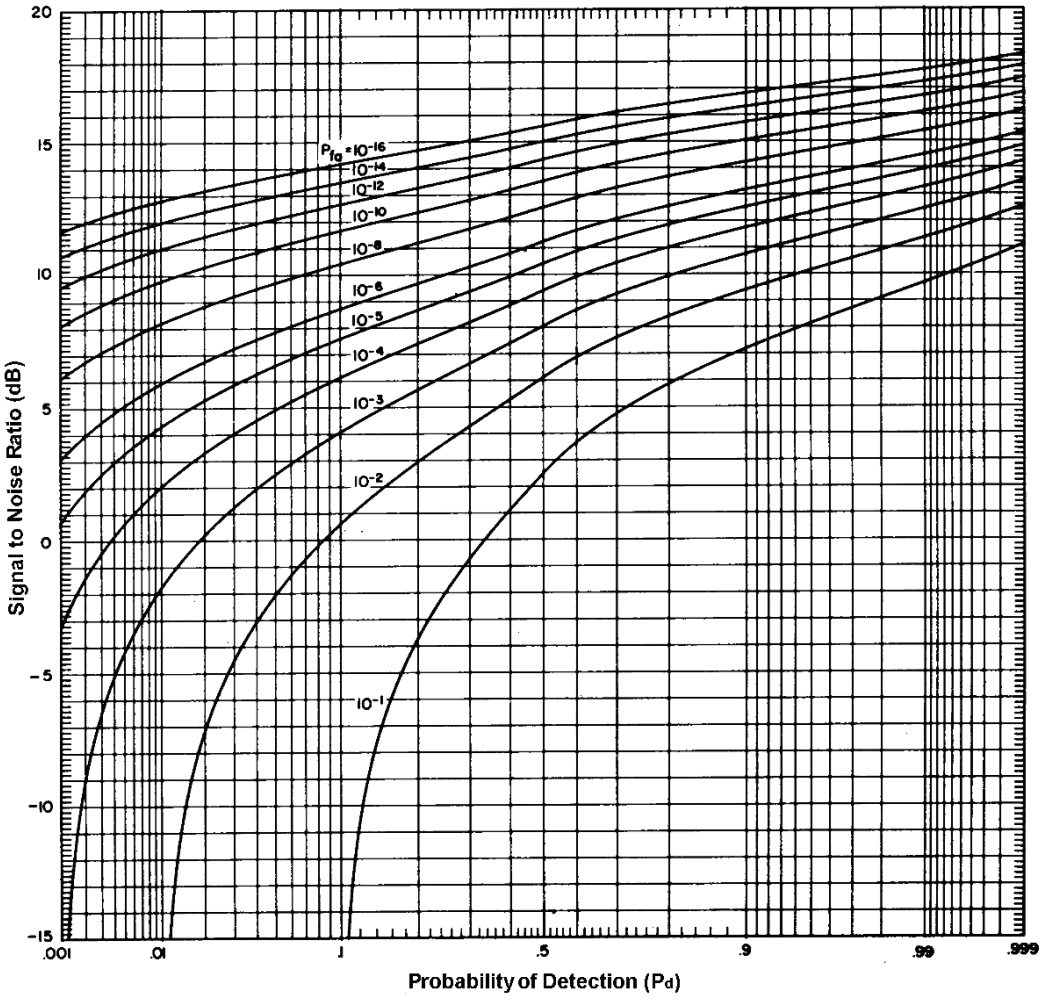


Figure 10.4: Detection probability as a function of signal to noise ratio with false alarm probability as a parameter

A typical radar system will operate with a detection probability of 0.9 and a probability of false alarm of 10^{-6}. The required signal to noise ratio can be read directly off the graph as 13.2dB. Note that this is for a single pulse of a steady sinusoidal signal in Gaussian noise with no detection losses.

```
% Pd and Pfa Determined Numerically by running a time domain simulation
% Generate noise and signals for SNR=13dB
a = (1:1000000);
x = randn(size(a));
sigi = 6.31*sin(a/1000);
y = randn(size(a));
sigq = 6.31*cos(a/1000);
% Determine the envelope
c = sqrt(x.^2+y.^2);
csig = sqrt((x+sigi).*(x+sigi)+(y+sigq).*(y+sigq));
%Plot the distributions
edges=(0:0.1:12);
n=histc(c,edges);
ns=hist(csig,edges);
plot(edges,n,edges,ns);
grid
% Determine the probabilities for different thresholds
vt=5;
% Look for noise peaks above the threshold
nfa = find(c>vt);
```

```
pfa = length(nfa)/1000000
% Look for S+N peaks above the threshold
nd = find(csig>vt);
pd = length(nd)/1000000
```

10.5.Detector Loss Relative to an Ideal System

An envelope detector is used by a radar system when the phase of the received pulse is unknown. This is called non-coherent detection, and it results in a slightly higher SNR requirement than the curves above show.

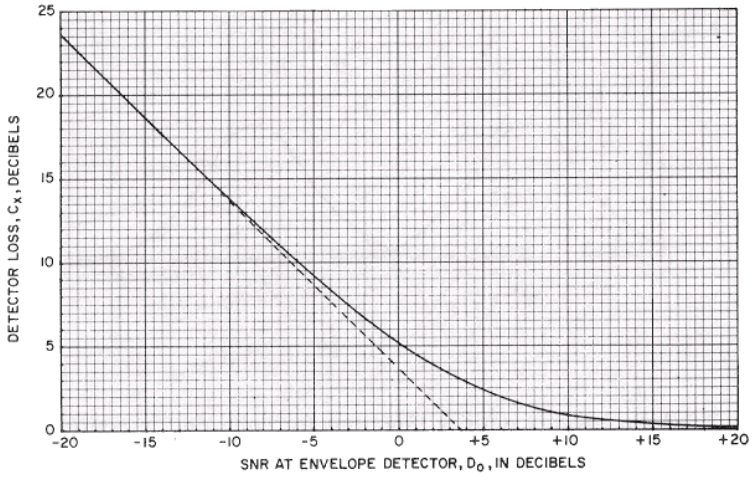


Figure 10.5: Envelope detector loss as a function of signal to noise ratio

This loss factor C_x is approximately

$$C_x(1) \approx \frac{SNR(1) - 2.3}{SNR(1)}, \tag{10.12}$$

where $C_x(1)$ – Loss in SNR,
$SNR(1)$ – The pre detector single pulse SNR required to achieve a particular P_d and P_{fa}.

The graph shows that for good SNR, the detector loss is very small. In the case where the $P_d = 0.9$ and the $P_{fa} = 10^{-6}$ it is only about 0.4dB. This effect is known as small signal suppression as it becomes much more pronounced as the SNR decreases.

One of the advantages of coherent detection is that it has zero response to the quadrature noise component, whereas, this component is translated into phase modulation after envelope detection with the result that the effective SNR in the latter case is degraded. However the phase modulation is very small for high SNR, and few advantages can be gained by coherent detection for SNR > 10dB (C_x = 1dB) over envelope detection.

10.6. The Matched Filter

To achieve the best possible SNR, the characteristics of the IF filter must be matched to those of the signal pulse.

The peak signal to (average) noise power ratio of the output response of the matched filter is equal to twice the received signal energy, *E,* divided by the single-sided noise power per Hz, N_o

$$\left(\frac{\hat{S}}{N}\right)_{out} = \frac{2E}{N_o}, \tag{10.13}$$

where $\hat{S}$ – Peak instantaneous signal power seen during the matched filter response to a pulse (W),
N – Average noise power (W),
E – Received signal energy (J),
N_o – Single sided noise power density (W/Hz).

The received energy is the product of the received power, *S*, as determined by the range equation and the pulse duration, τ

$$E = S\tau, \tag{10.14}$$

And the noise power density is the received noise power, *N,* divided by the bandwidth, β_{IF}

$$N_o = \frac{N}{\beta_{IF}}. \tag{10.15}$$

Substituting into (10.13)

$$\left(\frac{\hat{S}}{N}\right)_{out} = \frac{2S\tau}{N/B_{IF}} = \left(\frac{S}{N}\right)_{in} 2B_{IF}\tau. \tag{10.16}$$

When the bandwidth of the signal at IF is small compared to the centre frequency then the peak power is approximately twice the average power in the received pulse. So the output SNR is

$$\left(\frac{S}{N}\right)_{out} \approx \left(\frac{S}{N}\right)_{in} B_{IF}\tau. \tag{10.17}$$

The matched filter should not be confused with the circuit theory concept of matching that maximises power transfer rather than SNR.

In practise a matched filter implementation is often hard to achieve exactly, so compromises are made as shown in the following table:

Table 10.1: Efficiency of non-matched filters

Input Signal Shape	Matched Filter Characteristic	Optimum $B.\tau$	Loss in SNR compared to Matched Filter (dB)
Rectangular Pulse	Rectangular	1.37	0.85
Rectangular Pulse	Gaussian	0.72	0.49
Gaussian Pulse	Rectangular	0.72	0.39
Gaussian Pulse	Gaussian	0.44	0 (matched)
Rectangular Pulse	Single tuned circuit	0.4	0.88
Rectangular Pulse	Two cascaded tuned circuits	0.613	0.56
Rectangular Pulse	Five cascaded tuned circuits	0.672	0.5

10.7.Integration of Pulse Trains

The relationships developed earlier between SNR, P_d and P_{fa} apply to a single pulse only. However, it is possible to improve the radar performance by averaging a number of returns.

For example, as a search-radar beam scans, the target will remain in the beam sufficiently long for more than one pulse to hit it. This number, known as hits per scan, can be calculated as follows

$$n_b = \frac{\theta_b f_p}{\dot{\theta}_s} = \frac{\theta_b f_p}{6\omega_m}, \tag{10.18}$$

where n_b – Hits per scan,
θ_b – Azimuth beamwidth (deg),
$\dot{\theta}_s$ - Azimuth scan rate (deg/s),
ω_m – Azimuth scan rate (rpm).

For a typical ground based radar with an azimuth beamwidth or 1.5°, a scan rate of 5rpm and a pulse repetition frequency of 30Hz, the number of pulses returned from a single point target is 15.

The process of summing all these hits is called integration, and it can be achieved in many ways some of which were discussed in Chapter 5. If integration is performed prior to the envelope detector, it is called *pre-detection* or *coherent* integration, while if integration occurs after the detector, it is called *post-detection* or *non-coherent* integration.

Pre-detection integration requires that the phase of the signal be preserved if the full benefit of the summing process is to be achieved and because phase information is destroyed by the envelope detector, post-detection integration, though easier to achieve, is not as efficient.

If n pulses are perfectly integrated by a coherent integration process, the integrated SNR will be exactly n times that of a single pulse in white noise. However, in the non-coherent case, though the integration process is as efficient, there are the detector

losses discussed earlier that reduce the effective SNR at the output of the envelope detector.

Integration improves the P_d and reduces the P_{fa} by reducing the noise variance and thus narrowing the Noise and Signal + Noise PDFs as shown in the figure below

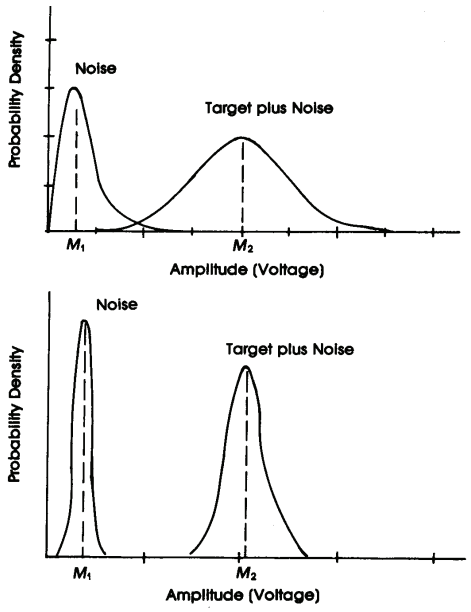


Figure 10.6: Effect of integration on signal and noise PDFs before and after integration

For n pulses integrated, the single pulse SNR required to achieve a given P_d and P_{fa} will be reduced. However, this results in increased detector losses, and hence a reduced effective integration efficiency.

The integration efficiency may be defined as

$$E_i(n) = \frac{SNR(1)}{nSNR(n)}, \tag{10.19}$$

where: $E_I(n)$ – Integration efficiency,
$SNR(1)$ – Single pulse SNR required to produce a specific P_d if there is no integration,
$SNR(n)$ – Single pulse SNR required to produce a specific P_d if n pulses are integrated perfectly.

The improvement in SNR if n pulses are integrated, post detection is thus $nE_I(n)$. This is the integration improvement factor, or the effective number of pulses integrated.

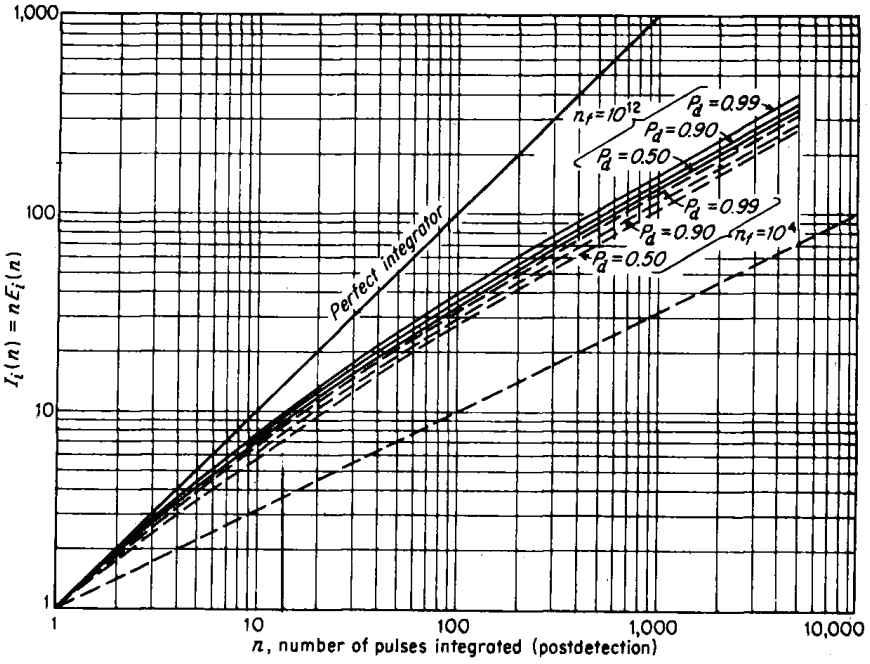


Figure 10.7: Integration improvement factor as a function of pulses integrated

The integration loss in dB is defined as follows

$$L_i(n) = 10\log_{10}\left(\frac{1}{E_i(n)}\right). \tag{10.20}$$

The integration improvement factor is not a sensitive function of either the P_d or the P_{fa} as can be seen by the clustering of the curves in the figure above.

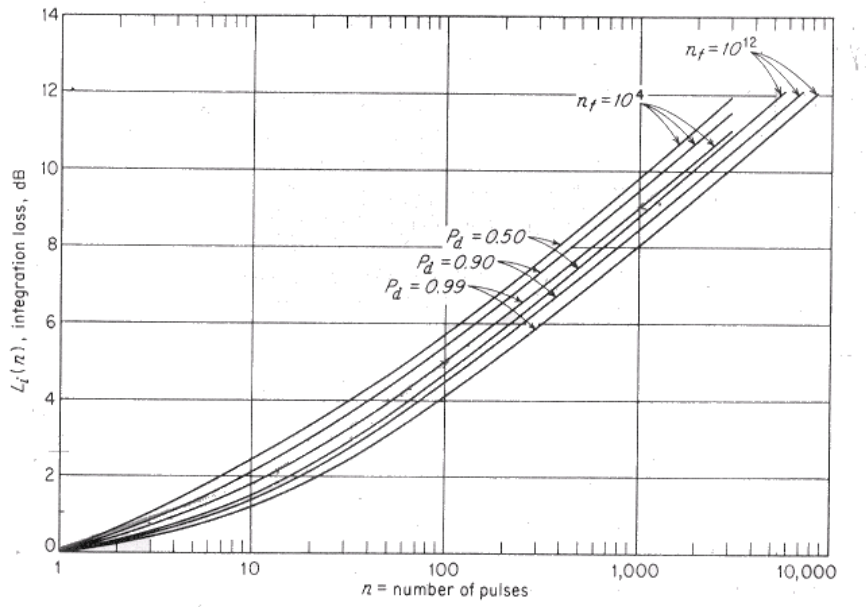


Figure 10.8: Integration loss as a function of the number of pulses integrated

10.8.Detection of Fluctuating Signals

The discussion in the previous section assumes that the signal amplitude does not vary from pulse to pulse during the integration period. However, from the discussion of target cross section in Chapter 9, it is obvious that the RCS of any moving target (with the exception of a sphere) will fluctuate with time as the target aspect as seen by the radar changes.

To properly account for these fluctuations, both the probability density function and the correlation properties with time must be known for a particular target and trajectory. Ideally, these characteristics should be measured for a target, but this is often impractical. An alternative is to postulate a reasonable model for the target fluctuations and to analyse the effects mathematically.

Four fluctuation models proposed by Swerling are used:

- **Swerling 1:**Echo pulses received from the target on any one scan are of constant amplitude throughout the scan, but uncorrelated from scan to scan. The PDF is given by

$$p(\sigma) = \frac{1}{\sigma_{av}} \exp \frac{-\sigma}{\sigma_{av}}, \tag{10.21}$$

  where σ_{av} is the average cross section over all target fluctuations.

- **Swerling 2:** The PDF is as for case 1, but the fluctuations are taken to be independent from pulse to pulse.
- **Swerling 3:** The fluctuations are independent from scan to scan, but the PDF is given by

$$p(\sigma) = \frac{4\sigma}{\sigma_{av}^2} \exp \frac{-2\sigma}{\sigma_{av}}. \tag{10.22}$$

- **Swerling:** The PDF is as for case 3, but the fluctuations are independent from pulse to pulse.
- **Swerling 5:** Non fluctuating

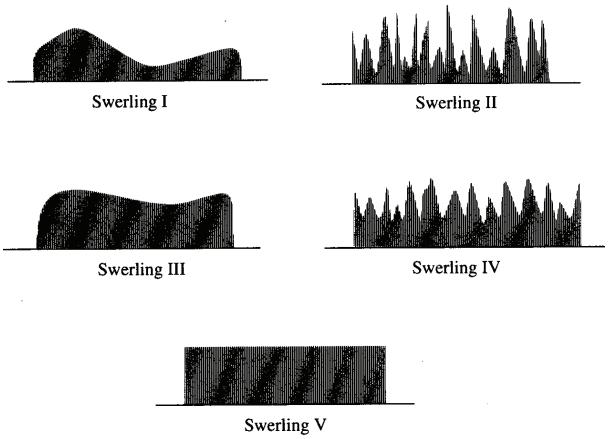


Figure 10.9: Radar returns for different Swerling fluctuations

The PDF for cases 1 and 2 is indicative of a target with many (>5) scatterers of equal amplitude. These are typical for complicated targets like aircraft, while the PDF for cases 3 and 4 is indicative of a target with one large scatterer and many small scatterers.

As one would expect, the single pulse SNR required to achieve a particular P_d (for $P_d > 0.4$) will be higher for a fluctuating target than for a constant amplitude signal. However, for $P_d < 0.4$, the system takes advantage of the fact that a fluctuating target will occasionally present echo signals larger than the average, and so the required SNR is lower.

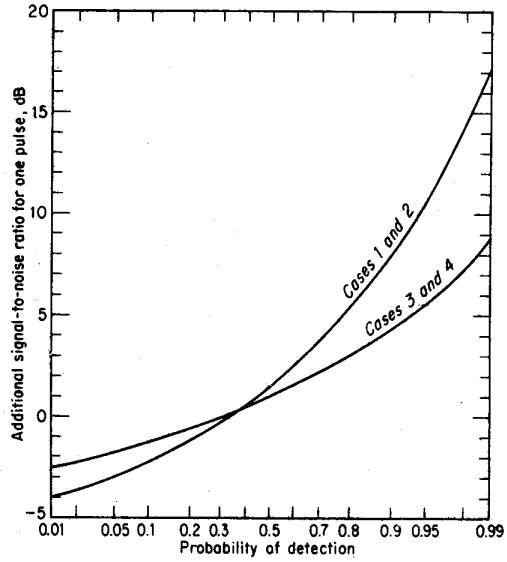


Figure 10.10: Effect of target fluctuation on required signal to noise ratio

To cater for fluctuating targets and integration, a further set of curves describing the integration improvement factor have been developed.

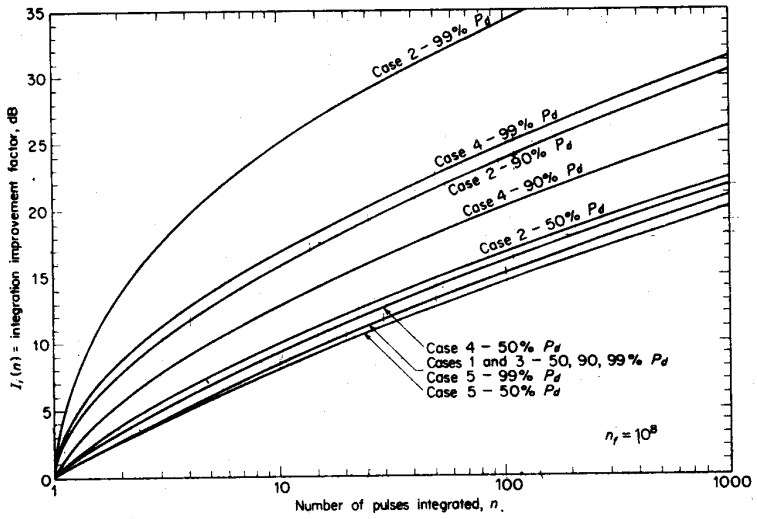


Figure 10.11: Integration improvement factor as a function of the number of pulses integrated for fluctuating targets

If these curves are examined in isolation, it would appear that the integration efficiency $E_I(n) > 1$ under certain conditions. One is not getting something for nothing as the single pulse SNR is much higher in these cases than it would be for the single pulse case.

The procedure for using the range equation when one of these Swerling targets is as follows:

1. Find the SNR for the single pulse, non-fluctuating case that corresponds to the P_d and P_{fa} required using the curves in Figure.10.4
2. For the specific Swerling target, find the additional SNR needed for the required P_d. using the curves in Figure 10.10
3. If n pulses are to be integrated, the integration improvement factor $I_I(n)=nE_I(n)$ is then found using the curves in Figure 10.11
4. The $SNR(n)$ and $nE_I(n)$ are substituted into the range equation along with σ_{av} and the detection range found.

10.9.Constant False Alarm Rate (CFAR) Processors

As can be seen from the tail in the Rayleigh distributed noise power, the false alarm rate is very sensitive to the setting of the detection threshold voltage. Changes in radar characteristics with time (ageing) and changes in the target background characteristics mean that a fixed detection threshold is not practical and so adaptive techniques are required to maintain a constant false alarm rate irrespective of the circumstances. These are called Constant False Alarm Rate (CFAR) processors.

The implementation of CFAR is not difficult, but in general the actual performance will be a function of the target, clutter and noise statistics.

For aircraft these are not a problem as the area around the craft is generally empty so good background statistics can be obtained. However for ground targets, the CFAR threshold is determined from the clutter statistics, which may not be homogeneous. This is made worse by the fact that targets (tanks etc) often hide at the edges of clutter boundaries to reduce the probability of detection.

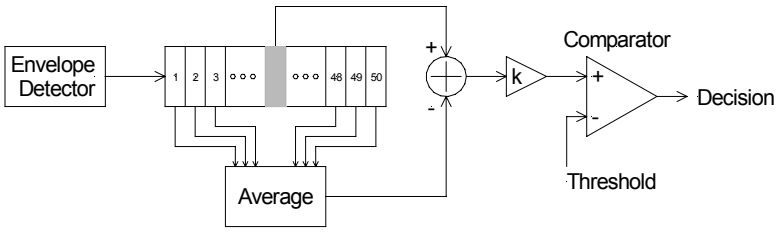


Figure 10.12: Multiple cell averaging CFAR

Needless to say, target detection using the CFAR process introduces additional losses as the statistics are incompletely characterised. For example a cell averaging process will exhibit a 3.5dB loss (compared to an ideal single pulse detector) if 10 cells are used in broadband noise or clutter with a Rayleigh PDF. This decreases to 1.5dB for 20 cells and 0.7dB for 40 cells. The loss decreases with increasing numbers of pulses integrated. For a 10 cell CFAR, with 10 pulses integrated the loss is only 0.7dB and it reduced to only 0.3dB for 100 pulses integrated.

CFAR can operate across range cells, cross-range cells or both, as shown in the figure below.

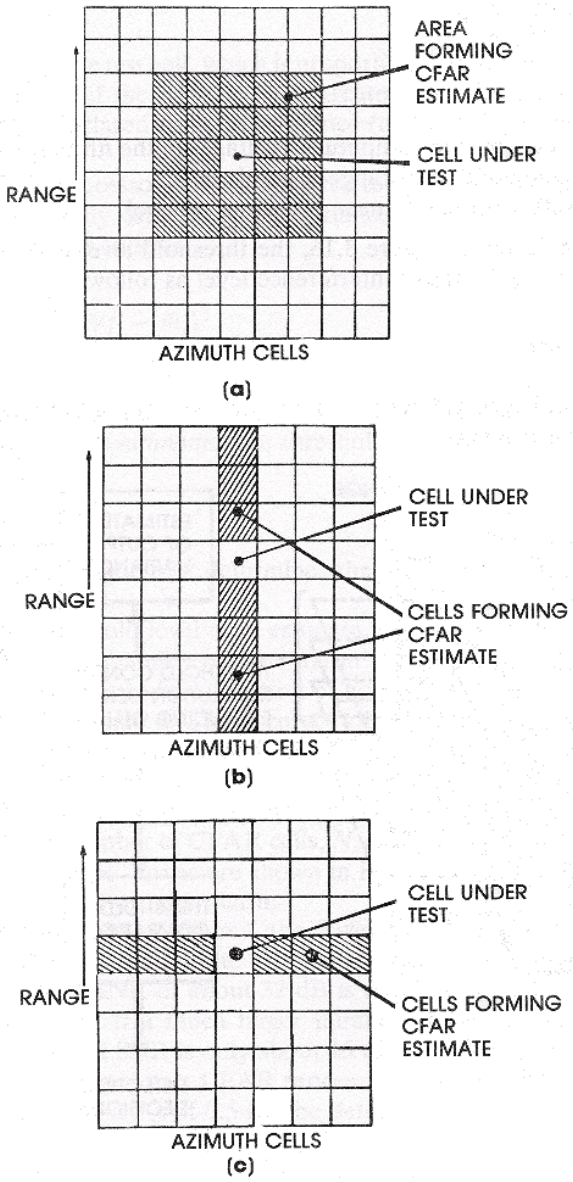


Figure 10.13: Constant false alarm rate options (a) area CFAR, (b) range Only CFAR and (c) azimuth angle only CFAR

In the following examples where a 25+25 point moving average is used to determine the mean value of the signal, it can be seen that the CFAR process detects the two targets without detecting the noise. However, a fixed threshold of 1.8 would perform just as well.

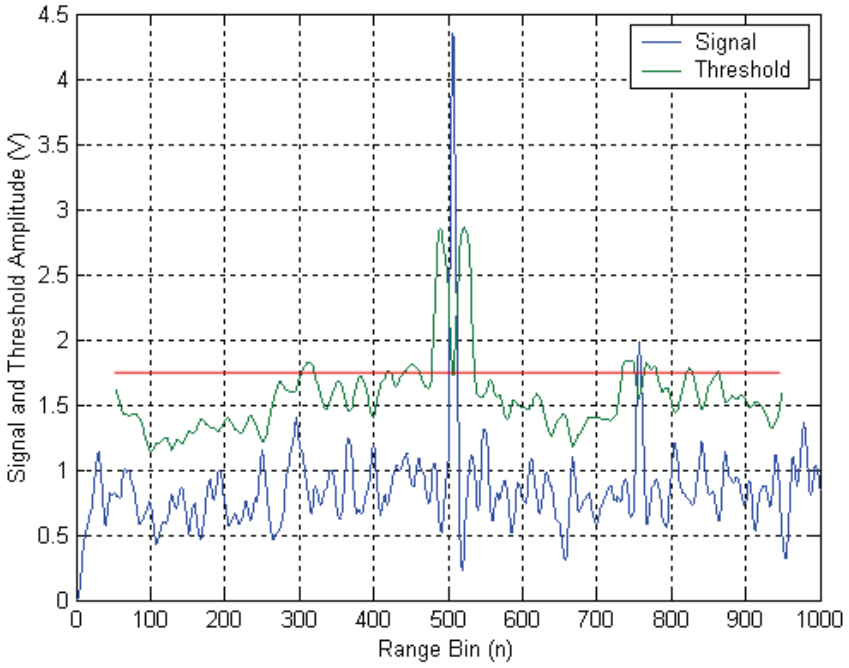


Figure 10.14: CFAR performance for a constant noise floor

In many radar systems, the noise floor varies with range, as shown in the following example. In this instance a fixed threshold does not perform adequately, but the CFAR process still does.

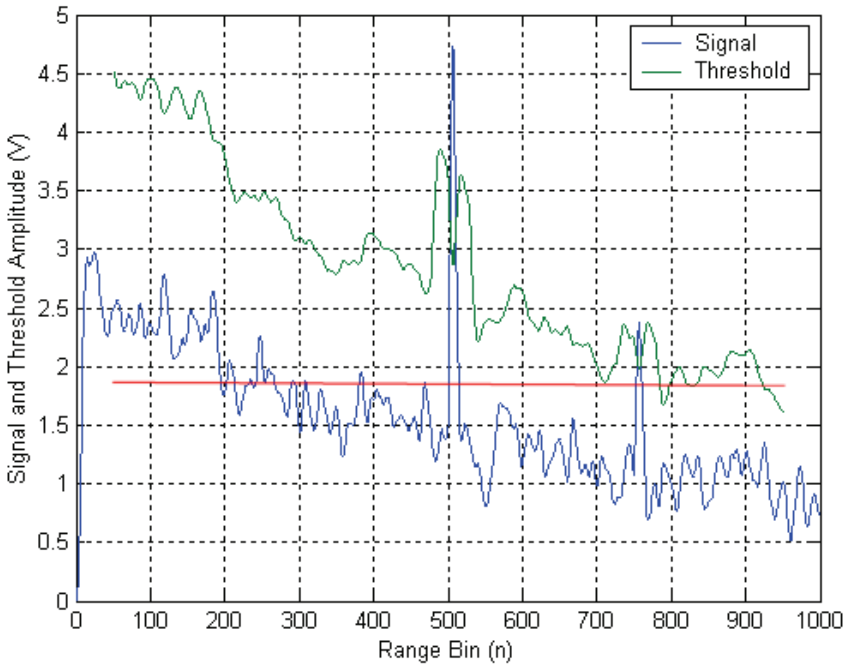


Figure 10.15: CFAR performance for a sloping noise floor

10.10. Air Traffic Control Radar Performance



Type:	2D air surveillance radar
Band:	L
Frequency:	1250 to 1350MHz
Peak Power:	5MW
Antenna size:	12.8×6.7m
Antenna Gain:	36dB (lower beam) 34.5dB (upper beam)
Beam shape:	Cosec^2
Elev beamwidth:	4° Cosec^2 to 40°
Azim beamwidth:	1.25°
Scan:	Mechanical
Scan Rate:	6rpm
PRF:	360pps
PRF Stagger:	Quadruple
Pulse width:	2μs
Noise Figure:	4dB

Calculate the theoretical detection range for a $1m^2$ aircraft target if the detection probability $P_d = 0.9$, and the mean time between false alarms is 9 hours.

Matched Filter Assumptions:	Rectangular Pulse Second order bandpass filter Loss 0.56dB $B.\tau = 0.613$

Hits per Scan: Use the formula $n_b = \frac{\theta_b f_p}{6\varpi_m} = \frac{1.25 \times 360}{6 \times 6} = 12.5$ [use $n_b = 10$]

False Alarm Probability: Use the formula $P_{fa} = \frac{1}{T_{fa}\beta}$

From the matched filter assumptions $B.\tau = 0.613$ and the pulsewidth $\tau = 2\mu s$, the IF bandwidth $\beta = 306kHz$.

$T_{fa} = 9$ hrs $= 32400s$

$$P_{fa} = \frac{1}{32.4 \times 10^3 \times 306 \times 10^3} = 10^{-10}$$

Single Pulse SNR: Use the curves in Figure.10.4 for $P_d = 0.9$ and $P_{fa} = 10^{-10}$

$SNR(1) = 15.2dB$

Fluctuating Target: Use the curves in Figure.10.10 for Swerling 2 for an aircraft

Additional SNR required 8dB

Pulse Integration: Use the curves in Figure.10.11 for 10 pulses integrated, a Swerling 2 target and $P_d = 90\%$ to obtain an improvement factor of 15dB

Total n Pulse SNR Required: Add up the requirements

$SNR(10) = 15.2+8-15 = 8.2dB$

Applying the Radar Range Equation

We assume the following losses:
Transmitter Line = L_{tx} = 2dB (incorporated into Tx power)
Receiver Line = L_{rec} = 2dB (incorporated into receiver noise figure)

Losses:

1D Scanning Loss = 1.6dB
Matched Filter = 0.56dB
CFAR Loss = 0.7dB
Misc. Loss = 1.3dB
Total L = 1.6+0.56+0.7+1.3 = 4.16dB

$$P_r = \frac{P_t G^2 \lambda^2 \sigma}{(4\pi)^3 R^4 L}$$

Transmitter Power: $10\log_{10}(5\times10^6) = 67\text{dBW}$

Less $L_{tx} = 2\text{dB}$
Radiated peak power $P_t = 65\text{dBW}$

Antenna gain: (Lower beam) 36dB

Radar cross section: $\sigma = 10\log_{10}(1) = 0\text{dBm}^2$

Constant: $10\log_{10}\frac{\lambda^2}{(4\pi)^3} = -45.7\text{dB}$

Received Power: Calculated from the range equation

$P_r = 65+36+36-45.7+0-4.16-40\log_{10}R$

$= 87.14-40\log_{10}R \quad \text{dBW}$

Receiver Noise: Use the equation $N = P_N F_N = kT_{sys}BF_N$

K = Boltzmann's Constant $1.38\text{x}10^{-23}$ J/deg

T_{sys} = Temperature 290K

$B = 306\times10^3$ Hz

We are given the receiver noise figure in dB

$10\log_{10}(F_n) = 4\text{dB}$

$$N = 10\log_{10}(kT_{sys}B) + 10\log_{10}(F_n) + L_{rec} = -149 + 4 + 2 = -143dBW$$

Received Signal to Noise Ratio

$$SNR_{rec} = P_r - N = 87.14 - 40\log_{10}R + 143.0 = 230.1 - 40\log_{10}R$$

This must equal the required single pulse SNR if 10 pulses are integrated $SNR(10)$ to achieve the specified P_d and P_{fa}.

$8.2 = 230.1 - 40\log_{10}R$

Solving for the Detection Range

Solve for R

$R = 352.8\text{km}$

Atmospheric Attenuation

The atmospheric attenuation α_{dB} at L-and is about 0.003dB/km (one way)

So the total two way attenuation over 353km is 2.1dB which will result in a significant reduction of the detection range.

These equations are best solved graphically using MATLAB.

The graph below shows two received power curves, one that does not take into account the atmospheric attenuation which intersects the *SNR*(10) curve at 353km.

The other graph which takes attenuation into account intersects the SNR(10) curve at 320km.

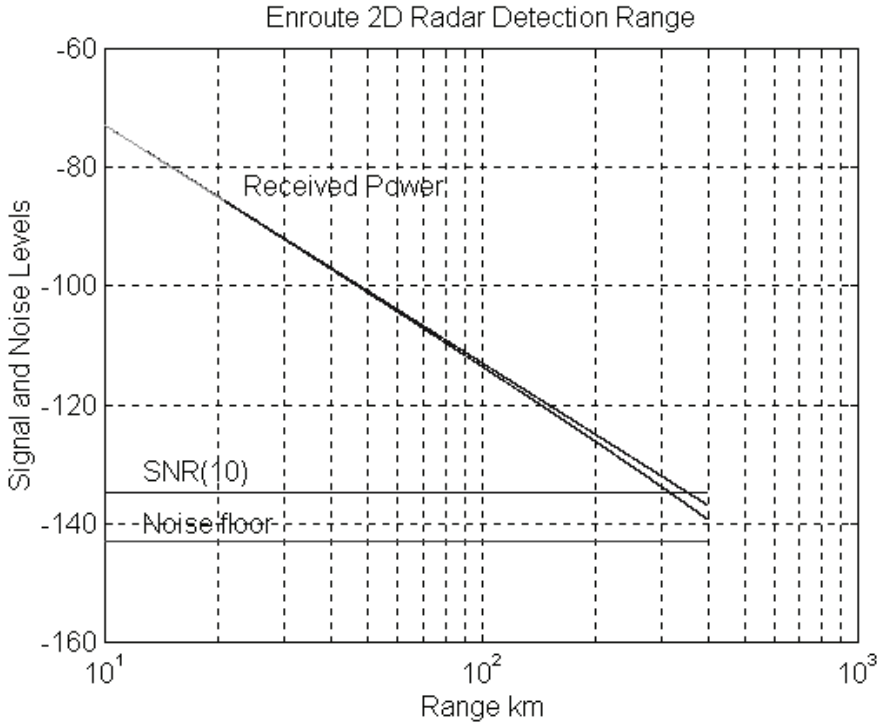


Figure 10.16: Graphical solution to the radar range equation including the atmospheric loss

10.10.1. Analysis Software

Radar performance analysis software based on Blake has been available for many years. In this case a package called RGCALC is used.

For the 2D radar it provides the following results.

```
Radar Name or Description -- Enroute 2D Air Surveillance Radar

     Radar and Target Parameters (inputs) --

     Peak Pulse Power (kilowatts) .................      5000.0
     Pulse Duration (usec) ........................      2.0000
     Transmit Antenna Gain (dB) ...................        36.0
     Receive Antenna Gain (dB) ....................        36.0
     Frequency (MHz) ..............................      1300.0
     Receiver Noise Factor (dB) ...................         4.0
     Bandwidth Correction Factor (dB) .............          .6
     Antenna Ohmic Loss (dB) ......................          .0
     Transmit Transmission Line Loss (dB) .........         2.0
     Receive  Transmission Line Loss (dB) .........         2.0
     Scanning-Antenna Pattern Loss (dB) ...........         1.6
     Miscellaneous Loss (dB) ......................         2.0
     Number of Pulses Integrated ..................          10
     Probability of Detection .....................        .900
     False-Alarm Probability (Negative Power of Ten)       10.0
     Target Cross Section (Square Meters) .........      1.0000
     Target Elevation Angle (Degrees) .............         .40
     Average Solar and Galactic Noise Assumed
     Pattern-Propagation Factors Assumed = 1

            **********************************

     Calculated Quantities (Outputs) --

     Noise Temperatures, Degrees Kelvin --
            Antenna (TA) ..........................      117.2
            Receiving Transmission Line (TR) ......      169.6
            Receiver (TE) .........................      438.4
            TE X Line-Loss Factor = TEI ...........      694.9
            System (TA + TR + TEI) ................      981.7
     Two-Way Attenuation Through Entire Troposphere (dB)    3.1

      Swerling     Signal-     Tropospheric  Range,    Range,
      Fluctuation  to-Noise    Attenuation,  Nautical  Kilometers
      Case         Ratio, dB   Decibels      Miles
      ---------    ----------  -----------   --------  ----------

          0           6.76         2.77        163.7       303.2

          1          15.28         2.13        104.0       192.6

          2           7.91         2.70        153.9       285.0

          3          11.33         2.45        128.2       237.4

          4           7.38         2.73        158.3       293.2
```

Note that the predicted performance using RGCALC for a Swerling 2 target is only 285km, this lower range is because it makes the assumption that the optimum matched filter has $B.\tau = 1$, and hence calculates a wider IF bandwidth by a factor of 1/0.613 than was made above.

This adds 2.1dB to the noise floor which accounts for the difference

10.11. References

[1] D.Barton, *Modern Radar Systems Analysis,* Artech, 1988
[2] M.Skolnik, *Introduction to Radar Systems,* McGraw Hill, 1980
[3] N.Currie (ed), *Principles and Applications of Millimeter-Wave Radar,* Artech, 1987
[4] Radar Detection, http://www.interscience.wiley.com:83/eeee, 09/03/2001.
[5] RGCALC Range Analysis Software
[6] B.Mahafza, *Radar Systems Analysis and Design Using MATLAB,* ChapMan & Hall, 2000

Chapter 11.

High Range-Resolution Techniques

11.1.Classical Modulation Techniques

The range resolution of a sensor is defined as the minimum separation (in range) of two targets or equal cross section that can be resolved as separate targets. It is determined by the bandwidth of the transmitted signal. The bandwidth, Δf, is generated by widening the transmitter bandwidth using some form of modulation

- Amplitude modulation
- Frequency modulation
- Phase modulation

11.2.Amplitude Modulation

A special case of the amplitude modulation technique is the classical pulsed radar where the amplitude is 100% for a very short period, and 0% the remaining time

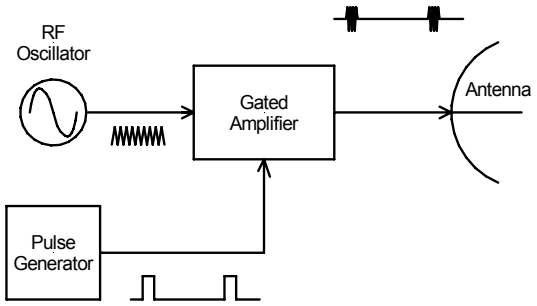


Figure 11.1: On-off amplitude modulation of a sine wave to produce pulses

11.2.1. Range Resolution

The range resolution is determined from the matched filter processing of the rectangular pulse. Consider the case that the transmitted signal consists of a constant frequency signal modulated by a rectangular pulse of width, τ. The sharp edges of the rectangular function in time generate an infinite frequency spectrum, a truncated version of which is shown in the figure below.

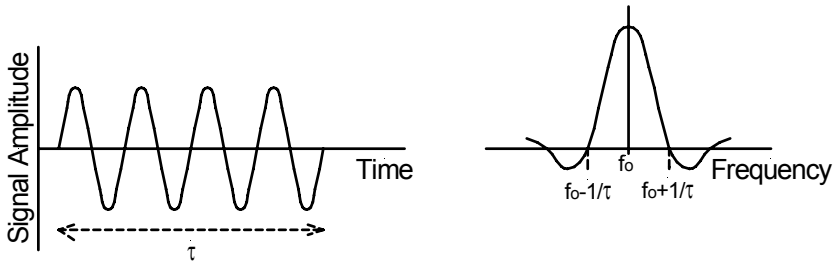


Figure 11.2: The relationship between the waveform and spectrum of a rectangular pulse

It can be seen from the frequency response that the 3dB (50%) bandwidth is just

$$\Delta f \approx \frac{1}{\tau} \tag{11.1}.$$

When a rectangular pulse is processed through a perfect matched filter (correlator) it produces a triangular output envelope 2τ wide at the base and with a well defined central peak as shown below.

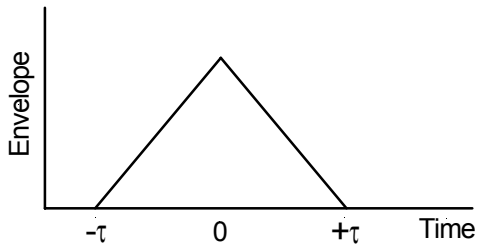


Figure 11.3: Matched-filter output for a generic rectangular pulse of duration τ

A second return is displaced from the first by a time delay of τ seconds. Depending on the relative phases of the two targets, the response envelope can take on any of the shapes shown below. For all phase angles the second peak is still identifiable and the two targets are said to be resolved in range.

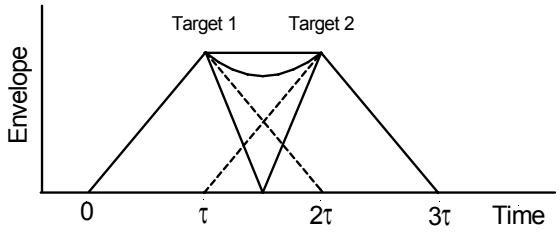


Figure 11.4: Matched-filter output of a pair of closely spaced targets showing the limits to the range resolution

The range resolution is determined by converting the time delay, τ, to the round trip time required to achieve that delay

$$\delta R = \frac{c\tau}{2} \tag{11.2}.$$

Using the relationship shown in Figure 11.2, the range resolution determined in terms of the pulse width can be rewritten in terms of the effective bandwidth of the signal

$$\delta R = \frac{c}{2\Delta f} \qquad (11.3).$$

Narrow pulse systems require large peak power (>10 MW) for long range operation and so special precautions must be taken to minimise the problems of ionisation and arcing within the waveguide for radar systems, or in the air for high power lasers. This makes it advantageous to generate a transmitted waveform that decouples the range resolution from the duration of the pulse.

11.3.Frequency & Phase Modulation

The inability of a fixed-frequency continuous-wave CW radar to resolve range is related to the narrow spectrum of its transmitted waveform. Frequency and phase modulation of the carrier are the most common techniques used to broaden this spectrum. Solutions involve lengthening the pulse to achieve large radiated energy, while still maintaining the wide bandwidth for good range-resolution. The received signal can then be processed using a matched filter that compresses the long pulse to a duration $1/\Delta f$.

The time-bandwidth product $\Delta f.\tau$ of the uncompressed pulse is used as a figure of merit for such "pulse compression" systems.

11.3.1. Matched Filter

A pulse-compression radar is the practical implementation of a matched-filter system as shown schematically in the figure below.

The coded signal can be described either by the frequency response $H(\omega)$ or as an impulse response $h(t)$ of the coding filter. The received echo is fed into a matched filter whose frequency response is the complex conjugate $H^*(\omega)$ of the coding filter. The output of the matched filter, $y(t)$ is the compressed pulse which is just the inverse Fourier transform of the product of the signal spectrum and the matched filter response

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 \exp(j\omega t) d\omega \,. \qquad (11.4)$$

A filter is also matched if the signal is the complex conjugate of the time inverse of the filter's impulse-response. This is often achieved by applying the time inverse of the received signal to the pulse-compression filter.

The output of this matched filter is given by the convolution of the signal $h(t)$ with the conjugate impulse response $h^*(-t)$ of the matched filter

$$y(t) = \int_{-\infty}^{\infty} h(\tau) h^*(t-\tau) d\tau \,. \tag{11.5}$$

In essence the matched filter results in a correlation of the received signal with a delayed version of the transmitted signal as shown in Figure 11.5c below.

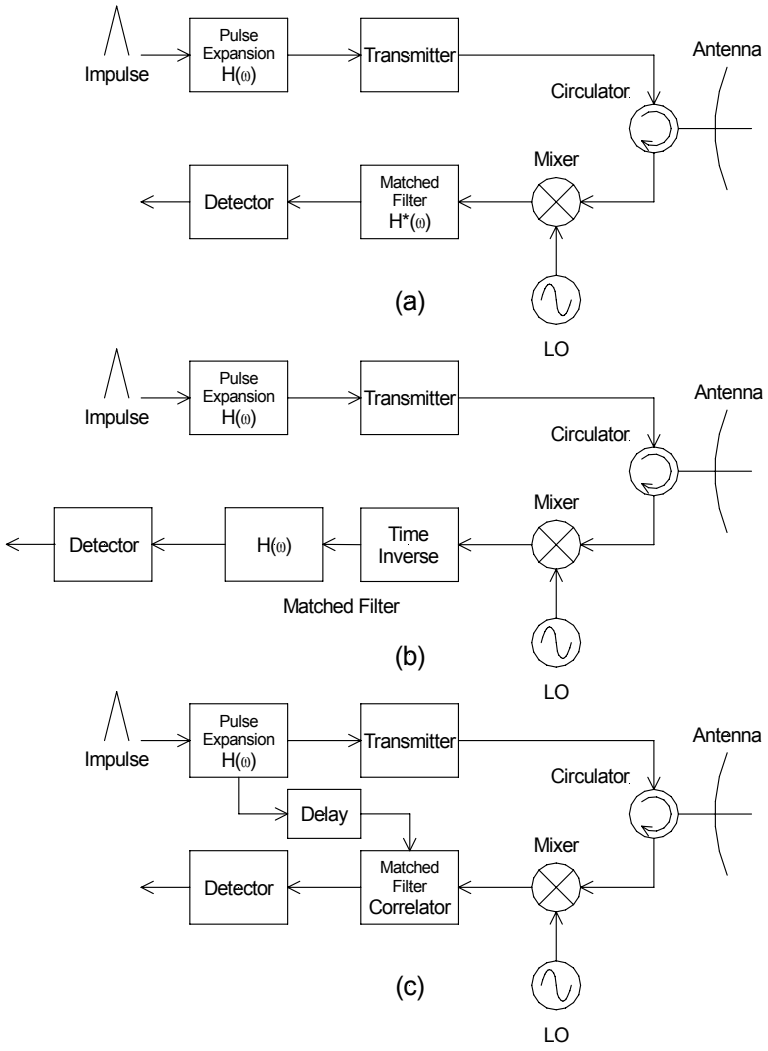


Figure 11.5: Matched-filter configurations for pulse compression using (a) conjugate filters, (b) time inversion and (c) correlation

The effects of this form of processing on two pulses with the same duration are shown in the following figure. In the continuous frequency (CF) example, the matched filter (correlation) response shows the triangular envelope described earlier. However, in the chirp example with the same duration, the matched filter generates a sinc function with a much narrower peak, and hence a superior range resolution. It is shown later in this chapter that the range resolution is inversely proportional to the chirp bandwidth, Δf.

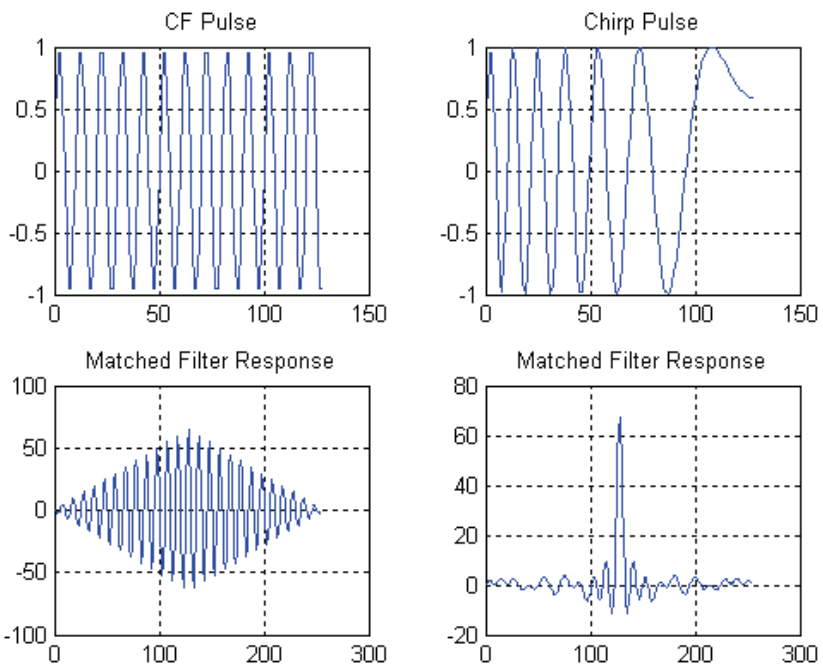


Figure 11.6: Comparison between the ultimate resolution of a rectangular constant frequency pulse and a chirp pulse of the same duration

11.4.Phase-Coded Pulse Compression

A simple way to understand pulse compression with a matched filter is to consider the binary phase-shift keying (BPSK) modulation technique. In this modulation the code is made up of *m* chips which are either in-phase, 0° (positive), or out-of-phase, 180° (negative), with a reference signal as shown in the figure below.

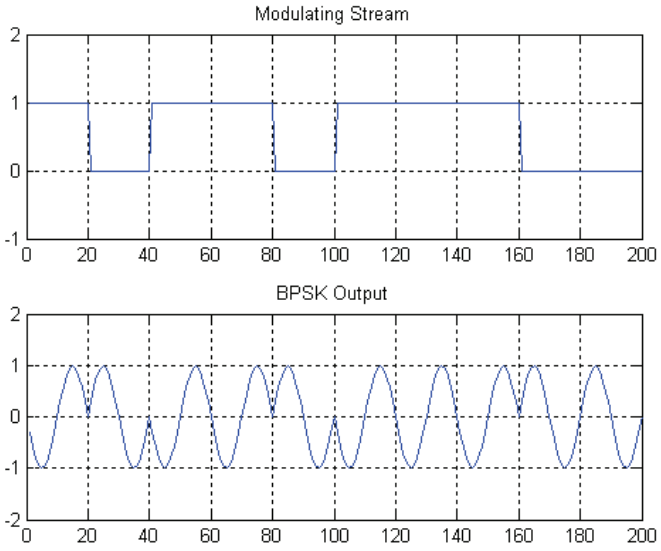


Figure 11.7: Example of binary phase shift keying using one cycle per bit

Demodulation is achieved by multiplying the incoming RF signal by a coherent carrier (a carrier that is identical in frequency and phase to the carrier that originally modulated the BPSK signal). This produces the original BPSK signal plus a signal at twice the carrier which can be filtered out. However, a more common technique that is used widely by the radar fraternity is shown in the figure below.

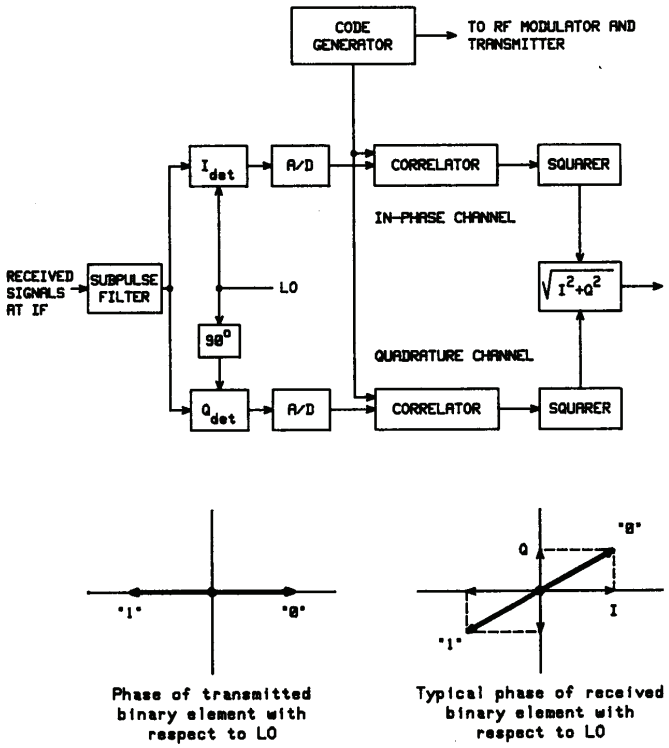


Figure 11.8: BPSK receiver & demodulator

The received signals are bandpassed by a filter matched to the data rate, the outputs are then demodulated by I and Q detectors. These detectors compare the phase of the received signal to the phase of the Local Oscillator which is also used in the RF modulator.

Though the phase of each of the transmitted signals is 0° or 180° with respect to the LO, on receive the phase will be shifted by an amount dependent on the round trip time and the Doppler velocity. For this reason, two processing channels are generally used, one which recovers the in-phase signal and one which recovers the quadrature signal.

These signals are converted to digital by the Analog to Digital (A/D) converters, correlated with the stored binary sequence and then combined.

The primary advantage of this configuration is that it utilises the coherence of the system to produce two quadrature receive channels. If only one channel is implemented, then there is a loss in effective signal to noise ratio (SNR) of 3dB

The echo is compressed by correlation with the stored reference which is the discrete equivalent of the matched filter process described earlier.

Special cases of these binary codes are the Barker codes where the peak of the autocorrelation function is N (for a code of length N) and the magnitude of the maximum peak sidelobe is 1. The problem with the barker codes is that none with lengths greater than 13 have been found.

Barker code sequences are called optimum, because, for zero Doppler shift, the peak to sidelobe ratio is +/-n after matched filtering (where n is the number of bits).

Table 11.1: Barker code sequences

Code length	Code Elements	Sidelobe Level (dB)
2	+- or ++	-6
3	++-	-9.5
4	++-+ or +++-	-12
5	+++-+	-14
7	+++--+-	-16.9
11	+++---+--+-	-20.8
13	+++++--++-+-+	-22.3

A five chip Barker code, +++-+, will have a filter matched to the chip length τ_c with bandwidth $\beta = 1/\tau_c$, which will take the classic $(sin\ x)/x$ transfer function shown in Chapter 2. This is followed by a tapped delay line having four delays τ_c, the outputs of which are weighted by the time reversed code +-+++ and summed prior to envelope detection as shown below.

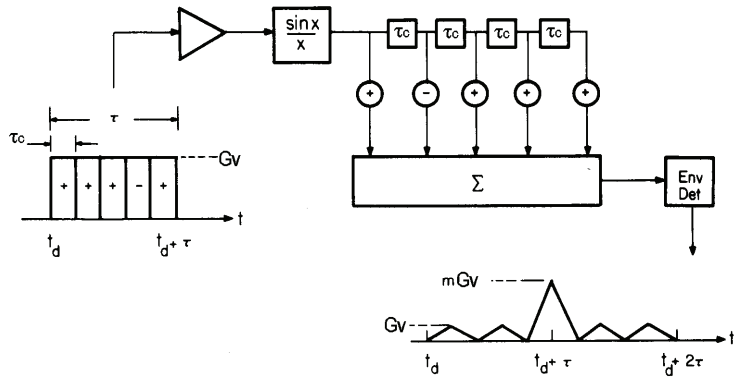


Figure 11.9: Diagram to illustrate the concept of phase-coded pulse compression for a five bit Barker code

The output consists of m-1 time sidelobes of unit amplitude G_v and a main lobe with amplitude mG_v each of width τ_c. The ratio of the transmitted pulsewidth to the output pulsewidth is $\tau/\tau_c = \beta\tau$ which is the pulse-compression ratio. The relative sidelobe power level is $1/m^2$ = -13dB.

The Barker code is the only code that has equal sidelobes at this low level, but this only applies along the zero-Doppler axis. If the target that produced this echo pulse is moving toward or away from the radar, then the phase of the echo will change due to the changing range. This will change the phase relationship between the chips on the

expanded pulse and modify the way that they combine on compression. Examination of the ambiguity diagram for a 13-bit Barker code below, shows that the main-lobe decays quickly and sidelobes increase rapidly with increasing Doppler.

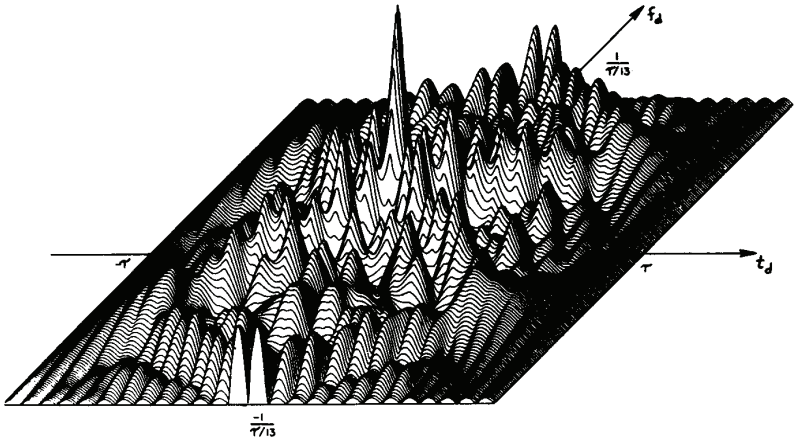


Figure 11.10: Ambiguity diagram for a 13-bit Barker code showing the "thumbtack" main lobe decaying into a sea of increasing delay and Doppler sidelobes

From a resolution perspective, the CW phase-coded techniques offer the same performance as their analogue counterparts. Their performance is limited when trying to transmit and receive simultaneously but it would be possible to develop an interrupted version which might have some merit. Processing still remains a potential problem as the computationally expensive autocorrelation process is required to extract the range information from the return echo.

11.4.1. Pseudo Random Codes

Optimal Binary Sequences

The definition of an optimal binary sequence is one whose peak sidelobe of the aperiodic autocorrelation function is the minimum possible for a given code length.

Most of the optimal codes are found by computer searches, however the search time becomes prohibitively long as N increases, and it is often easier to resort to the use of other non-optimal sequences so long as they posses the desired correlation effects.

Maximal length sequences that are particularly useful are those that can be obtained from linear feedback shift registers. These have a structure similar to random sequences and therefore possess desirable autocorrelation functions. They are often called pseudo-random (PR) or pseudo-noise (PN) sequences.

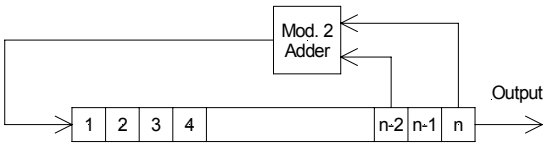


Figure 11.11: Shift register generator

A typical shift register generator is shown in the figure above. The N stages of the register are pre loaded with all 1s or a combination of 1s and 0s (all 0s is not used as it results in an all 0 output). The outputs from specific individual stages of the shift register are summed by modulo-2 addition to form the new input to the shift register.

Table 11.2: Optimal binary codes

Length of Code (N)	Magnitude of Peak Sidelobe	Number of Codes	Code (octal notation for N > 13)
2	1	2	11, 10
3	1	1	110
4	1	2	1101, 1110
5	1	1	11101
6	2	8	110100
7	1	1	1110010
8	2	16	10110001
9	2	20	110101100
10	2	10	1110011010
11	1	1	11100010010
12	2	32	110100100011
13	1	1	1111100110101
14	2	18	36324
15	2	26	74665165
16	2	20	141335
17	2	8	265014
18	2	4	467412
19	2	2	1610445
20	2	6	3731261
21	2	6	5204154
22	3	756	11273014
23	3	1021	32511437
24	3	1716	44650367
25	2	2	163402511
26	3	484	262704136
27	3	774	624213647
28	2	4	1111240347
29	3	561	3061240333
30	3	172	6162500266
31	3	502	16665201630
32	3	844	37233244307
33	3	278	55524037163
34	3	102	144771604524
35	3	222	223352204341
35	3	322	526311337707
37	3	110	1232767305704
38	3	34	2251232160063
39	3	60	4516642774561
40	3	114	14727057244044

Octal	Binary
0	000
1	001
2	010
3	011
4	100
5	101
6	110
7	111

Modulo-2 addition depends only on the number of 1s being added. If it is odd, the sum is 1, and if it is even, the sum is 0. The shift register is clocked, and the output at any stage is the binary sequence. When the feedback connections are properly chosen, the output is a sequence of maximal length N where $N = 2^n$-1, where n is the number of stages of the shift register. There are a total of M maximal length sequences that

can be obtained from a generator with n stages, where M is given by the following formula:

$$M = \frac{N}{n}\Pi\left(1 - \frac{1}{p_i}\right), \tag{11.6}$$

where p_i are the prime factors of N.

The number of different sequences that exist for a given n is important particularly in applications such as collision avoidance where a number of different radar units will be sharing the same area, and the potential for mutual interference exists.

From a radar perspective a BPSK sequence of length N will have a time-bandwidth product of N where the bandwidth of the system is determined by the clock rate. This allows for the generation of large time-bandwidth products (which result in good range resolution) from registers having a small number of stages. By altering the clock rate, the length and feedback connections on the shift register, it is possible to produce, without additional hardware, waveforms of various pulse lengths, bandwidths and time-bandwidth products to suit most radar requirements. The following table lists the length and number of maximal length sequences obtained from shift registers of various lengths along with the feedback connection required to generate one of the sequences

Table 11.3: Maximum length sequences

Number of Stages (n)	Length of Maximal Sequence (N)	Number of Maximal Sequences (M)	Feedback-stage Connections
2	3	1	2,1
3	7	2	3,2
4	15	2	4,3
5	31	6	5,3
6	63	6	6,5
7	127	18	7,6
8	255	16	8,6,5,4
9	511	48	9,5
10	1023	60	10,7
11	2047	176	11,9
12	4095	144	12,11,8,6
13	8191	630	13,12,10,9
14	16383	756	14,13,8,4
15	32767	1800	15,14
16	65535	2048	16,15,13,4
17	131071	7712	17,14
18	262143	7776	18,11
19	524287	27594	19,18,17,14
20	1048575	24000	20,17

If the shift register is left in continuous operation, then a continuous repeating waveform is generated which can be used for continuous-wave operation. Aperiodic waveforms are obtained if the generator output is terminated after one complete sequence, and these are generally used for pulsed-radar applications. The autocorrelation functions of the two cases vary in terms of their sidelobe structure.

Maximal length sequences have characteristics which approach the three characteristics ascribed to truly random processes:

- the number of 1s is approximately equal to the number of 0s
- runs of consecutive 1s and 0s occur with about half the runs having length 1, a quarter of 2, an eighth of 3 etc
- The autocorrelation is thumbtack in nature (peaked in the centre and approaching zero elsewhere)

Maximal length sequences are an odd length, so to make them a power of 2 for processing purposes, a zero is inserted into the start or the end of the sequence. This results in degraded sidelobes.

11.4.2. Correlation

For binary sequences where the values are restricted to +/-1 the following approach is taken

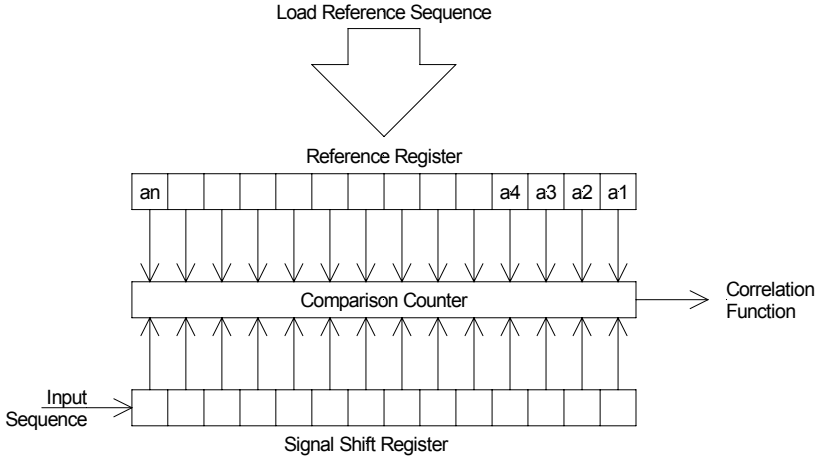


Figure 11.12: Digital correlation

Circular Correlation

For correlation on the two long sequences, the Fourier transforms must be taken, followed by the product of the one series with the complex conjugate of the other, and finally, the inverse Fourier transform completes the procedure as shown in the figure below.

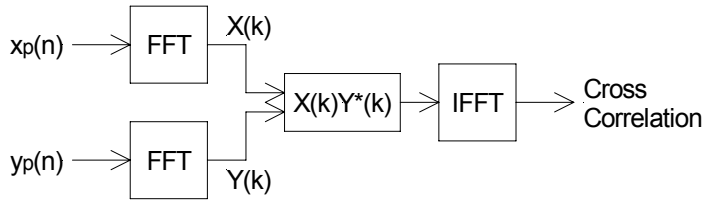


Figure 11.13: Cross correlation using the Fourier transform method

The transmitted sequence is loaded into the reference register, and the input sequence is continuously clocked through the signal shift register. A comparison counter forms a sum of the matches and subtracts the mismatches between corresponding stages of the shift registers on every clock cycle to produce the correlation function.

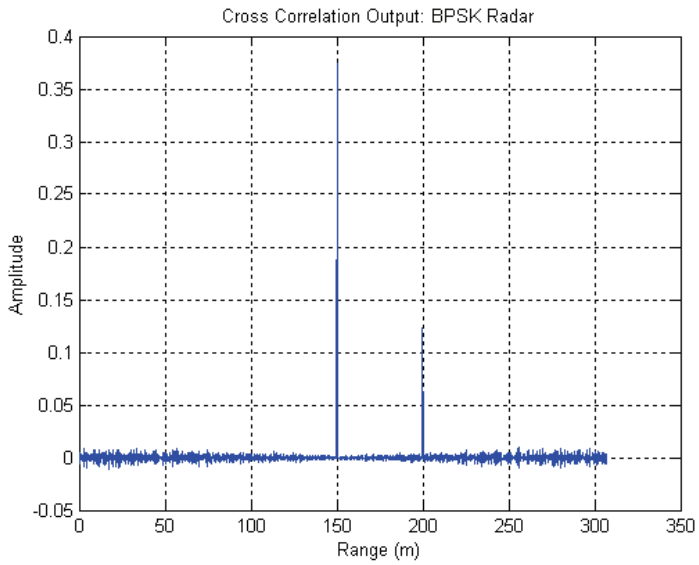


Figure 11.14: Cross correlation showing two targets with different amplitudes and at different ranges: BPSK radar noise sequence generated using a 12bit shift register with 4096 points

As was mentioned earlier, the Maximal Length series must be an odd number, and by padding with zeros degrades the range sidelobe performance. To test this, the correct (unpadded) series was generated and the correct correlation function performed on 4095 points with the following incredible results.

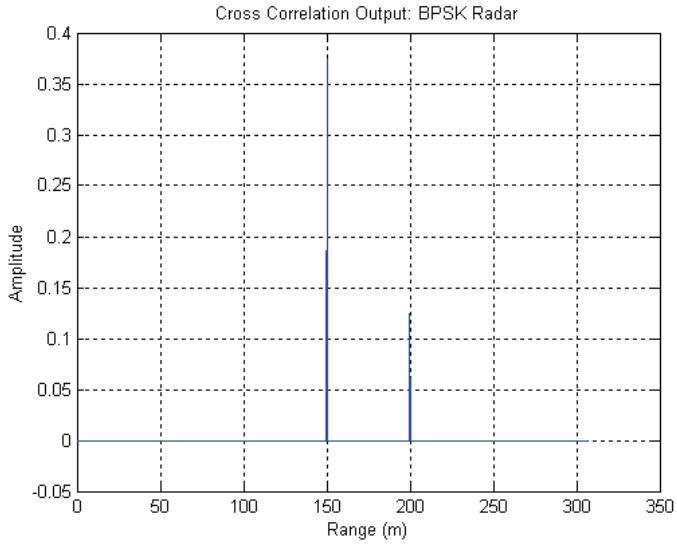


Figure 11.15: Cross correlation showing two targets with different amplitudes and at different ranges: BPSK radar noise sequence generated using a 12bit shift register with 4095 points

If the figure is examined carefully, it can be seen that the sidelobe level is constantly flat with a value slightly smaller than zero (-1.22×10^{-4}).

11.5.SAW Based Pulse Compression

In a pulse-compression system such as that shown below, a very brief pulse consisting of a range of frequencies passes through a dispersive delay-line (SAW expander) in

which its components are delayed in proportion to their frequency. In the process the pulse is stretched, for example a 10ns pulse may be lengthened by a factor of 100 to a duration of 1μs before it is up-converted, amplified and transmitted.

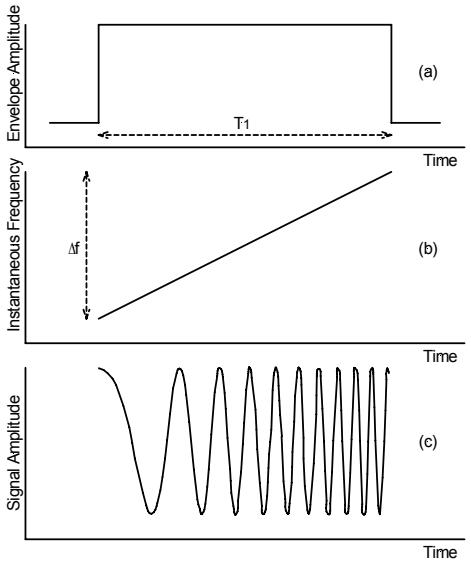


Figure 11.16: Linear chirp pulse (a) transmitter pulse envelope, (b) transmitter pulse frequency and (c) transmitted pulse RF waveform

The echo returns from the target are down-converted and amplified before being fed into a pulse-compression network that retards the echo by amounts that vary inversely with frequency to reduce the signal to its original 10ns length. The compressed echo yields nearly all of the information that would have been available had the unaltered 10ns pulse been transmitted.

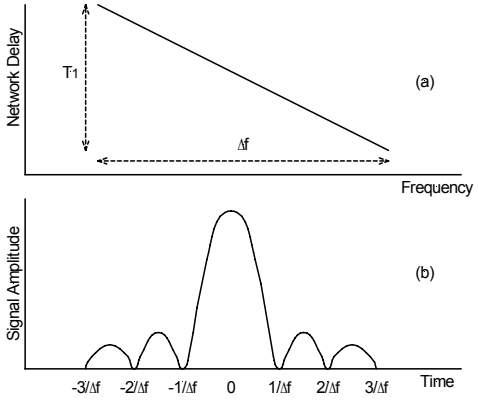


Figure 11.17: Chirp pulse compression characteristics

A slight sacrifice in range resolution (≈1.3) is the penalty incurred in reducing the range sidelobes from –13.2dB with no weighting to –43dB with Hamming weighting.

The SNR gain achieved is approximately equivalent to the pulse time-bandwidth product, $\beta.\tau$. Even though using surface acoustic wave technology to implement the pulse expansion and compression functions limits the maximum $\beta.\tau$ product to about 100, it is the most common method in use because it is both compact and robust.

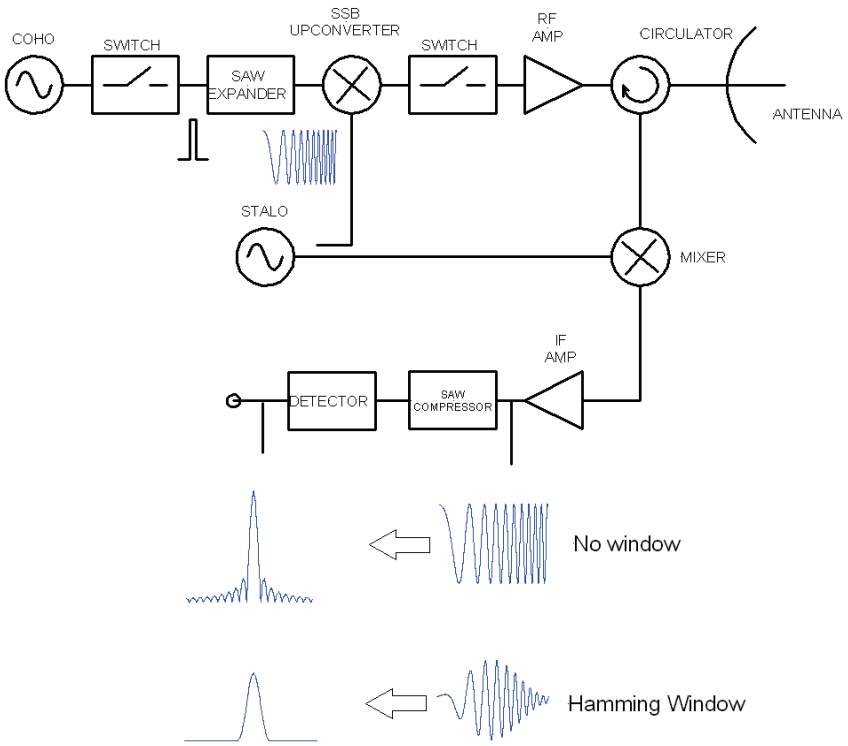


Figure 11.18: Conceptual diagram of a linear-chirp pulse-compression radar

In essence a SAW-based pulse-compression system is similar to an interrupted-FMCW system with the exception that its duty cycle is generally much smaller than 50%. This means that for the same pulse-repetition frequency, the time-bandwidth product, $\beta.\tau$, will be smaller, and hence the SNR gain will be lower for the same transmitter power.

Commercial injection locked amplifiers (ILAs) operating in the millimetre wave band use either pulsed or CW IMPATT diodes with a breakpoint occurring at pulse widths of about 100ns below which powers of up to 22W are available. Unfortunately, at pulsewidths exceeding this value, in the regime that would be required for the imaging radar, output powers are limited to 250mW which are the same as those available for the FMICW configuration. Hence the performance of a SAW-based system would be poorer than that of the FMICW counterpart because $\beta.\tau$ is smaller and the transmit power is the same.

11.6.Step Frequency

Step frequency modulation, also known as step-chirp, provides a piece-wise approximation of the linear chirp signal. It consists of a sequence of different frequencies spaced Δf Hz apart with duration $\tau_f = 1/\Delta f$. The total length of the transmission is

$$\tau_t = \frac{N}{\Delta f} = N\tau_f , \tag{11.7}$$

and the transmitted bandwidth β_t is

$$\beta_t = N\Delta f = \frac{1}{\tau_c}, \tag{11.8}$$

making the pulse-compression ratio

$$\rho = \frac{\tau_t}{\tau_c} = \tau_t \beta_t . \tag{11.9}$$

The step-frequency modulation code is not as Doppler tolerant as the linear FM code. Large grating lobes appear in the compressed pulse sidelobes at Doppler shifts that are odd multiples of $\Delta f/2$. Some techniques including Costas coding, nonlinear FM and amplitude modulation have been developed to improve the sidelobe performance.

The stepped-frequency technique generally relies on a phase-locked oscillator to generate the transmitted signals. This ensures that though the same homodyne process is applied as in the FMCW technique, the magnitude of the phase-noise is lower by a factor of about 30dB at an offset of 100kHz from the carrier.

This method results in improved performance at short range for low transmit power but any attempt to increase the transmitter power significantly can result in degraded performance because of mixer saturation due to transmitter leakage.

If an interrupted version were to be developed, then the frequency would have to remain constant for the round-trip time to the target, and hence the synthesis of the required range resolution that requires N samples would require N times the period required by FMICW to synthesise the same return.

11.7.Frequency Modulated Continuous Wave Radar

11.7.1. Operational Principles

The schematic block diagram below shows the structure of a homodyne radar[3]. In this case, the CW signal is modulated in frequency to produce a linear chirp which is radiated toward a target through an antenna.

[3] a CW radar in which the microwave oscillator serves as both the transmitter and local oscillator

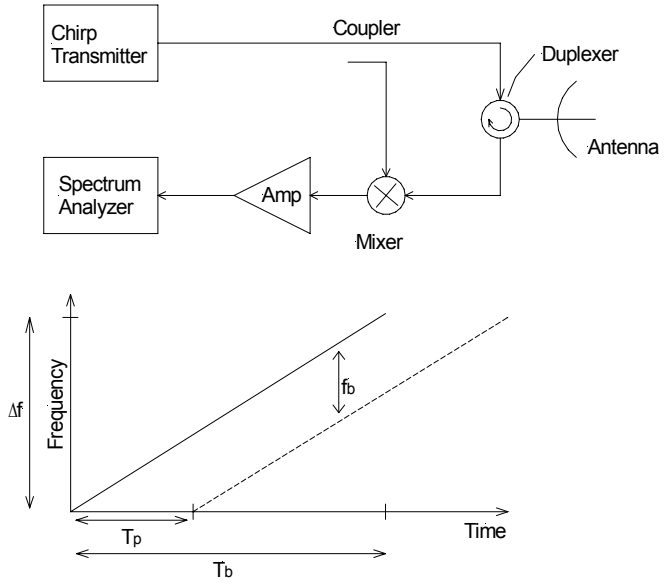


Figure 11.19: Schematic diagram illustrating the FMCW concept

In this diagram, the echo received T_p seconds later by the same antenna is mixed with a portion of the transmitted signal to produce a beat signal at a frequency f_b. From the graphical representation of this process, it is clear that the frequency of this signal will be proportional to the round-trip time T_p. It can be seen that FMCW is just a subset of the standard stretch processing technique in which the LO chirp is equal to the transmitted chirp.

For an analytical explanation, the change in frequency, ω_b, with time or chirp, can be described as

$$\omega_b = A_b t\,, \tag{11.10}$$

substituting into the standard equation for FM results in

$$v_{fm}(t) = A_c \cos\left[\omega_c t + A_b \int_{-\infty}^{t} t dt\right], \tag{11.11a}$$

$$v_{fm}(t) = A_c \cos\left[\omega_c t + \frac{A_b}{2} t^2\right]. \tag{11.11b}$$

This analysis assumes that the frequency continues to increase indefinitely, but in practise the transmitter has a limited bandwidth and the chirp duration is limited.

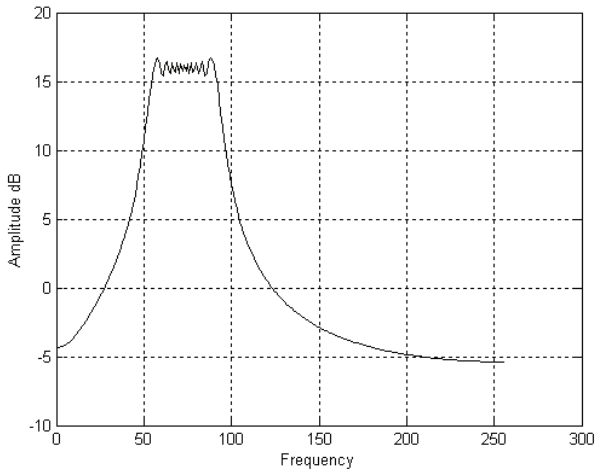


Figure 11.20: Frequency domain representation of a linear FM chirp

In FMCW systems, a portion of the transmitted signal is mixed with the returned echo by which time the transmit signal will be shifted from that of the received signal because of the round-trip time T_p

$$v_{fm}(t-T_p) = A_C \cos\left[\omega_c(t-T_p) + \frac{A_b}{2}(t-T_p)^2\right]. \tag{11.12}$$

Calculating the product of (12.15) and (12.16),

$$v_{fm}(t-T_p)v_{fm}(t) = A_c^2 \cos\left[\omega_c t + \frac{A_b}{2}t^2\right]\cos\left[\omega_c(t-T_p) + \frac{A_b}{2}(t-T_P)^2\right]. \tag{11.13}$$

Equating using the trigonometric identity for the product of two sines,

$$\cos A \cos B = 0.5[\cos(A+B) + \cos(A-B)], \tag{11.14}$$

$$v_{out}(t) = \frac{A_c^2}{2}\left[\begin{array}{l}\cos\left\{\left(2\omega_c - A_b T_p\right)t + A_b t^2 + \left(\frac{A_b}{2}T_p{}^2 - \omega_c T_p\right)\right\} \\ + \cos\left\{A_b T_p.t + \left(\omega_c T_p - \frac{A_b}{2}T_p{}^2\right)\right\}\end{array}\right]. \tag{11.15}$$

The first cosine-term in (11.15) describes a linearly increasing FM signal (chirp) at about twice the carrier frequency with a phase shift that is proportional to the delay time T_p. This term is generally filtered out actively, or more usually in millimetre-wave radar systems because it is beyond the cut-off frequency of the mixer and subsequent receiver components. The second cosine-term describes a beat signal at a fixed frequency.

This can be determined by differentiating, with respect to time, the instantaneous phase term as shown,

$$f_b = \frac{1}{2\pi}\frac{d}{dt}\left[A_b T_p t + \left(\omega_c T_p + \frac{A_b}{2}T_p^2\right)\right], \tag{11.16}$$

$$f_b = \frac{A_b}{2\pi}T_p. \tag{11.17}$$

It can be seen that the signal frequency is directly proportional to the delay time T_p, and hence is directly proportional to the round-trip time to the target as postulated.

The spectrum shown below includes both the fixed and chirp terms for illustrative purposes, but in general only the low frequency component is output.

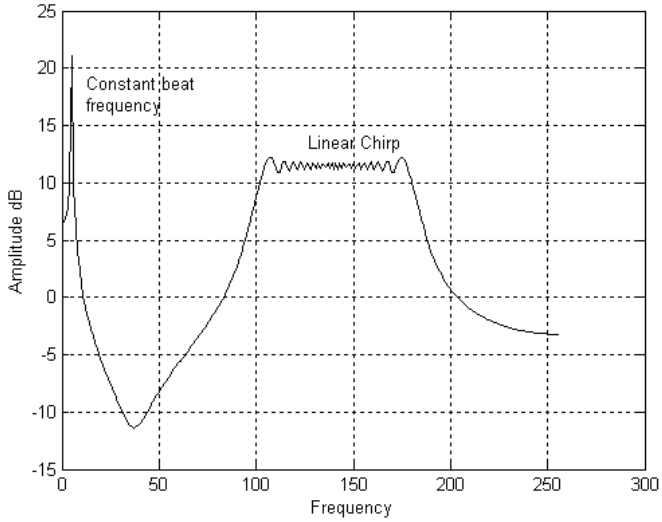


Figure 11.21: Frequency domain representation of the FMCW receiver output including both the high and low frequency components after mixing but before filtering

In these examples, and for most FMCW implementations, spectral analysis is performed using the standard FFT. However there are other techniques that can be used. These include autoregressive (AR), autoregressive moving average (ARMA), minimum-entropy methods and spectral-parameter estimation including the now famous MUSIC algorithm.

11.7.2. Matched Filtering

Linear chirp is the most Doppler-tolerant code that is commonly implemented in analog form today as it concentrates most of the volume of the ambiguity function into the main lobe and not into the Doppler or delay sidelobes.

In the previous section it was shown that the output of a matched filter is just the correlation of the received signal with a delayed version of the transmitted signal. In the FMCW case this function is implemented by taking the product of the received signal with the transmitted signal and filtering to obtain a constant frequency beat, as discussed. The spectrum is then determined using the Fourier transform or a similar spectral estimation process.

If the chirp duration is T_b seconds, then the spectrum of the beat signal will be resolvable to an accuracy of $2/T_b$ Hz (between minima) as determined by the

ambiguity function at zero Doppler, $\chi_o(t_d,0)$. This assumes that $T_b >> T_p$ so that the signal duration $\tau \approx T_b$, as shown in Figure 11.19. It is common practice to define the resolution bandwidth of a signal, δf_b, between its 3dB (half power) points, which in this case fall within the $1/T_b$ region centred on f_b.

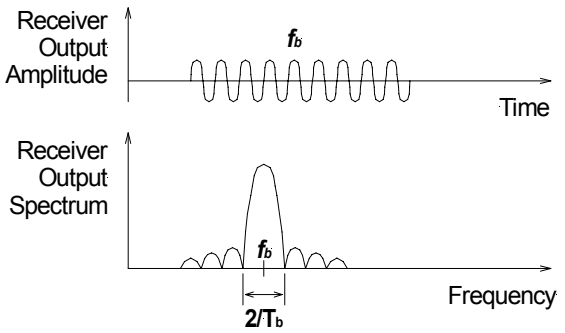


Figure 11.22: Spectrum of the truncated sinusoidal signal output by an FMCW radar

The rate of change of frequency (chirp slope) in the linear case is constant and equal to the total frequency excursion, Δf, divided by the chirp time, T_b. The beat frequency can then be calculated

$$f_b = \frac{A_b}{2\pi} T_p = \frac{\Delta f}{T_b} T_p . \tag{11.18}$$

From the basics of radar, the round-trip time T_p to the target and back can be written in terms of the range as

$$T_p = \frac{2R}{c}, \tag{11.19}$$

and substituting into (11.18) gives the classical FMCW formula that relates the beat frequency and the target range

$$f_b = \frac{\Delta f}{T_b} \frac{2R}{c} . \tag{11.20}$$

For a frequency resolution, δf, (11.20) can be used to show that the range and range resolution, δR, is

$$R = \frac{T_b c}{2\Delta f} f_b , \tag{11.21}$$

$$\delta R = \frac{T_b c}{2\Delta f} \delta f_b . \tag{11.22}$$

It was shown earlier that $\delta f_b = 1/T_b$, which when substituted into (11.22) results in a closed relationship between the total transmitted bandwidth and the range resolution

$$\delta R = \frac{c}{2\Delta f}. \tag{11.23}$$

This is intuitively quite satisfying at it represents the FMCW equivalent of the classical pulsed radar range-resolution equation where $\tau = 1/\Delta f$.

11.7.3. The Ambiguity Function

The Ambiguity Function for linear FM pulse compression, Stretch[4] and for FMCW, is reproduced below. It shows that there is a strong cross-coupling between the Doppler shift and the measured range. For a target with radial velocity, v_r, the magnitude of the coupling can be determined by replacing the echo signal in (12.12) with its Doppler shifted counterpart,

$$v_{fm}(t - T_p) = A_c \cos\left[\omega_c(t - T_p) + \frac{A_b}{2}(t - T_p)^2 - \frac{2v_r}{c}\omega_c(t - T_p)\right]. \tag{11.24}$$

Processing as before to determine the new beat frequency f_b, it is just the old beat frequency offset by the Doppler shift

$$f_b = \frac{2v_r}{c} f_c - \frac{A_b}{2\pi} T_p = f_d - \frac{A_b}{2\pi} T_p. \tag{11.25}$$

In this case, the ambiguity function can be expressed as

$$\left|\chi_o(t_d, f_d)\right|^2 = \left[\frac{\sin\left[\pi\left(f_d - \frac{A_b}{2\pi}t_d\right)\left(T_b - |t_d|\right)\right]}{\pi\left(f_d - \frac{A_b}{2\pi}t_d\right)\left(T_b - |t_d|\right)}\left(1 - \frac{|t_d|}{T_b}\right)\right]^2 \quad \text{for } -T_b \le t_d \le T_b \tag{11.26}$$

$$= 0 \qquad \text{for } |t_d| > T_b$$

For a typical FMCW radar with a 150MHz chirp over a 1ms interval ($A_b \approx 10^{12}$), the beat frequency in the absence of the Doppler shift is about 1MHz at a range of 1km. The Doppler shift at 94GHz is 625Hz per m/s which equates to just less than the theoretical range resolution of the waveform in this case. Higher velocities, from a moving vehicle, for example, would introduce significant errors in the measured range which would need to be accounted for.

[4] A wideband linear FM pulse is transmitted and the return echo is down-converted using a frequency modulated LO of identical or slightly different FM slope. If the slopes are identical the output frequency from a single target is constant. If the slopes are slightly different then a pulse with a reduced chirp bandwidth is produced.

A cut along the Doppler axis is similar to that of the single pulse because the pulse width is the same, only the modulation is different. A cut along the time delay axis changes considerably as it is now much narrower and corresponds to the compressed pulse width $\tau_c = 1/\Delta f$.

In the figure below an increasing Doppler shift results in a decreasing measure of the range because a rising-frequency chirp is used. For a decreasing-frequency chirp, the sense of the function is reversed to produce a mirror image of this ambiguity diagram. By combining the two slopes using a triangular modulation, it is possible to obtain an unbiased estimate of the target range and of the Doppler shift.

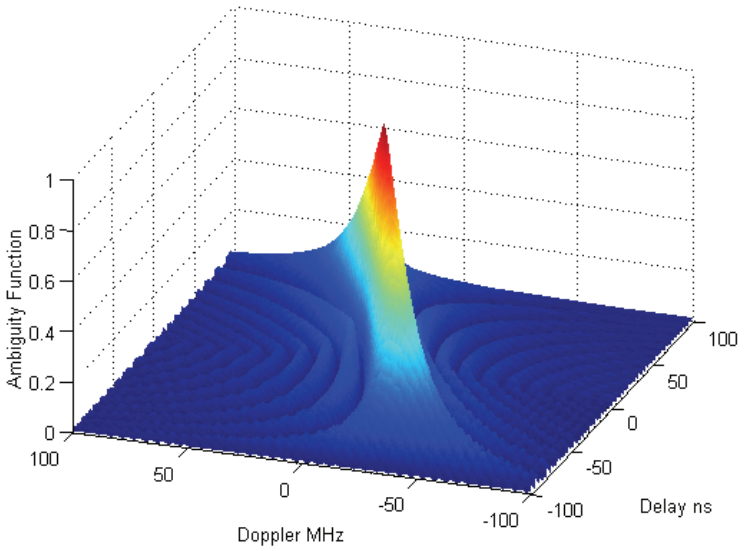


Figure 11.23: Linear FM up chirp ambiguity diagram for a 100ns duration signal showing the interaction between delay and Doppler

A moving target will therefore superimpose a Doppler frequency shift on the beat frequency as shown in the figure below.

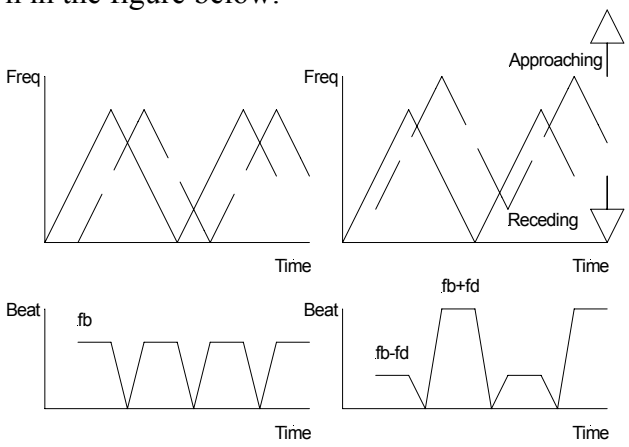


Figure 11.24: Effects of Doppler shift on beat frequency

One portion of the beat frequency will be increased and the other portion will be decreased. For a target approaching the radar, the received signal frequency is

increased (shifted up in the diagram) decreasing the up-sweep beat frequency and increasing the down-sweep beat frequency

$$f_{b(\text{up})} = f_b - f_d, \tag{11.27}$$

$$f_{b(\text{dn})} = f_b + f_d. \tag{11.28}$$

The beat frequency corresponding to range can be obtained by averaging the up and down sections $f_r = [f_{b(\text{up})} + f_{b(\text{dn})}]/2$.

The Doppler frequency (and hence target velocity) can be obtained by measuring one half of the difference frequency $f_d = [f_{b(\text{up})} - f_{b(\text{dn})}]/2$.

The roles are reversed if $f_d > f_b$.

11.7.4. Effect of a Non-Linear Chirp

As shown conceptually in the figure below, if the chirp is not linear, the standard matched filter assumptions for resolution are not satisfied and the range resolution will suffer.

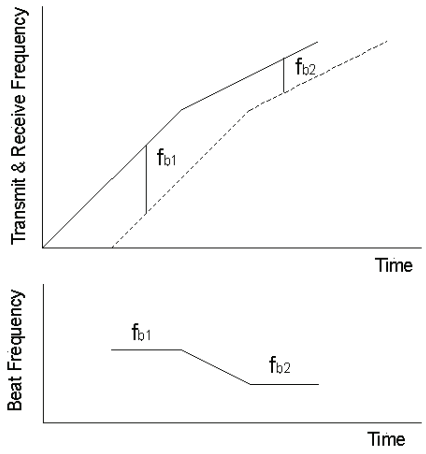


Figure 11.25: Effect of chirp nonlinearity on the beat frequency

The analysis presented thus far assumes that the chirp is completely linear with time. However, in most practical applications this is not true and it can be shown that if the non-linearity is quadratic in nature then the range resolution becomes proportional to the slope linearity and the range to the target

$$\delta R = R.Lin\,, \tag{11.29}$$

where the linearity, *Lin*, is defined as the change in chirp slope, $S = df/dt$, normalised by the minimum slope.

$$Lin = \frac{S_{\max} - S_{\min}}{S_{\min}}\,. \tag{11.30}$$

This sensitivity to slope linearity is one of the fundamental problems that limits the resolution of real FMCW radar systems.

11.7.5. Open Loop Linearisation

A common method to perform this function uses the programmed correction stored in an EPROM and then clocking this data through a digital to analog converter (DAC). Because of the varying characteristics of the VCO with temperature, either the temperature must be controlled, or a number of different curves must be implemented and referenced appropriately. The latter is easily achieved, as shown in the figure below where the upper bits of the lookup table address are driven by a digital representation of the temperature and the lower bits address a particular voltage entry in that table.

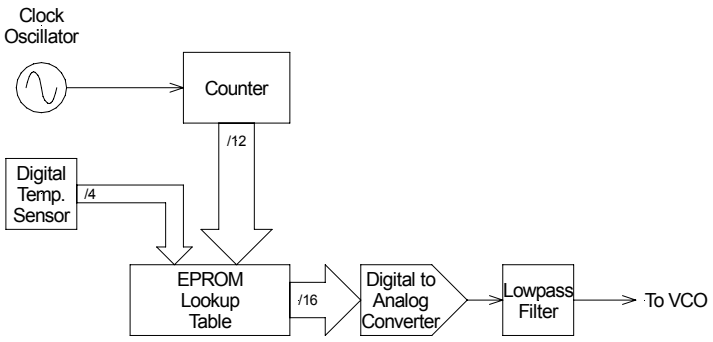


Figure 11.26: Schematic diagram of a linear chirp generator based on a lookup table

Glitches that are generated during some of the DAC transitions generate noise and clock harmonics on the RF signal which are difficult to remove by filtering, and it is only since a new generation of low glitch power DACs has became available that this technique has become feasible.

An all-analog configuration as shown below which can reduce the nonlinearity of a well-behaved VCO by a factor of 10. This alternative uses an analog multiplier chip to produce a quadratic voltage that is added to the linear ramp to perform the correction. A DC offset is often included in the circuit to set the start frequency.

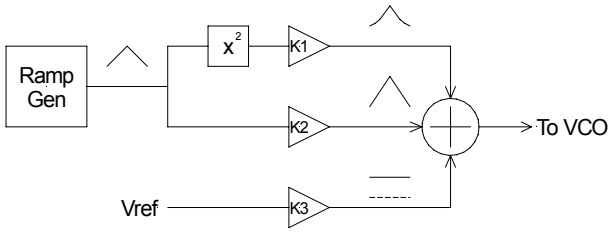


Figure 11.27: Quadratic frequency chirp correction circuit using an analog multiplier chip

11.7.6. Determining the Effectiveness of Linearisation Techniques

The obvious method to determine the effectiveness of a linearisation technique is to examine the beat-frequency spectrum for a point target, at a reasonable range (>500m). However, this is often not practical, and so an alternative more compact method is required.

In essence all a FMICW radar does is mix a portion of the transmitted signal with the received signal to produce a beat signal, the frequency of which is proportional to the

range. As the name implies, a delay-line discriminator performs the same function using an electrical delay-line rather than the genuine round-trip delay to a target and back.

The most basic delay-line is simply a length of coaxial or fibre-optic cable, but these are usually too bulky for practical applications. In general, surface acoustic wave (SAW) or bulk acoustic wave (BAW) devices are used to fulfil this function as shown in the figure below.

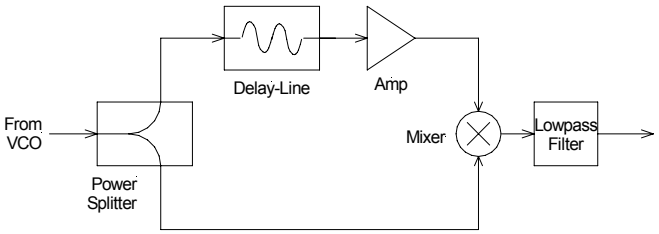
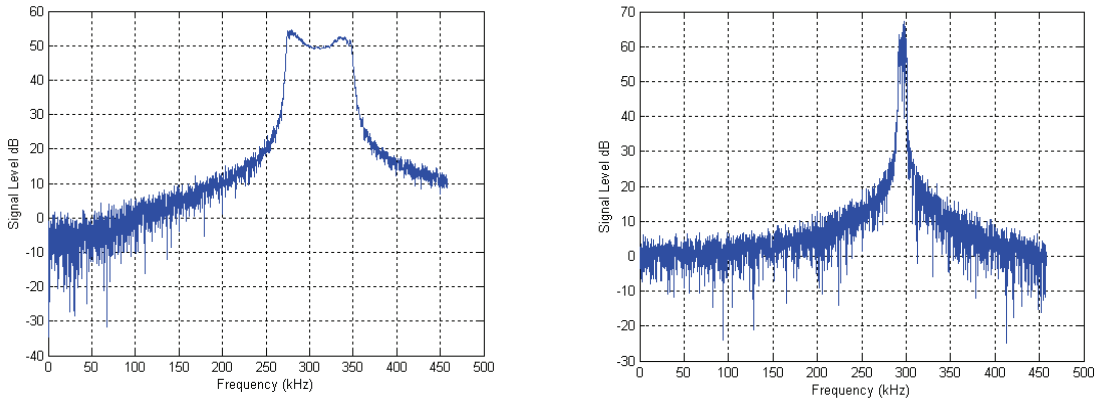


Figure 11.28: Schematic diagram of a delay-line discriminator

Disadvantages of SAW delay-lines are their high insertion loss (>35dB) limited bandwidth (<300MHz) and an operating frequency of less than 1GHz. For millimetre-wave radar applications, the VCO frequency must be down-converted to an appropriate IF (700MHz) to take advantage of commercially available components.

The spectrum of the output of the discriminator is then examined to determine the effectiveness of the linearization process. The centre frequency defines the chirp slope, and the 3dB bandwidth, the linearity. In the following figure, the discriminator outputs are shown for the Hughes VCO both completely unlinearised and after open loop linearization. Note that the width of the signal is reduced from 80kHz to 10kHz which implies an improvement in linearity from 0.26 to just over 0.03.



(a) (b)

Figure 11.29: Discriminator output spectra for (a) an unlinearised Hughes VCO and (b) after open-loop correction

11.7.7. Implementation of Closed-Loop Linearisation

The delay-line discriminator can be used as a feedback element to close the linearization loop using a classical phase-locked loop, if it is remembered that the loop must maintain a constant rate-of-change of frequency and not frequency as is more usual.

The delay-line discriminator is effectively a differentiator in the frequency domain and produces a constant output frequency if the frequency slope (rate-of-change of frequency) is constant. Thus to close the loop correctly, an integrator must be implemented in the feedback path to produce the loop structure shown in the figure below.

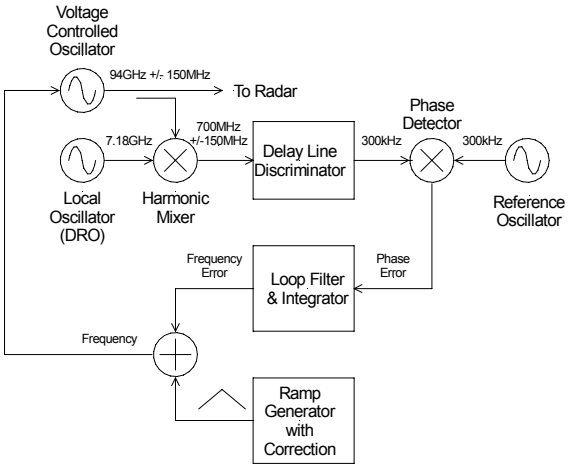
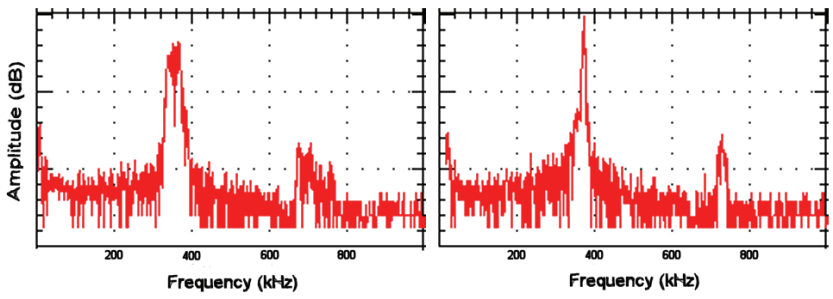


Figure 11.30: Schematic diagram showing the process of chirp linearisation based on a combination of open-loop correction and closed-loop delay-line discriminator output feedback

The implementation of a loop filter that exhibits the appropriate locking bandwidth, low phase-noise and good suppression of spurious signals requires careful design and layout. Even so, it is nearly impossible to eliminate the spurious signals from the receiver spectrum completely.

The following figure shows the discriminator output for open-loop and closed-loop linearisation



(a) **(b)**

Figure 11.31: Measured delay line discriminator output spectra (a) Unlinearised chirp and (b) Linearised chirp

The combined range resolution due to the swept bandwidth and the non linearity's can be determined as follows

$$\delta R_{tot} = \sqrt{\delta R^2 + \delta R_{lin}^2} \,. \tag{11.31}$$

11.7.8. Extraction of Range Information

Multiple targets result in more than one beat frequency being present in the received signal, so a simple counter can no longer be used to determine the range

Range gating must be performed using some spectral analysis technique

- Bank of band-pass filters
- Swept band-pass filter (Spectrum Analyser)
- Digitisation and FFT processing

FFT Processing

Fourier analysis shows that the power spectrum of a truncated sine wave will have sidelobes only 13.2dB lower than the main lobe. This is not adequate for high resolution systems as it results in "leakage" of the return from one target to contaminating and even overwhelming the returns from adjacent smaller targets.

Implementation of the FFT is therefore almost always preceded by a windowing function to reduce the sidelobe level. This is discussed in detail later in this chapter,

If the signal is observed for a time T_d then the width of the FFT frequency bin $W = 1/T_d$ and main lobe width, as shown earlier is twice that. The 3dB bandwidth of the filter produced by the FFT process is 0.89 bins for no windowing (rectangle), increasing to 1.3 bins for a Hamming window.

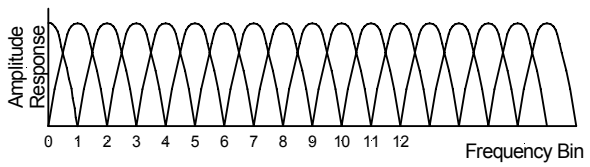


Figure 11.32: Filter bank implemented using the FFT

11.7.9. Problems with FMCW

The primary problems with FMCW all relate to transmitting and receiving simultaneously as the transmitted power can be more than 100dB higher than the received echo, so if even a small fraction of the transmitted power leaks into the receiver it can saturate or even damage the sensitive circuitry.

The performance of even well designed systems used to be degraded by 10-20dB compared to that which is achievable with pulsed systems. This limitation can be minimised by ensuring that there is good isolation between the receive and transmit antennas by separating them and using low antenna sidelobe levels.

Modern signal processing techniques and hardware can also be used to cancel the leakage power in real time, and good performance can be obtained.

11.8.Stretch

In Stretch a linear FM pulse is transmitted and then the return echo is demodulated by down-converting using a frequency modulated LO signal of identical or slightly different FM slope. If the identical slope is used then the echo spectrum corresponds to the range profile. This is a form of pulse compression intermediate between standard pulse compression and FMICW.

If the slope of the LO is different to that of the transmitted chirp, then the output of the Stretch processor comprises signals with a reduced chirp. These can then be processed using a standard SAW pulse compression system to produce target echoes as described in the previous section.

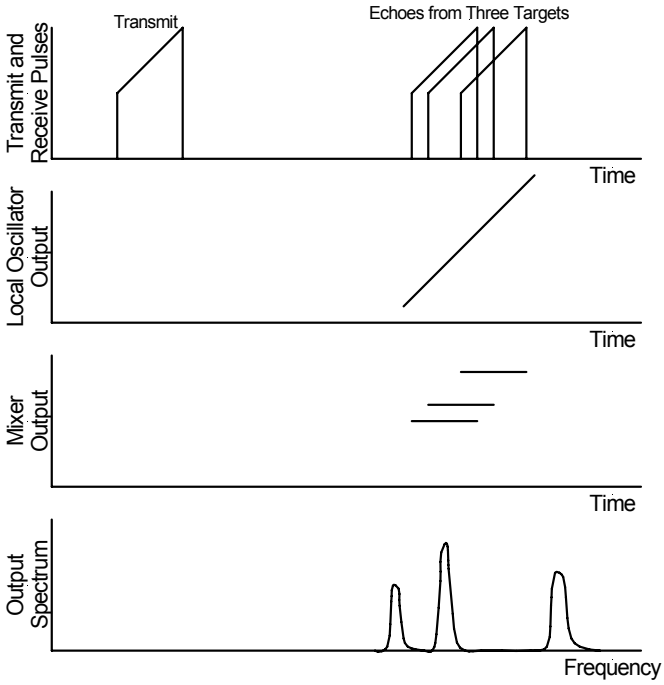


Figure 11.33: Stretch processing of received overlapping chirp echoes

11.9.Interrupted FMCW

Known as IFMCW or FMICW, this involves interrupting the FMCW signal to eliminate the requirement for good isolation between the transmitter and the receiver. It is generally implemented with a transmission time matched to the round trip propagation time. This is followed by a quiet reception time equal to the transmission time.

A duty factor of 0.5 reduces the average transmitted power by 3dB but the improved performance due to reduced system noise improves the SNR by more than the 3dB lost.

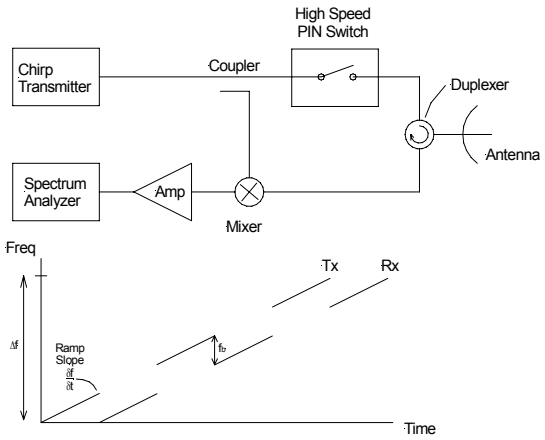


Figure 11.34: FMICW principle of operation

11.9.1. Disadvantages

The major problems are the limited minimum range due to the finite switching time of the transmitter modulator and the need to know the target range to optimise the transmit time.

For imaging applications where a whole range of frequencies are received, maintaining a fixed 50% duty cycle is sub-optimum except at one range

FFT processing of the interrupted signal results in large numbers of spurious components that can interfere with the identification of the target return as shown in the following figure.

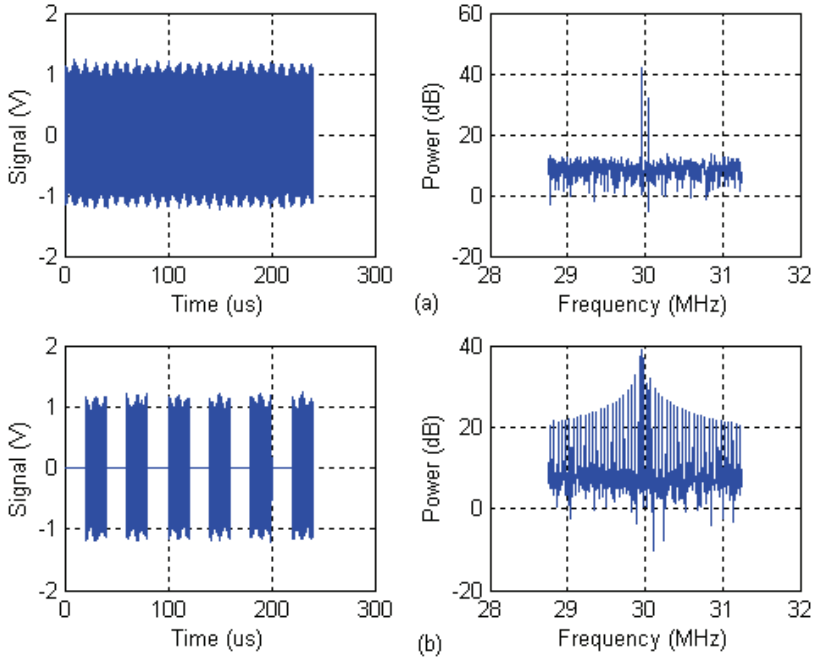


Figure 11.35: Comparison between the received signals and spectra for two closely spaced targets of different amplitudes for (a) an FMCW radar and (b) an FMICW radar with a deterministic interrupt sequence

11.9.2. Optimising for a Long Range Imaging Application

The Tx time is optimised for the longest range of interest (where the SNR will be lowest)

The shorter ranges will suffer from the following problems:

- Reduced illumination time -> lower SNR
- Reduced chirp bandwidth -> poorer range resolution
- Sub-optimal windowing -> higher range sidelobes

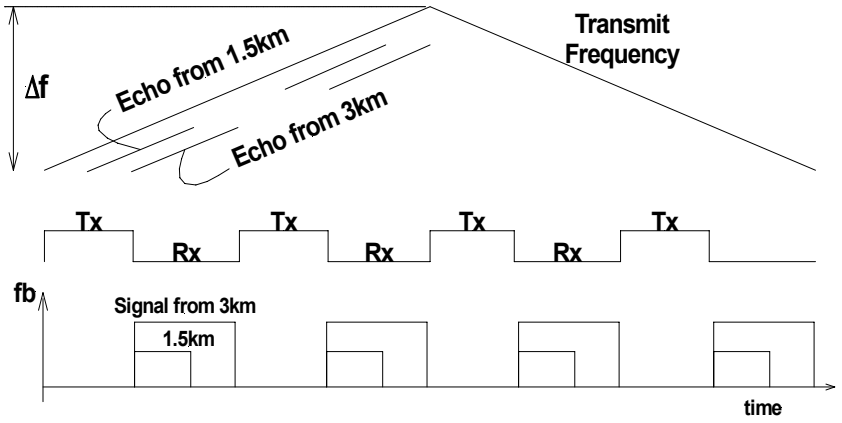


Figure 11.36: FMICW waveform optimised for 3km

The degradation in range resolution at short range is compensated for by the improved cross-range resolution (constant beamwidth) so the actual resolution (pixel area) remains constant.

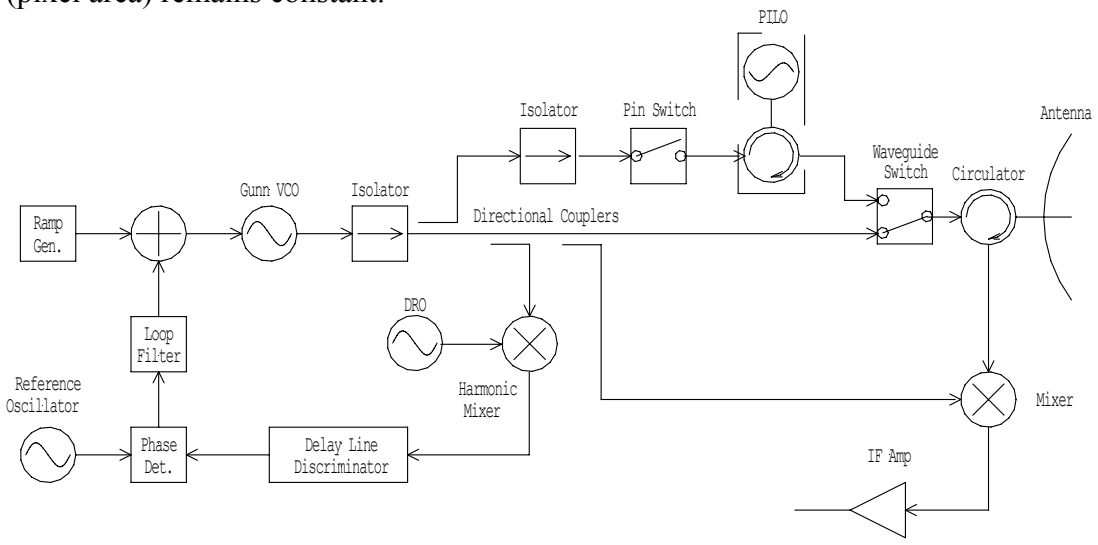


Figure 11.37: FMICW radar front-end block diagram

11.10. Sidelobes and Weighting for Linear FM Systems

The spectrum of a truncated sine wave output by an FMCW radar for a single target, has the characteristic $|sin(x)/x|$ shape as predicted by Fourier theory.

The range sidelobes in this case are only 13.2dB lower than the main lobe which is not satisfactory as it can result in the occlusion of small nearby targets as well as introducing clutter from the adjacent lobes into the main lobe. To counter this unacceptable characteristic of the matched filter, the time domain signal is mismatched on purpose. This mismatch generally takes the form of amplitude weighting of the received signal.

One method to do this is to increase the FM slope of the chirp pulse near the ends of the transmitted pulse to weight the energy spectrum which will result in the desired low sidelobe levels after the application of the matched filter. This is easy to achieve in surface acoustic wave (SAW) based systems. A more conventional method that is often used in digital systems is to apply the function to the signal amplitude prior to processing to achieve the same ends as shown in the figure below.

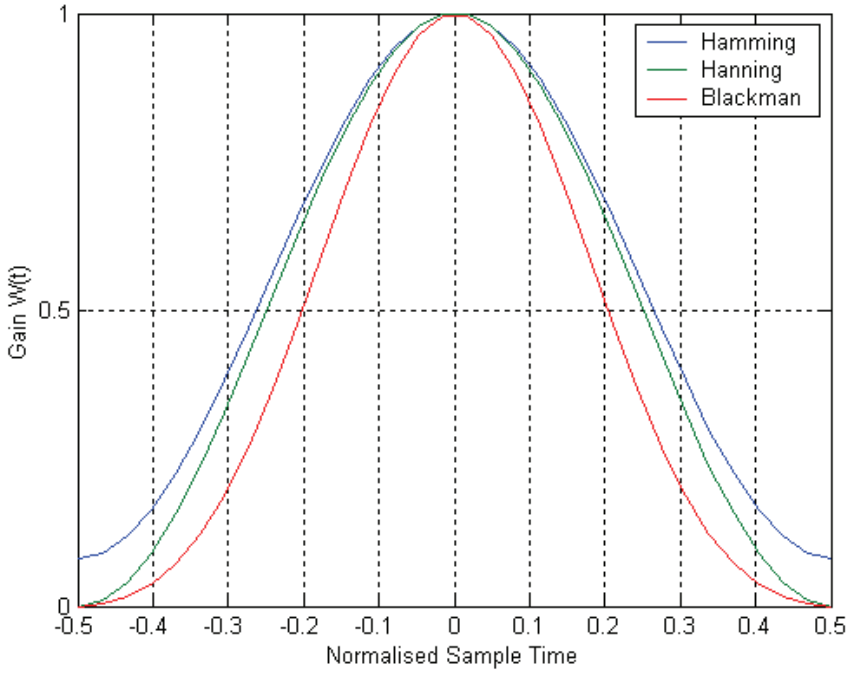


Figure 11.38: Weighting function gains

The reduction in sidelobe levels does come at a price though: the main lobe amplitude is also marginally reduced in amplitude, and it is also widened quite substantially as summarised in the following table.

Table 11.4: Properties of some weighting functions

Window	Rectangle	Hamming	Hanning	Blackman
Worst Sidelobe (dB)	-13.2	-42.8	-31.4	-58
3dB Beamwidth (bins)	0.88	1.32	1.48	1.68
Scalloping Loss (dB)	3.92	1.78	1.36	1.1
SNR Loss (dB)	0	1.34	1.76	2.37
Main Lobe Width (bins)	2	4	4	6
a0	1	0.54	0.50	0.42
a1		0.46	0.50	0.50
a2				0.08
$W(n)=a0-a1\cos[2\pi(n-1)/(N-1)]+a2\cos[4\pi(n-1)/(N-1)]$				

The rectangular, or uniform, weighting function provides a matched filter operation with no loss in SNR, while the weighting in the other cases introduces a tailored mismatch in the receiver amplitude characteristics with an associated loss in SNR which can be quite substantial.

In addition to providing the best SNR, uniform weighting also provides the best range resolution (narrowest beamwidth), but this characteristic comes with unacceptably high sidelobe levels. The other weighting functions offer poorer resolution but improved sidelobe levels with falloff characteristics that can accommodate almost any requirement as seen below.

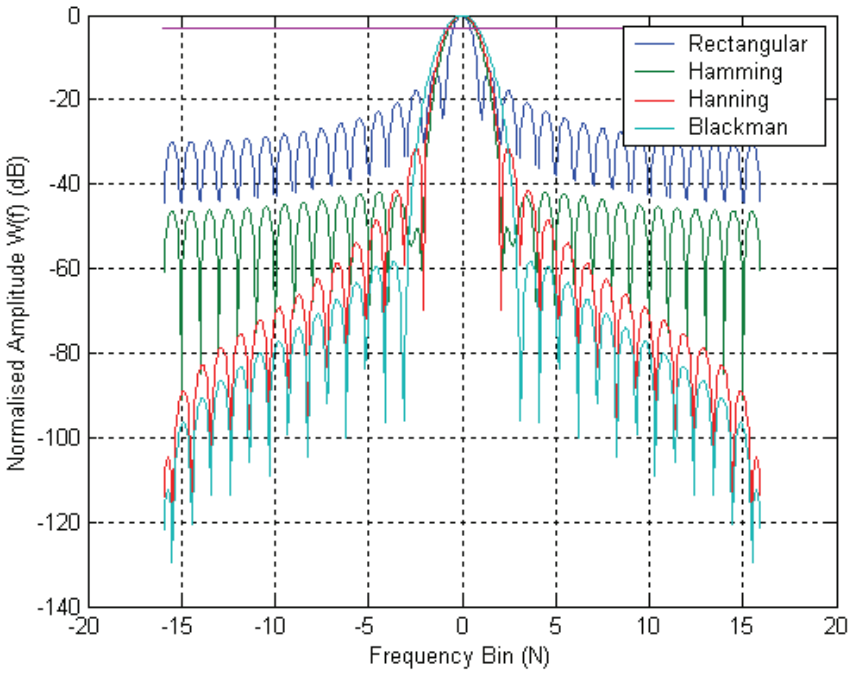


Figure 11.39: Normalised weighting function amplitude spectra for different window functions

Of particular interest are the Hamming and Hanning weighting functions which offer similar loss in SNR and resolutions, but with completely different sidelobe characteristics. As can be seen in Figure12.44, the former has the form of a cosine-squared-plus-pedestal, while the latter is just a standard cosine squared function. In the Hamming case, the close-in sidelobe is suppressed to produce a maximum level of -42.8dB but that energy is spread into the remaining sidelobes resulting in a falloff of only 6dB/octave, while in the Hanning case, the first sidelobe is higher, -31.4dB, but with a falloff of 18dB/octave.

For most FMCW applications, the Hamming window is used as it provides a good balance between sidelobe levels (-42.8dB), beamwidth (1.32 bins) and loss in SNR compared to a matched filter (1.34dB). For imaging applications, where a large dynamic range of target reflectivities is expected, then the Hanning window with its superior far-out sidelobe performance is the function of choice.

11.11. FMCW Radar Systems

Many short range sensors operate using FMCW principles. One common example is the Krohne level radar which operates at X-Band (10GHz) and used FMCW techniques.



- Frequency 8.5 to 9.9GHz
- Range 0.5m to 100m (longer if required)
- Swept bandwidth 1GHz
- Closed loop linearisation
- Linearity correction 98%
- Accuracy <+/-0.5% for still target
- Fourier processing
- Rate of change of level <10m/min

Figure 11.40: Krohne radar

We have built a number of systems at W-Band including simple range-only radars used for measuring the depths of orepasses as described in earlier in these notes.

One of the more complex units is the dual polar 94GHz unit built to measure the characteristics of vehicles in two orthogonal linear polarisations shown in the figure below.

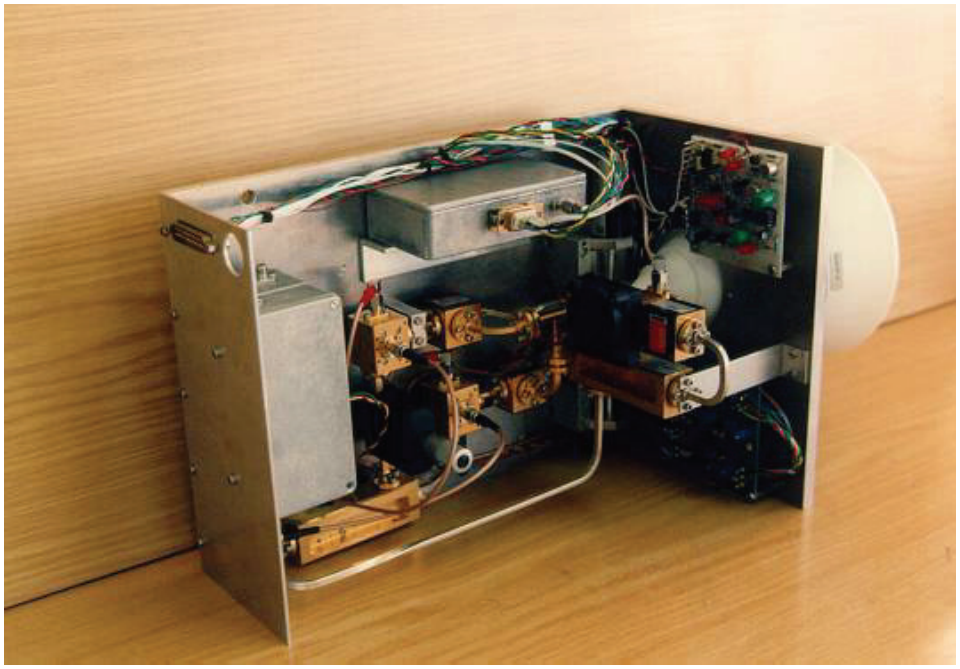


Figure 11.41: Dual polar W-band FMCW radar

We have also developed a modular high range-resolution FMCW system operating at 77GHz for the following applications

- Imaging radar for unmanned aerial vehicle (ANSER)
- Imaging radar for stope fill monitoring
- Imaging radar for shovel/dragline visualisation

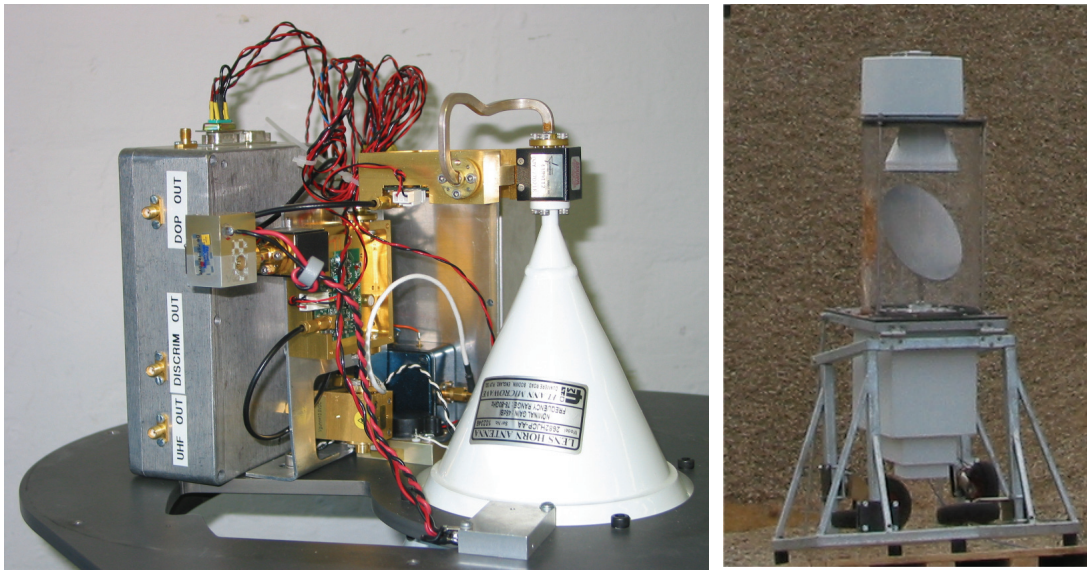


Figure 11.42: High resolution 77GHz radar for imaging

11.12. Application: Brimstone Antitank Missile

The Brimstone Missile is one of the guided missiles developed for the Longbow Apache AH-64D attack helicopter

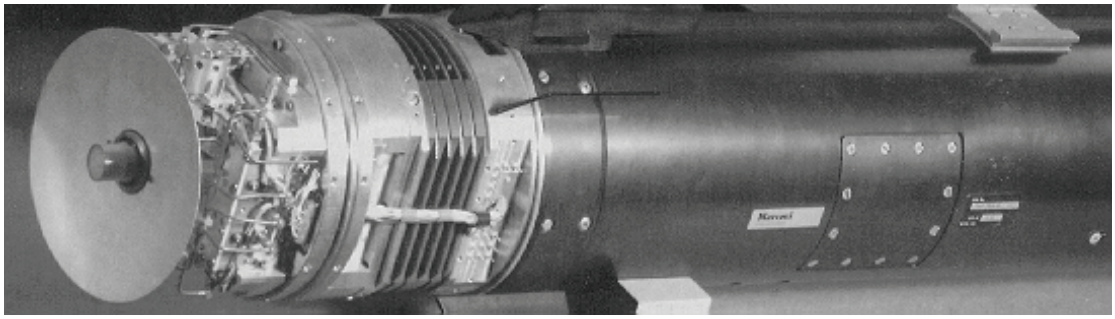


Figure 11.43: The Brimstone missile with radome removed showing the FMCW seeker

11.12.1. System Specifications

- Length: 1.8m Diameter: 178mm
- Mass: 50kg
- Operation: 24hr, day/night, all weather
- Mode: Totally autonomous, fire-and-forget, lock-on after launch (LOAL)
- Resistant to camouflage, smoke, flares, chaff, decoys, jamming
- Operational Range: 8km
- Designation: Accepts any or no target information
- Motor: Boost/coast, burns for 2.75s with a thrust of 7.5kN
- Guidance: Digital autopilot, 2 gyros (25°/hr drift), 3 accelerometers

11.12.2. Seeker Specifications (known)

- 94GHz active radar
- Low power, narrow beam
- Dual polar, dual look
- Fast 96002 processor
- Detection/ classification software

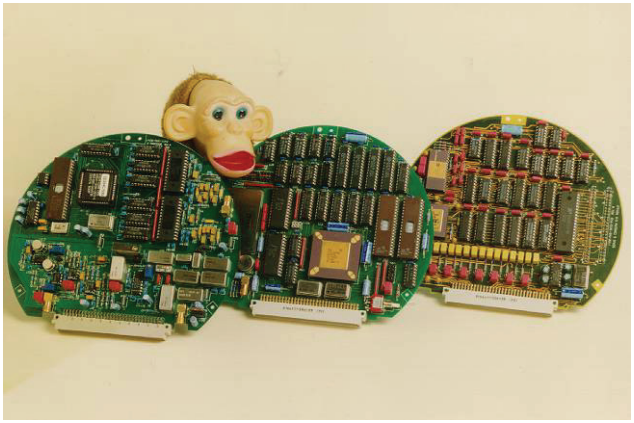


Figure 11.44: Processor for an FMCW seeker and a monkey

11.12.3. Operational procedure (LOAL)



Figure 11.45: Missile engagement options

The operational procedure for lock-on after launch is as follows:

- Rough target designations including, range bearings and rates downloaded to missile
- Missile fired in general direction of target
- Updates designation from initial positions and rates
- Flies up to 7km toward target using INS guidance only
- In the last 1km it activates the radar seeker and searches for target
- Search footprint scans search box in 200ms

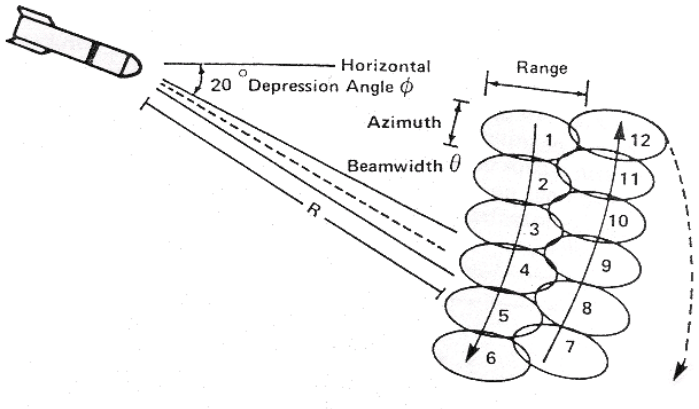


Figure 11.46: Push broom search relies on forward motion of missile and antenna scan in azimuth to cover a swathe of ground

- Acquisition algorithms map all targets in box (exclude trucks)
- Track-while-scan enables optimum decision on target priority
- Algorithm selects ADU or MBT
- Moving armour given the highest priority

11.12.4. System Performance (speculated)[5]

Target Detection and Identification

Target identification is based on a combination of the high range-resolution and polarisation characteristics of the radar echo. The system transmits horizontal polarisation (H) and receive vertical (V) and horizontal (H) returns and the range gate size is matched to the radar bandwidth for high resolution ≈0.5m. This puts between 6 and 10 range cells on a typical MBT (3m × 5m).

Doppler processing is used to distinguish moving targets.

Radar Front End

To make the radar low probability of intercept (LPI), the transmit power will be low and spread spectrum. This almost certainly implies FMCW operation.

FMCW operation through a single antenna generally limits the transmit power to less than 50mW. However with good matching and active leakage compensation, transmit powers can be as high as 1W. We believe that the Brimstone transmit power P_{tx} ≈ (100mW) 20dBm as it includes an injection locked amplifier stage

Transmitter swept bandwidth Δf = 300MHz to meet the 0.5m range resolution requirement,

$$\delta R_{chirp} = \frac{c}{2\Delta f} = \frac{3\times10^8}{2\times300\times10^6} = 0.5m$$

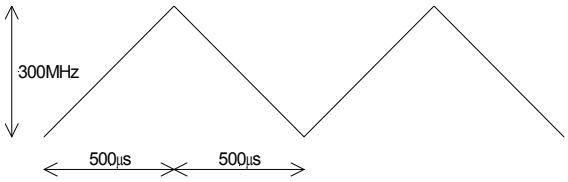
.

To allow for Doppler processing a triangular waveform will be used as shown below

For an operational range of 1km with a 0.5m bin size, 2000gates are required. It is speculated that a 4096pt FFT will produce 2048bins for both the co and cross polar receive channels.

Because the time available to perform a search is limited, the data rate will be as high as possible, however, there is a limit to the speed that the loop linearisation and the ADC can operate. We will assume a total sweep time of 1ms (500μs for each the up and down sweeps).

[5] Because of the limited digitisation and processing power available when the Brimstone was developed, it uses fewer gates for search and detection, before performing higher resolution processing for the target identification phase.



The beat frequency for an FMCW radar is given by the following equation

$$f_b = \frac{\delta f}{\delta t}.T_r = \frac{\delta f}{\delta t}\frac{2R}{c} = \frac{300\times10^6}{500\times10^{-6}}\frac{2\times1000}{3\times10^8} = 4MHz$$

Using the Nyquist criterion, the minimum sample rate required to digitise a signal with a 4MHz bandwidth is 8MHz. Because of non brick-wall anti aliasing filter characteristics, the sample rate is generally 2.5× making the sample rate 10MHz

To ensure sufficient dynamic range, an ADC with at least 12bits of resolution is required.

A total of 5000 samples can be taken over each the up and the down sweep, this is just about perfect for the 4096 point FFT because the sweep linearity is generally not good at the start and the end.

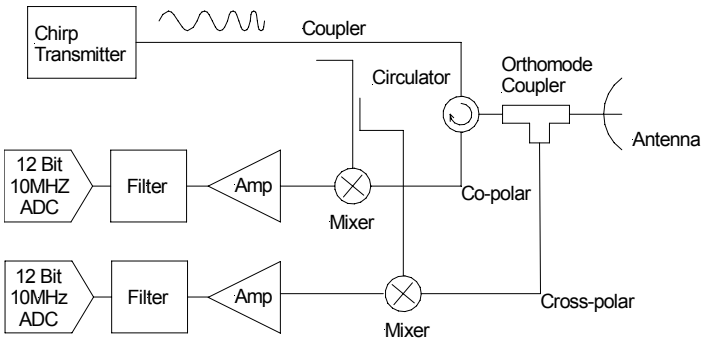


Figure 11.47: Brimstone seeker schematic diagram

Antenna and Scanner

For a missile diameter of 178mm, the antenna cannot be much more than 160mm in across.

For λ=3.2mm at 94GHz, the 3dB beamwidth will be

$$\theta_{3dB} = \frac{70\lambda}{D} = \frac{70\times3.2}{160} = 1.4\deg$$

The antenna uses an interesting Cassegrain configuration with a scanned parabolic mirror as shown in the figure.

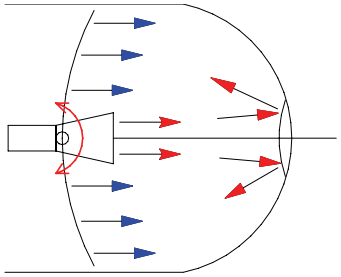


Figure 11.48: Schematic of the derivative Cassegrain antenna used by Brimstone

The gain of the pencil beam antenna will be approximately

$$G = \frac{4\pi\eta A}{\lambda^2} = \frac{4\pi \times 0.6 \times \pi \times 0.08^2}{0.00319^2} = 14897 \quad (41.7\text{dB})$$

The critical aspect is the sub-reflector beam shaping that allows a limited scan using the parabolic prime reflector without generating large sidelobes

At a range of 1km, the width of the footprint will be 24.5m and the length of the footprint will be a function of the operational height at an operational range of 1km.

Table 11.5: Relationship between radar height and beam footprint length

height (m)	angle1 (deg)	angle2 (deg)	x2 (m)	footprint (m)
10	0.57	1.97	290.29	709.71
20	1.15	2.55	449.83	550.17
30	1.72	3.12	550.67	449.33
40	2.29	3.69	620.13	379.87
50	2.86	4.26	670.87	329.13
100	5.71	7.11	801.64	198.36

To limit the amount of potential shadowing of the target area due to trees and undulating terrain, while maintaining a reasonable size footprint on the ground, an operational height of 50m would be reasonable. This results in a footprint length of 330m

It can be assumed that a single mechanical scan takes place in the 200ms search time

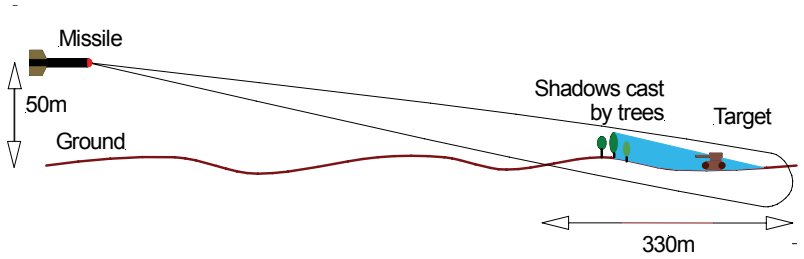


Figure 11.49: Shadow effects due to low grazing angle

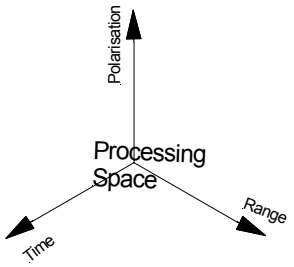
Because the missile is coasting, it will have limited lateral acceleration capability, and so it is pointless searching beyond the boundaries that the missile can reach.

It is reasonable to assume that a square search area of 330×330m will be covered. At a range of 1000m, this equates to an angular scan of about 18° if the antenna beamwidth is considered. To scan 18° in 200ms requires an angular rate of 90°/s

Signal Processing

The time-on-target for a beamwidth of 1.4° and an angular rate of 90°/s is 15.5ms. For a total sweep time of 1ms, a total of nearly 16 hits per scan occurs.

This allows for 16 pulse integration to improve the signal to noise ratio if it is required, it also gives the processor more information to identify the target type



Each target can be identified using the following information

- 5-10 gates that span it in range
- 16 time slices
- 2 orthogonal polarisations

This is sufficient information to discriminate between a truck and a main battle tank (MBT)

Signal to Clutter Ratio: Clutter Levels

Single look signal to clutter ratio (SCR) is determined from the target RCS, the clutter reflectivity σ^o and the area of a range gate.

The following graphs show measured clutter reflectivity data at 94GHz for grass and crops.

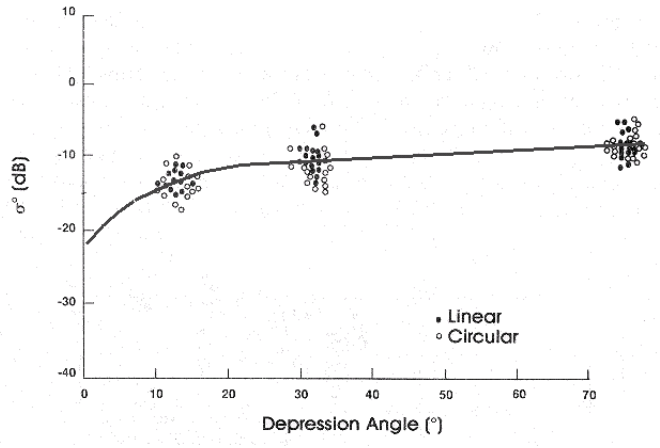


Figure 11.50: Clutter reflectivity for grass & crops at 94GHz

At a grazing (depression) angle of between 3° and 4° the mean reflectivity of grass will be about –20dBm^2/m^2.(reduces to dB).

The clutter cross section is the product of the clutter reflectivity σ^o and the area of the gate footprint $\tau.R.\theta_{3dB}$ on the ground for flat terrain (the beamwidth must be in radians) as described in Chapter 9.

$$\sigma_{clut} = \sigma^o \tau R \theta_{3dB} = -20 + 10\log_{10}(0.5 \times 1000 \times 1.4 \times \frac{\pi}{180}) = -9\text{dBm}^2$$

Because tanks commanders are aware that they are vulnerable when out in the open, they tend to make use of the available local cover, and will position themselves on the borders of lines of trees.

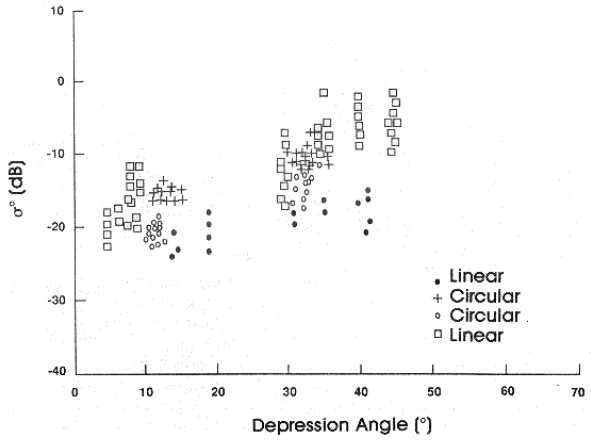


Figure 11.51: Clutter reflectivity for a deciduous tree canopy at 94GHz

The reflectivity of lines of trees observed broadside is much higher than that of the canopy, as shown in the following image which shows rows of pine trees between orchards, and a double line of eucalyptus straddling a railway line.

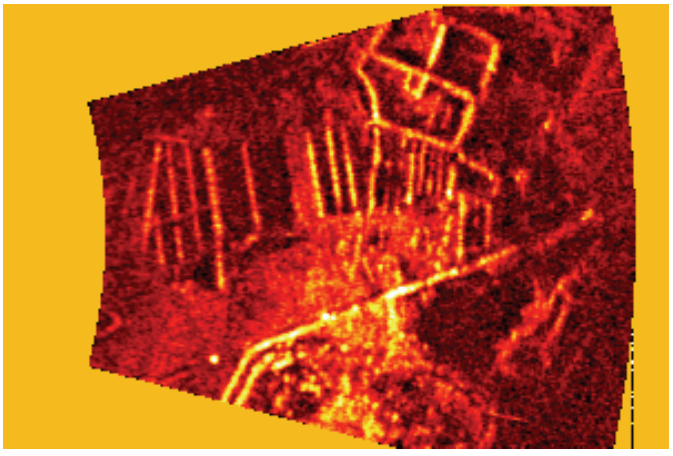


Figure 11.52: 94GHz radar image of trees and scrub gives an indication of the difficulties inherent in detecting small targets in ground clutter

Measurements made by us indicate that the mean reflectivity of deciduous trees is typically –10dBm^2/m^2.

The clutter RCS in this case is product of the area of trees illuminated by the radar and the reflectivity. Because of the narrow gate, there will be areas where the tree reflectivity is very strong and areas where it is very low.

There will also be areas where the tank is sticking out from under the tree, in which case the clutter level is determined by the ground clutter only.

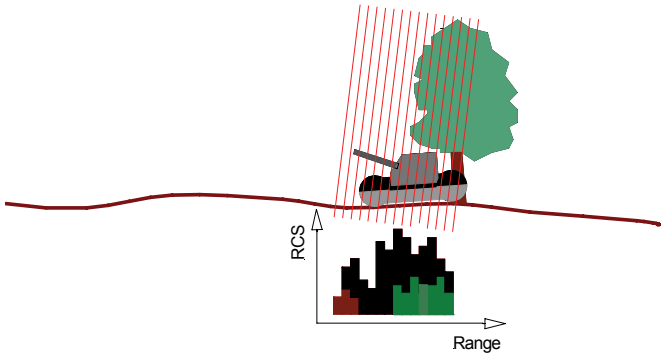


Figure 11.53: RCS profile of tank under a tree

If a 4m hedge of trees the width of the range gate is illuminated, then the RCS will be as calculated

$$\sigma_{clut} = \sigma^o hR\theta_{3dB} = -10 + 10\log_{10}(4 \times 1000 \times 1.4 \times \frac{\pi}{180}) = +10\text{dBm}^2$$

In general, however, a much smaller section of the tree will be illuminated, within a single gate. For a tree 4m tall and 3m wide, roughly elliptical in shape, a maximum area of $8m^2$ will be illuminated

$$\sigma_{clut} = \sigma^o A = -10 + 10\log_{10}(8) = -1dBm^2$$

Target Levels

The RCS of a tank depends on the observation angle as shown in the figure below



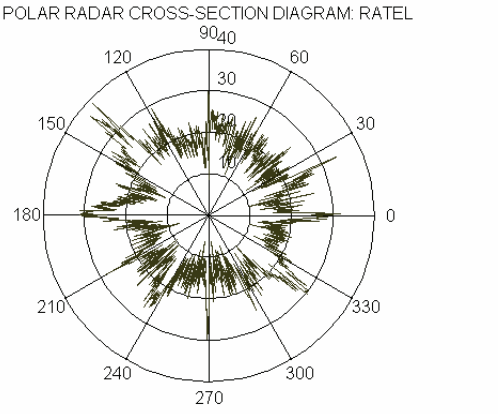


Figure 11.54: Radar cross section of a Ratel (Armoured Personnel Carrier)

The maximum RCS can reach 40dBm^2 and the minimum seldom falls below 10dBm^2. Hence, to ensure that the vehicle is always detected irrespective of the angle, then the 10dBm^2 threshold must be selected.

Signal to Clutter Ratio

In open ground the SCR is then

$$SCR = \sigma_{\tan} - \sigma_{clut} = 10 - (-10) = 20\text{dB}$$

For the tank under the tree, the worst case will be

$$SCR = \sigma_{\tan} - \sigma_{clut} = 10 - 10 = 0\text{dB}$$

Typical SCR will be more reasonable

$$SCR = \sigma_{\tan} - \sigma_{clut} = 10 - (-1) = 11\text{dB}$$

Without resorting to the statistics of the variation in tank RCS and that of trees, it can be seen that if the range bin is sufficiently narrow, parts of the tank will be visible if it is parked on the border of a row of trees.

When the radar is looking for a moving target, the clutter signals (because they are static) are suppressed.

Signal to Noise Ratio

The signal to noise ratio is determined using the characteristics of the radar and the target as they are related in the radar range equation.

The total noise at the output of the receiver N can be considered to be equal to the noise power output from an ideal receiver multiplied by a factor called the Noise Figure, NF. NF_{dB}.≈15dB for an FMCW radar.

In this case β is the bandwidth of a single bin output by the FFT and widened by the window function 1.3×5MHz/2048 ≈ 3kHz

$$N_{dB} = 10\log_{10} P_N NF = 10\log_{10} kT_{sys}\beta + NF_{dB} = -154dBW\,.$$

Because the transmitter power is in mW, this value is generally converted from dBW to dBm by adding 30dB

$$N_{dB} = -154 + 30 = -124dBm\,.$$

Writing the range equation for a monostatic radar system in dB

$$10\log_{10}(P_r) = 10\log_{10}(P_t) + 10\log_{10}\left(\frac{\lambda^2}{(4\pi)^3}\right) + 20\log_{10}(G) + 10\log_{10}(\sigma)$$
$$-10\log_{10}(L) - 40\log_{10}(R) - 2\alpha R_{km}$$

This is best tackled in MATLAB as the attenuation α is a function of the weather conditions

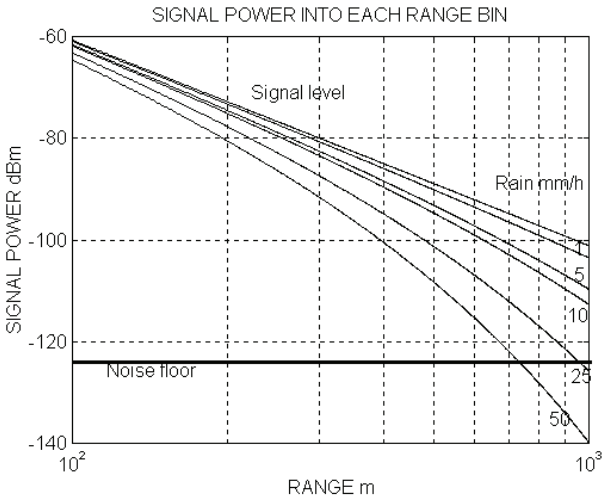


Figure 11.55: Brimstone performance in adverse weather

The signal to noise ratio is sufficient for detection up to a rain rate of about 10mm/hr.

Target Identification: Doppler Processing

The bandwidth of each bin output by the FFT is about 3kHz. This is equivalent to a Doppler velocity of

$$v_r = \frac{f_d \lambda}{2} = \frac{3 \times 10^3 \times 0.00319}{2} = 4.8\text{m/s}\,.$$

Because a Doppler shift causes an upward shift for half the sweep and a downward shift for the other, the range profiles generated by the up and down sweeps will diverge. For a target with a radial velocity of 4.8m/s this will be 2 bins, and will increase to 6 bins at a speed of 50km/h which is reasonable for a tank on the move.

A simple form of moving target discrimination is obtained by taking the difference between the up-sweep and the down-sweep range profiles. Static targets will cancel if the correct shift to compensate for the missile velocity is applied, but moving targets will appear as two large peaks as shown in the figure.

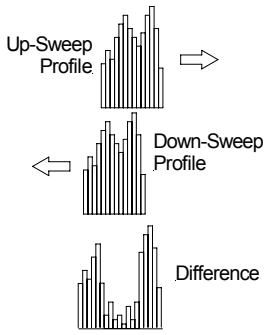


Figure 11.56: Moving target detection

Target Identification: Other Techniques

Different target types are identified by the differences in their co and cross-polar signatures.

Targets with lots of corners and attachments tend to reflect signals after more than one bounce, and that rotates the polarisation.

Because there are lots of scatterers each rotating the polarisation by a different amount, the overall return will have a random polarisation that is uniformly spread. The signal is said to be depolarised.

Smooth targets reflect with a single bounce, so the polarisation is not rotated.

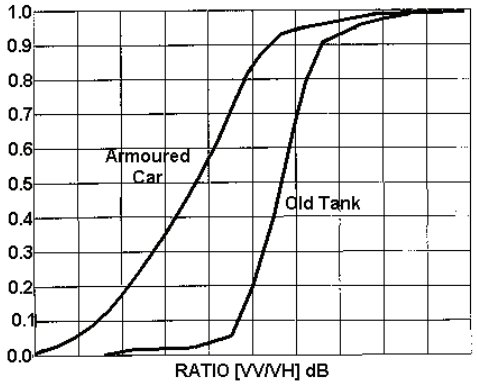


Figure 11.57: Polarisation ratio used to identify vehicles

The Results



Figure 11.58: Anti tank missile scoring a direct hit

11.13. Application: Ground Penetrating Radar

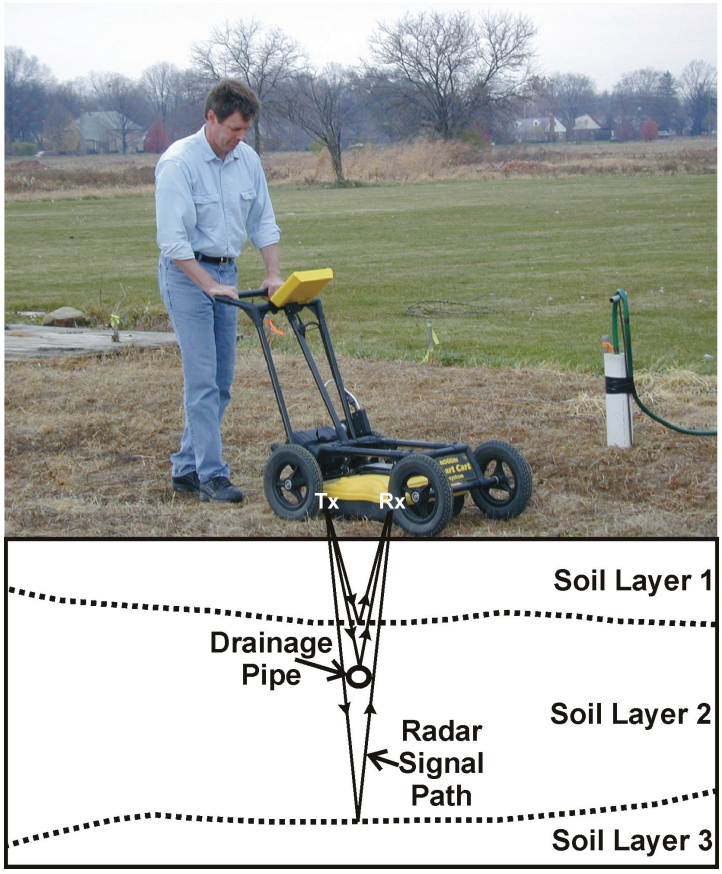


Figure 11.59: Ground penetrating radar deployment

11.13.1. Overview

GPR operates by transmitting a wide band low frequency electromagnetic signal into the earth.

A typical GPS signal may span the frequency range from 100MHz to 1GHz or higher. This can be generated using stepped frequency methods generated by Direct Digital Synthesis (DDS), or using a fast impulse or a fast rising/falling edge.

The main problem with GPR is to couple this wide-band energy into an antenna because most antennas are resonant, and so have bandwidths of <10%.

One method to broaden the bandwidth of an antenna is to resistively load it. This also has the effect of reducing its efficiency. GPR antennas often have efficiencies of <1% (the rest of the power is dissipated as heat). And so to compensate for this low efficiency, high voltage pulses (≈1kV) are often generated.

A low frequency (<2GHz) is selected as the absorption of EM radiation by rock is proportional to frequency.

The reflection coefficient ρ as the EM wave propagates, at normal incidence, from one non-magnetic material to another within the solid is given by the following

$$\rho = \left| \frac{Z_2 - Z_1}{Z_2 + Z_1} \right|$$

$$\rho = \left| \frac{\sqrt{\frac{\mu_o}{\varepsilon_o}} \left(\frac{1}{\sqrt{\varepsilon_{r2}}} - \frac{1}{\sqrt{\varepsilon_{r1}}} \right)}{\sqrt{\frac{\mu_o}{\varepsilon_o}} \left(\frac{1}{\sqrt{\varepsilon_{r2}}} + \frac{1}{\sqrt{\varepsilon_{r1}}} \right)} \right|$$

$$\rho = \left| \frac{\frac{1}{\sqrt{\varepsilon_{r2}}} - \frac{1}{\sqrt{\varepsilon_{r1}}}}{\frac{1}{\sqrt{\varepsilon_{r2}}} + \frac{1}{\sqrt{\varepsilon_{r1}}}} \times \frac{\sqrt{\varepsilon_{r2}}\sqrt{\varepsilon_{r1}}}{\sqrt{\varepsilon_{r2}}\sqrt{\varepsilon_{r1}}} \right| \tag{11.32}$$

$$\rho = \left| \frac{\sqrt{\varepsilon_{r1}} - \sqrt{\varepsilon_{r2}}}{\sqrt{\varepsilon_{r1}} + \sqrt{\varepsilon_{r2}}} \right|$$

where: Z_1 – Characteristic impedance of material 1 $\sqrt{\mu_o / \varepsilon_o \varepsilon_{r1}}$,

Z_2 – Impedance of the material 2 $\sqrt{\mu_o / \varepsilon_o \varepsilon_{r2}}$,

Z_o – Characteristic impedance of free space $\sqrt{\mu_o / \varepsilon_o}$.

The reflection coefficient into the material from the air where $\varepsilon_{r1} = 1$

$$\rho = \left| \frac{Z_1 - Z_o}{Z_1 + Z_o} \right| = \frac{\sqrt{\varepsilon_{r1}} - 1}{\sqrt{\varepsilon_{r1}} + 1} . \tag{11.33}$$

Absorption of EM radiation by solids is determined by their relative dielectric constant, ε_r, and the loss tangent, $\tan\delta$, of the material. For most materials this is a function of frequency.

The attenuation in dB for propagation through an unbounded dielectric material is

$$\alpha_d = 27.3\sqrt{\varepsilon_r} \tan\delta \frac{d}{\lambda_o} , \tag{11.34}$$

where; α_d – One way attenuation (dB),

ε_r – Relative dielectric constant,

$\tan\delta$ - Loss tangent,

d – Distance (m),

λ_o – Wavelength (m).

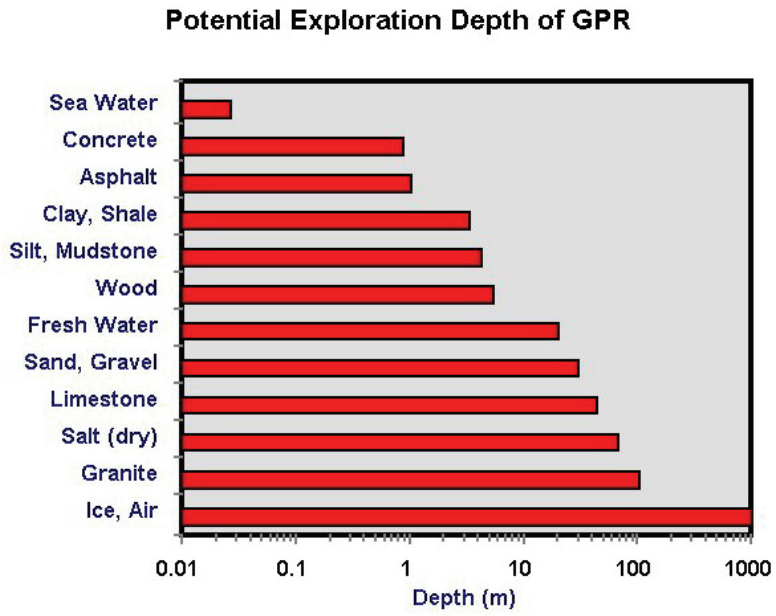


Figure 11.60: Potential exploration depth for GPR

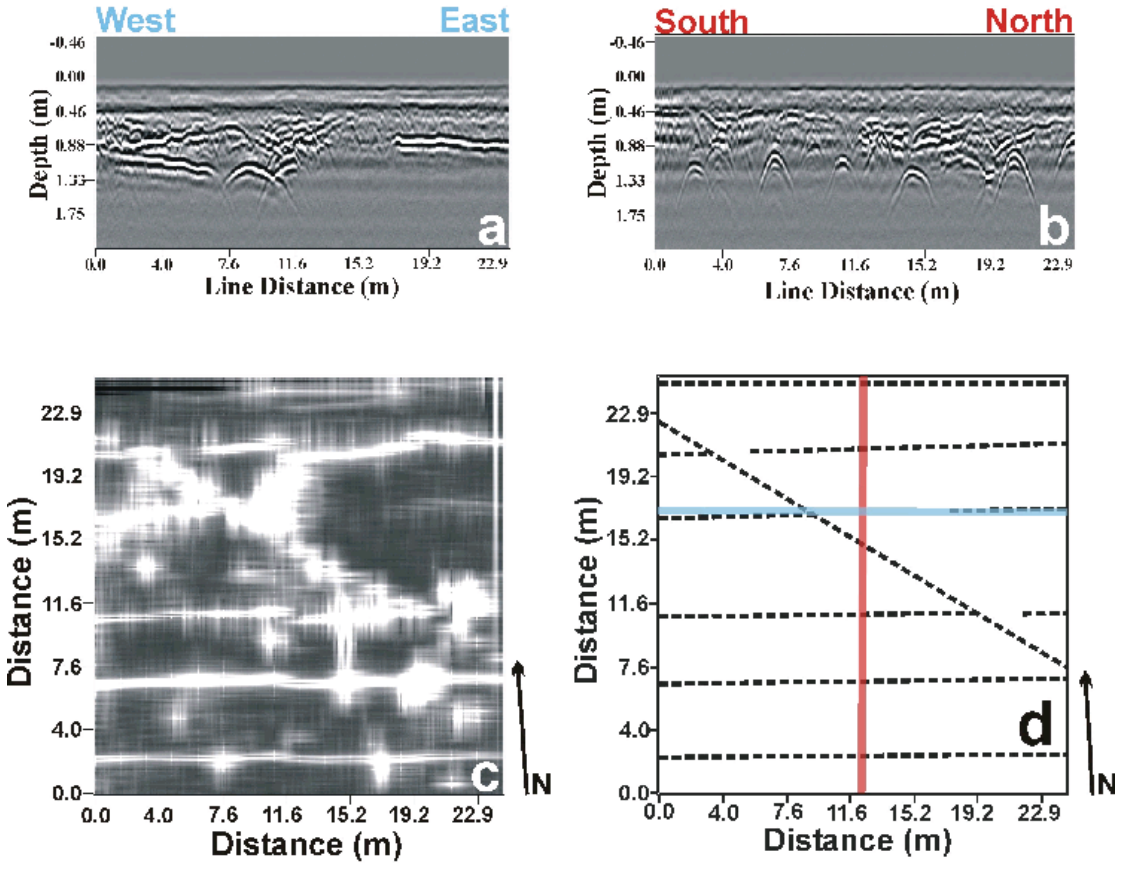


Figure 11.61: GPR results for agricultural drainage pipe location. (a) and (b) are GPR images along orthogonal axes. (c) is a reflectivity map for depths between 0.9 and 1.4m and (d) is an interpreted map of the area showing the positions of the two cuts

Table 11.6: Applications of GPR

Engineering	Geotechnical	Mining	Environment
Bedrock profile	Karst topography	Reef delineation	Buried tanks & drums
Sinkholes	Low density zones	Fault delineation	Contamination plumes
Leaks from services	Sedimentary layers	Depth of weathering	Geological structures
Void detection	Old excavations	Water fissures	Dam situation
Service detection	Depth of fill	Dykes	Bridge scour
		Fracture mapping	

11.13.2. Application Example

A GPR designed to find nodules of ruby embedded in rock generates a pulse with an amplitude of 500V and a duration of 0.5ns. What is the received signal level from the ruby nodule described below

Rock with the following properties

$\varepsilon_{r1} - 2.25$
$\tan\delta$ - 0.005

A nodule of pure ruby with a diameter of 10cm is at a depth of 2.5m

$\varepsilon_{r2} - 6.6$
Tanδ - 0.001

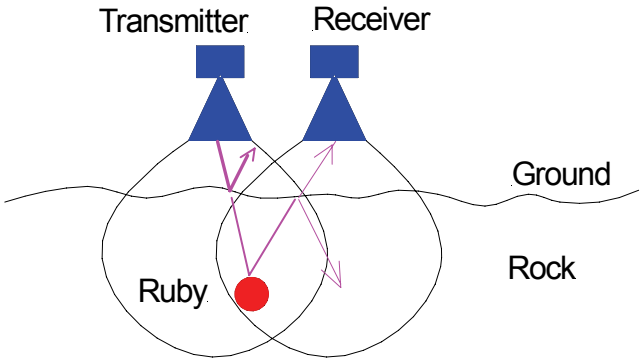


Figure 11.62: Operational scenario

If the radiated signal amplitude (E field) is 1 and the reflection coefficient as the EM wave enters the rock is

$$\rho = \frac{\sqrt{\varepsilon_{r1}} - 1}{\sqrt{\varepsilon_{r1}} + 1} = 0.2 .$$

Then the amplitude of the reflected signal is 0.2 and the transmission coefficient is 1-ρ = 0.8, so the amplitude of the transmitted signal is 0.8.

The reflection coefficient as the EM wave strikes the nodule will be

$$\rho = \left| \frac{\sqrt{\varepsilon_{r1}} - \sqrt{\varepsilon_{r2}}}{\sqrt{\varepsilon_{r1}} + \sqrt{\varepsilon_{r2}}} \right| = 0.26 .$$

So the amplitude of the reflected signal will be 0.8×0.26 = 0.208

Back at the surface, the transmission coefficient is still 0.8, so the amplitude of the signal that enters the receiver antenna is 0.208×0.8=0.1664

The power is proportional to the square of the amplitude so the received echo power compared to the transmitted power in dB is

$$\frac{P_{rec}}{P_{tx}} = 20\log_{10}(0.1664) = -15.5dB .$$

Note that the propagation velocity is reduced by the square root of the dielectric constant

$$c_* = \frac{c}{\sqrt{\varepsilon_r}} \text{ m/s.}$$

The range resolution is a function to the pulsewidth of the signal transmitted and the propagation velocity in the rock

$$\delta R = \frac{c_* \tau}{2} = 0.05m .$$

The transmitter power P_{tx} is proportional to the square of the voltage divided by the circuit impedance. For $Z = 50\Omega$ and $V = 500$V, the transmitter power is

$$P_{tx} = 10\log_{10}\frac{V^2}{Z} = 37dBW .$$

Assuming that the antenna transmits uniformly over the lower hemisphere, it will have a gain of 3dB. This will be reduced by 20dB for an efficiency of 1% to –17dB for both receiver and transmitter.

For a rectangular pulse with a duration τ, the spectrum will have the form shown below. For τ = 0.5ns, the maximum frequency at the first zero is 2GHz, and the average frequency over the band 0Hz to 2GHz will be 374MHz.

See Chapter 7 for an analysis of the Micro Impulse Radar which can be used as a Ground Penetrating Radar

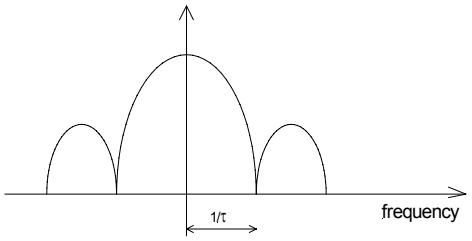


Figure 11.63: Spectrum of pulsed GPR

For f_{ave} = 374MHz, the wavelength λ_{ave} = 0.8m will be used in the range equation.

Because the diameter of the nodule is small compared to the wavelength, to calculate the scattering cross-section, the Rayleigh formula is used. This is modified because the effects of the dielectric have already been considered.

$$\sigma_{dB} = 10\log_{10}\frac{128\pi^5}{3}\frac{a^6}{\lambda^4} = -33dBm^2 .$$

The attenuation per metre (one way)

$$\frac{\alpha_d}{d} = 27.3\sqrt{\varepsilon_r}\tan\delta\frac{1}{\lambda_o} = 27.3\times 1.5\times 0.005\times\frac{1}{0.8} = 0.256dB/m .$$

For a total distance travelled of 2.5×2 = 5m, the attenuation will be 0.256×5 = 1.28dB

Applying the radar range equation with the losses due to attenuation and transmission coefficients etc, the received power is

$$P_{rec} = P_{tx} + 2G_{ant} + 10\log_{10}\frac{\lambda^2}{(4\pi)^3} + \sigma_{dB} - L ,$$

P_{rec} = 37 – 2×17 – 34.9 – 33 – 15.5 – 1.28 = –81.7dBW.

Assuming a 50Ω input impedance, the received echo will have an amplitude of 0.58mV.

The matched filter bandwidth should be about 2GHz, so the thermal noise level will be

$$N = 10\log_{10} kTB = 10\log_{10}\left(1.38\times 10^{-23}\times 290\times 2\times 10^9\right) = -111dBW .$$

A wideband amplifier will have a noise figure of about 4dB, so the final noise level will be –107dBW

The received signal to noise ratio will be

SNR = –81.7+107 = 25.3dB

This should be sufficient to see the target quite easily

Because the antenna beamwidth is very wide, target angular resolution is poor. As the radar unit is dragged over the ground, the apparent range to the nodule will change, and an hyperbolic echo will result.

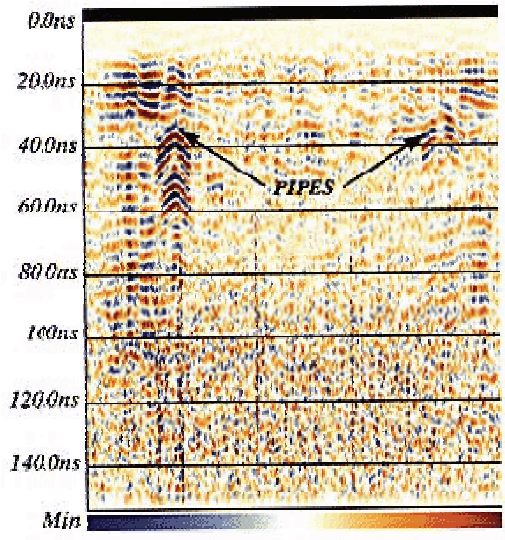


Figure 11.64: GPR image of point targets as the radar is dragged along the ground

11.14. Application: 2D Medical Ultrasound

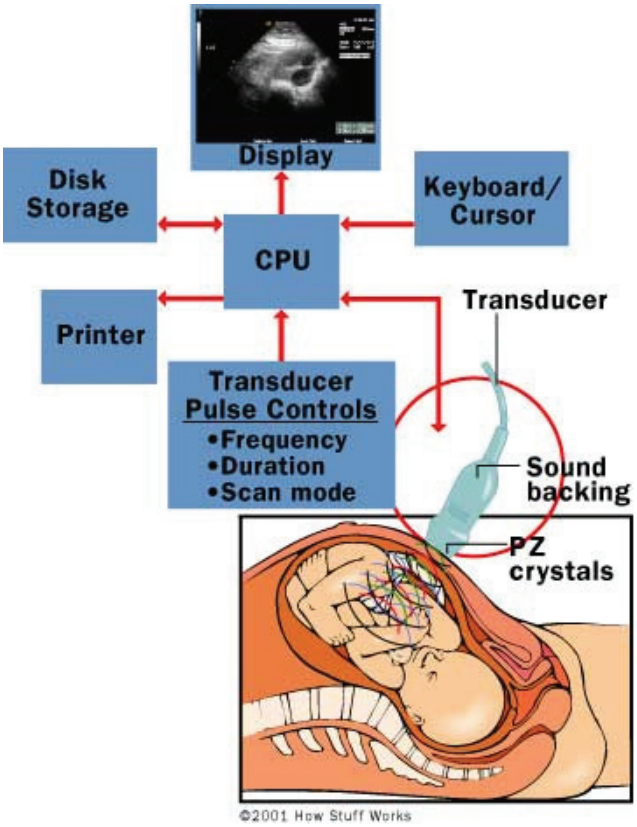


Figure 11.65: 2D ultrasound components

Ultrasonic scanning in medical diagnosis uses the same principle as sonar. Pulses of high-frequency ultrasound, generally between 1 and 5MHz, are created by a piezoelectric transducer and directed into the body. As the ultrasound traverses various internal organs, it encounters changes in acoustic impedance, which cause reflections.

The speed of sound in the tissue (mostly water) is about 1540m/s

The amount and time delay of the various reflections can be analysed to obtain information regarding the internal organs. In the B-scan mode, a linear array of transducers is used to scan a plane in the body, and the resultant data is displayed on a television screen as a two-dimensional plot of range and angles with intensity encoding for reflected signal strength.



Figure 11.66: 2D ultrasound hardware and application

There are many different probe types of which a sample of 3 are shown in the figure above. The shape of the probe determines the field of view.

Because the penetration depth of high frequency (high resolution) ultrasound is limited, probes are often designed for insertion into the body via its various orifices.

Probes with multiple transducers can be phased to steer the beam

The A-scan technique uses a single transducer to scan along a line in the body, and the echoes are plotted as a function of time. This technique is used for measuring the distances or sizes of internal organs.

The M-scan mode is used to record the motion of internal organs, as in the study of heart dysfunction.

Greater resolution is obtained in ultrasonic imaging by using higher frequencies. A limitation of this property of waves is that higher frequencies tend to be much more strongly absorbed.

11.14.1. Applications

Most medical applications were developed specifically to reduce the risks to the foetus of ionising X-rays. They include the following:

- Measuring foetus size to establish due date,
- Determining foetus position to see whether it is breech or head-down for birth,
- Checking the placenta to see that it is properly formed and not obstructing the cervix,
- Counting the number and sex of foetuses,
- Detecting whether a fertilised egg has implanted in the fallopian tubes (ectopic pregnancy),
- Determining the volume of amniotic fluid,
- Monitoring foetus during specialised procedures such as amniocentesis.

Non Obstetric uses for ultrasound are also common:

- Looking for tumours on ovaries and breast,
- Imaging the heart to identify abnormal structures or functions,
- Measuring blood flow (Doppler see Chapter 14),
- Seeing kidney stones,
- Early detection of prostate cancer.



Figure 11.67: 2D ultrasound scan of a 12 week foetus

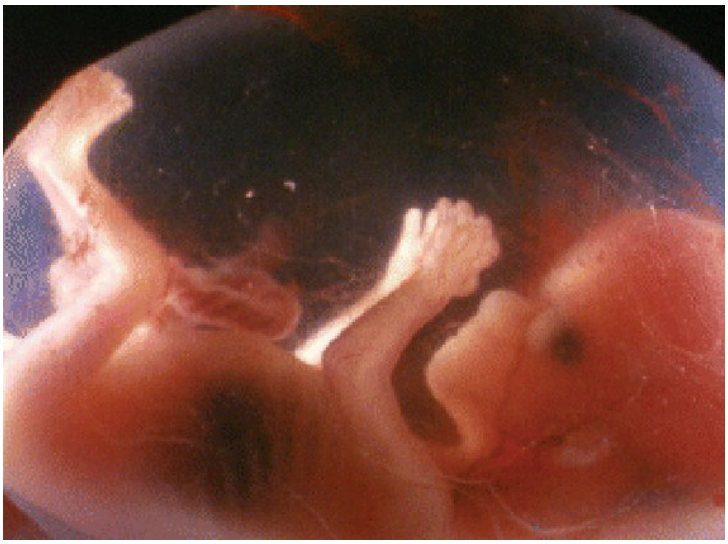


Figure 11.68: Photograph of a foetus of similar age (12 weeks)

11.14.2. Dangers

Two potential dangers exist. They are localised heating due to the absorption of energy, and the formation of bubbles (cavitation) where the ultrasound induces dissolved gases to leave solution.

11.15. References

[1] Electronic Distance Measurement, http://geomatrics.eng.ohio-state.edu/GS400_Notes/Notes_10/GS400_Notes_10.html, 07/03/2001.

[2] Electronic Distance Measurement, http://www.solent.ac.uk/hydrography/notes/horizon/extend/ extend3.html, 07/03/2001.

[3] MRA-101 Tellurometer, http://www.gmat.unsw.edu.au/final_year_thesis/f_pall/html/e7.html, 07/03/2001.

[4] CA-1000 Tellurometer, http://www.gmat.unsw.edu.au/final_year_thesis/f_pall/html/e10.html, 07/03/2001.

[5] MA-100 Tellurometer, http://www.gmat.unsw.edu.au/final_year_thesis/f_pall/html/e8.html, 07/03/2001.

[6] Tellurometer MA-200. Plessey Tellumat Brochure.

[7] GEC-Marconi, Brimstone Presentation to Atlas Aviation, 07/07/1994

[8] Rockradar, http://www.isi.co.za, 25/02/1999

[9] Radarscan, http://radarscan.com, 6/02/1999.

[10] How Stuff Works, How Ultrasound Works, http://howstuffworks.com/ultrasound.html. 04/04/2001.

[11] M.Skolnik, *Rasar Handbook 2nd Ed.* McGraw Hill. 1990

[12] AD8302, http://www.analog.com/UploadedFiles/Data_Sheets/797075782AD8302_a.pdf

[13] Agricultural Drainage Pipe Detection http://www.ag.ohio-state.edu/~usdasdru/radar.htm,

Chapter 12.
High Angular Resolution Techniques

12.1.Introduction

For imaging systems (not null steering trackers), the angular resolution is limited by the beam divergence. Beam divergence θ_{3dB} (beamwidth) is a function of the wavelength and the aperture size

$$\theta_{3dB} = \frac{k\lambda}{d}, \tag{12.1}$$

where θ_{3dB} – 3dB Beamwidth ,
k – Constant (70 for ° and 1.22 for rad),
λ - Wavelength (m),
d – Aperture diameter (m).

The cross range resolution δ_{XR} is the product of the beamwidth (rad) and range (m)

$$\delta_{XR} = R\theta_{3dB} = \frac{1.22R\lambda}{d} \text{ (m)} \tag{12.2}$$

As the wavelength is often fixed by atmospheric window, propagation effects or physical size constraints and it is difficult to make the antenna diameter, d, arbitrarily large to obtain the required angular resolution because of manufacturing limitations, a solution is to use two or more transceivers in an array to synthesise an effective linear aperture equal to the array baseline.

Other techniques that use the forward motion of the antenna to synthesise a larger aperture (in one dimension only) include Doppler Beam Sharpening and Synthetic Aperture processing

12.2.Phased Array Synthesis

This technique is used by:

- Phased array radars
- Phased array sonar (linear or 2D) underwater or in the air
- Long baseline radio telescopes

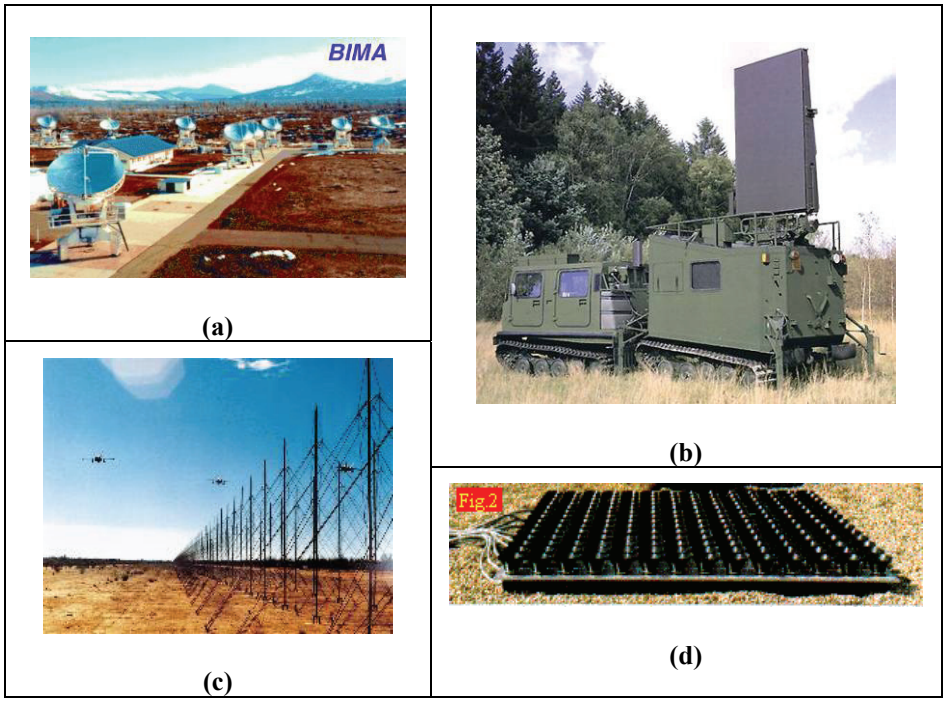


Figure 12.1: Phased array antennas (a) millimetre wave radio telescope (b) mortar locating radar (c) Jindalee over the horizon radar and (d) acoustic array for atmosphere profiling

The power received by each element is the sum of the received power scattered by target, P, from all the transmit elements. The outputs of all N elements are summed via lines of equal length, without delays or phase shifts to give E_a.

Since each element observes phase ϕ_k, the output after summing the signals from all N elements directly, with no phasing is given by:

$$E_a = \sum_{k=1}^{N} \sin(\omega t + \phi_k) \tag{12.3}$$

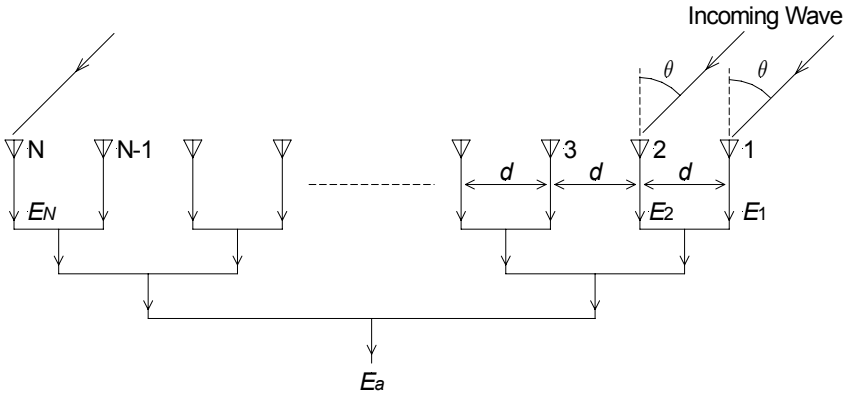


Figure 12.2: The Array beam forming process is achieved by summing the signals from all the array elements in phase

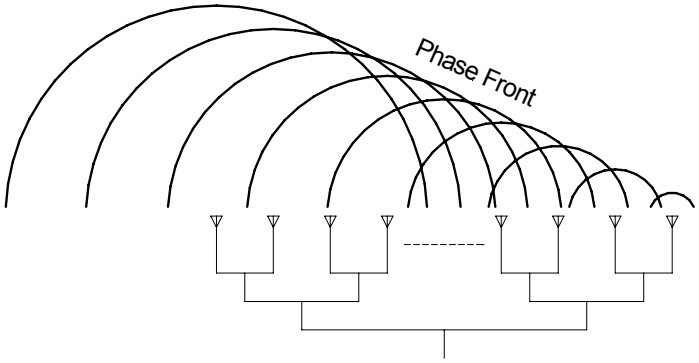


Figure 12.3: Radiation from individual elements with a linearly increasing phase shift combines to form a linear phase-front at an angle to the array as shown by the crests of the waves

For the example of a uniform phase front approaching the array at an angle θ, the relative phase shifts are 0, ψ, 2ψ, 3ψ, …Nψ as shown in Figure 19.4

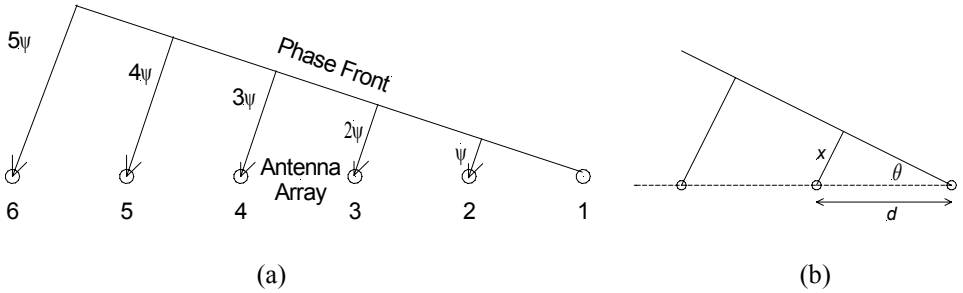


Figure 12.4: Schematic diagram showing (a) the relative phase differences for a plane wave approaching a phased array and (b) construction used to determine the magnitude of the phase difference

It is easy to see that the displacement, x, corresponds to the phase shift between sequential elements and that it can be determined from the array spacing, d, and the angle of arrival of the plane wave, θ

$$x = d\sin\theta \tag{12.4}$$

The phase shift, Ψ, that corresponds to the distance x is then

$$\psi = \frac{2\pi}{\lambda}x \tag{12.5}$$

By substituting for x into (12.5)

$$\psi = \frac{2\pi d}{\lambda}\sin\theta, \tag{12.6}$$

In general, however, the phase shifts are measured relative to the geometric centre of the array, and not element one. The geometric effect of this, is to displace the phase front represented by the line AB down to CD which passes through the geometric centre.

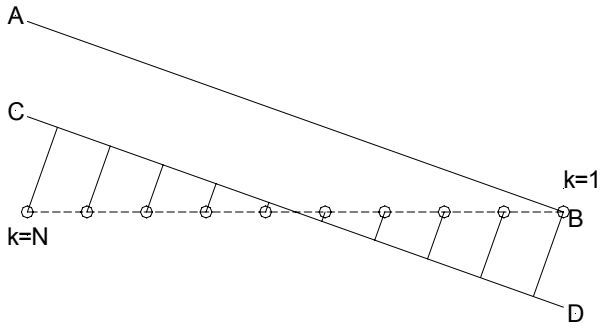


Figure 12.5: Diagram showing the geometric effect of measuring the phase front relative to the geometric centre of the phased array, rather than to element one

For the construction shown, for the phase front AB, the relative phase shift of the elements starting at point B (k=1) is summed as follows

$$E_a = \sum_{k=1}^{N} \sin\left[\omega t + (k-1)\psi\right] \tag{12.7}$$

For the phase front CD that passes through the geometric centre of the array, it can be seen that it is displaced by (N-1) Ψ/2. The output can be re written

$$\begin{aligned} E_a &= \sum_{k=1}^{N} \sin\left[\omega t + (k-1)\psi - \frac{N-1}{2}\psi\right] \\ &= \sum_{k=1}^{N} \sin\left[\omega t + k\psi - \frac{N+1}{2}\psi\right] \end{aligned} \tag{12.8}$$

For the six element array, N = 6, shown in Figure 19.4, the phase shifts become

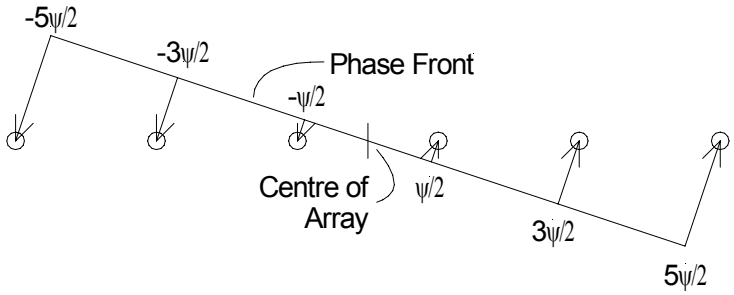


Figure 12.6: Displaced phase shift across array

12.2.1. Two Point Array

Examining the elements on either side of the array centre shown in Figure 19.6, it can be seen that the phase shifts are -ψ/2 and ψ/2 and the sum will be

$$E_a = \sin(\omega.t + \psi/2) + \sin(\omega.t - \psi/2) \tag{12.9}$$

Using the trig identity - $2\sin A\cos B = \sin(A+B) + \sin(A-B)$

$$E_a = \sin(\omega.t).2\cos(\psi/2) \tag{12.10}$$

From the trig identity - $\sin 2A = 2\sin A\cos A$, $2\cos A = \sin 2A/\sin A$ therefore (12.10) can be rewritten as

$$E_a = \sin(\omega.t)\frac{\sin(2\psi/2)}{\sin(\psi/2)} \tag{12.11}$$

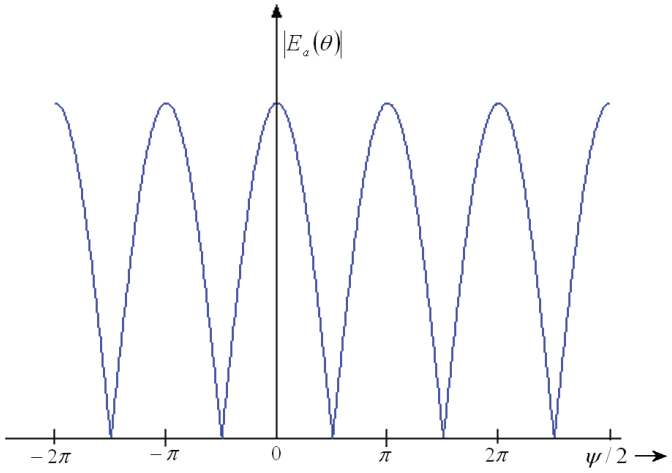


Figure 12.7: Field intensity pattern generated by two isotropic radiators

12.2.2. 4 Point Array

Once again, examining the elements on either side of the array in Figure 19.6, the sum will be

$$E_a = \underset{(1)}{\sin(\omega.t + 3\psi/2)} + \underset{(2)}{\sin(\omega.t + \Psi/2)} + \underset{(3)}{\sin(\omega.t - \psi/2)} + \underset{(4)}{\sin(\omega.t - 3\psi/2)} \tag{12.12}$$

Grouping terms (1),(3) and (2),(4)
and applying the trig identity - $2\sin A\cos B = \sin(A+B) + \sin(A-B)$

$$E_a = 2\sin(\omega.t).\cos(3\psi/2) + 2\sin(\omega.t).\cos(\psi/2) \tag{12.13}$$

Let $A+B = 3\psi/2$ and $A-B = \psi/2$ and applying the same trig identity

$$E_a = \sin(\omega.t).4\cos(2\psi/2)\cos(\psi/2) \tag{12.14}$$

Multiplying by $\sin(2\psi/2)/\sin(2\psi/2)$, simplifying and applying the trig identity $\sin 2A = 2\sin A\cos A$

$$E_a = \sin(\omega.t)\frac{2\sin(4\psi/2)\cos(\psi/2)}{\sin(2\psi/2)} \tag{12.15}$$

Applying the trig identity again, and simplifying, the sum for a 4 point array is

$$E_a = \sin(\omega.t)\frac{\sin(4\psi/2)}{\sin(\psi/2)} \tag{12.16}$$

12.2.3. The General Case

It can be seen that the only difference between the two element and the four element cases, is the size of the numerator. In the general case, for N elements, the numerator can therefore be rewritten as $\sin(N\psi/2)$ and the equation for the sum becomes

$$E_a = \sum_{k=1}^{N} \sin\left[\omega t + k\psi - \frac{N+1}{2}\psi\right] = \sin(\omega.t)\frac{\sin(N\psi/2)}{\sin(\psi/2)} \tag{12.17}$$

Substituting for ψ from (12.6)

$$E_a = \sin(\omega t).\frac{\sin\left\{\frac{N\pi d}{\lambda}\sin\theta\right\}}{\sin\left\{\frac{\pi d}{\lambda}\sin\theta\right\}}. \tag{12.18}$$

The second term of the equation is known as the amplitude factor of the array. The field intensity pattern, $E_a(\theta)$, for the antenna is the magnitude of this amplitude factor.

The general expression for the field intensity pattern is therefore.

$$\left|E_a(\theta)\right| = \left|\frac{\sin\left\{\frac{N\pi d}{\lambda}\sin\theta\right\}}{\sin\left\{\frac{\pi d}{\lambda}\sin\theta\right\}}\right|. \tag{12.19}$$

There are nulls where the numerator is zero: $\frac{N\pi d}{\lambda}\sin\theta$ = 0, +/-π, +/-2π etc

The denominator is also zero at $\frac{\pi d}{\lambda}\sin\theta$ = 0, +/-π, +/-2π etc

Applying L'Hopital's rule where E_a =0/0 it is found that E_a is a maximum where $\sin\theta$ = +/-nλ/d. These maxima will all have the value *N*.

The maximum where $\sin\theta$ = 0 is called the Main lobe, all the other lobes are called Grating lobes.

If d/λ = 0.5, the grating lobe does not appear for *n* = +/-1 in real space because $\sin\theta$>1 which is not possible. If d/λ = 1 the grating lobes appear at +/-90°, and as most real radiating elements do not radiate much at θ = 90°, the grating lobes are suppressed.

For a non scanning array, the best element spacing is $d = \lambda$ while for a scanned array the best spacing is $d < \lambda/2$.

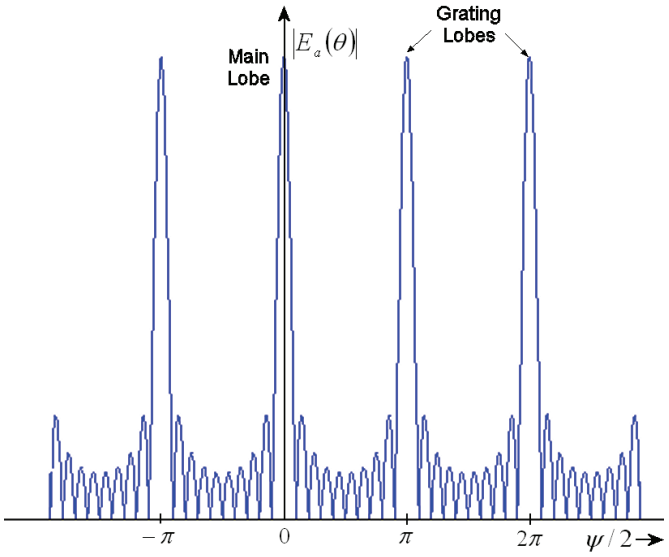


Figure 12.8: Ten element synthesis

12.3.The Radiation Pattern

12.3.1. Linear Array

The radiation pattern is defined as the normalised square of the amplitude factor

$$G_a(\theta) = \frac{|E_a|^2}{N^2} = \frac{\sin^2\left\{\frac{N\pi d}{\lambda}\sin\theta\right\}}{N^2\sin^2\left\{\frac{\pi d}{\lambda}swin\theta\right\}}. \qquad (12.20)$$

For $Nd = L$ (the length of the antenna) and $\mathrm{Sin}\alpha = \alpha$ (small angle aprox.)

$$G_a(\theta) \approx \frac{\sin^2\left\{\frac{\pi L}{\lambda}\sin\theta\right\}}{\left\{\frac{\pi L}{\lambda}\sin\theta\right\}^2}. \qquad (12.21)$$

For $d = \lambda/2$, then the half power beamwidth θ_{3dB} is as follows

$$\theta_{3dB} = \frac{1.73}{N} \text{ (rad)}. \qquad (12.22)$$

If N is sufficiently large, the antenna will be equivalent to a uniformly illuminated aperture, and the first sidelobe will be 13.2dB down.

For directive elements, the antenna pattern is the product of the element factor $G_e(\theta)$ and the array factor $G_a(\theta)$,

$$G(\theta) = G_e(\theta).G_a(\theta). \tag{12.23}$$

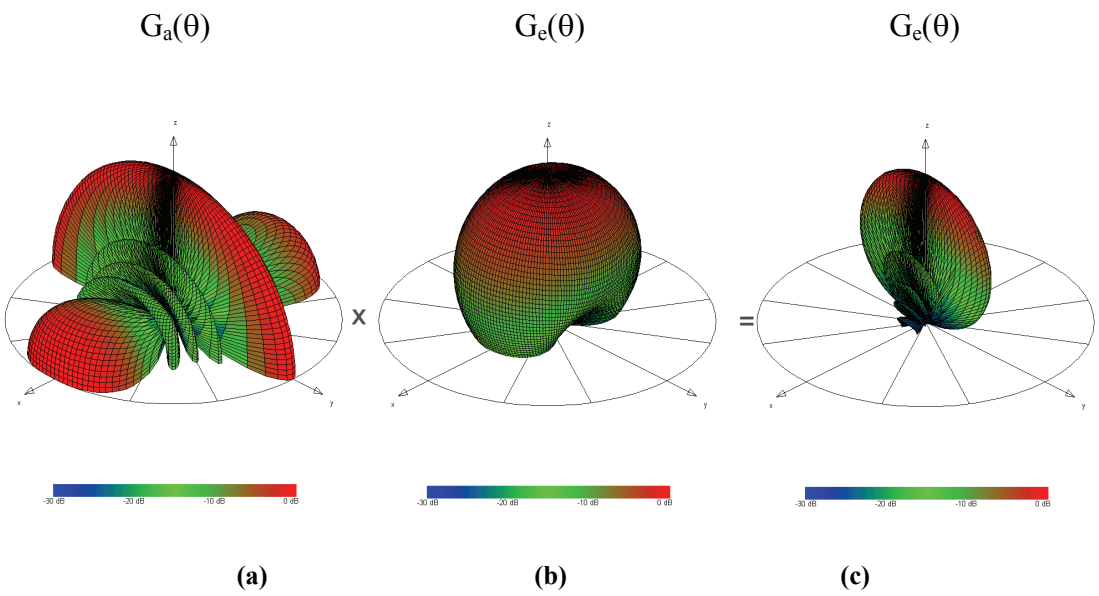


Figure 12.9: Effect of directive elements on the antenna pattern where (a) is the array factor, (b) is the element factor (c) is the final antenna pattern

12.3.2. Radiation pattern: 2D Rectangular Array

The radiation pattern may be approximated as the product of the patterns of the two planes that contain the principle axes of the antenna

$$G(\theta,\phi) = G(\theta)G(\phi), \tag{12.24}$$

$$G(\theta,\phi) = \frac{\sin^2\left\{\frac{N\pi d}{\lambda}\sin\theta\right\}\sin^2\left\{\frac{M\pi d}{\lambda}\sin\phi\right\}}{\sin^2\left\{\frac{\pi d}{\lambda}\sin\theta\right\}\sin^2\left\{\frac{\pi d}{\lambda}\sin\phi\right\}}. \tag{12.25}$$

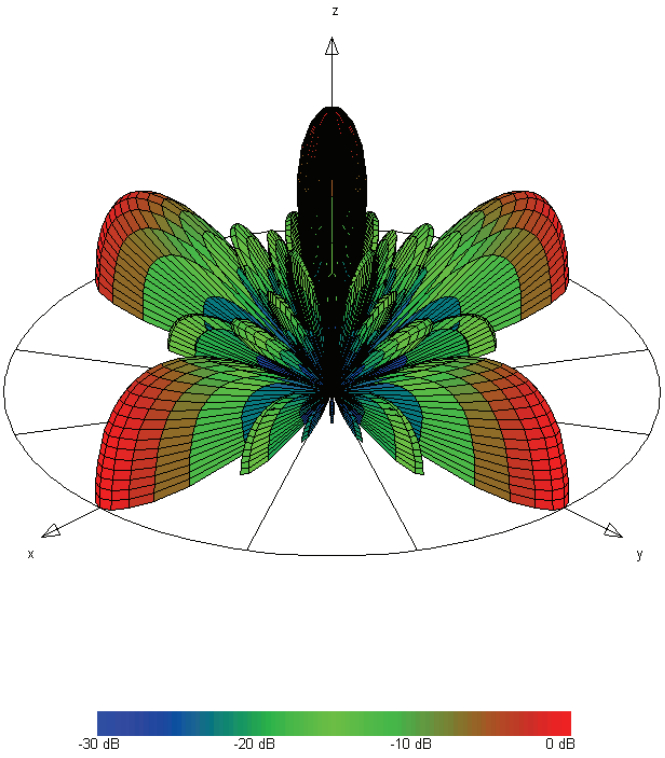


Figure 12.10: 3D polar pattern for a 5x5 element phased array

12.3.3. Antenna Gain

For large arrays, the non-scanned antenna gain can be approximated by the gain of a uniformly illuminated aperture

$$G_o = \frac{4\pi A}{\lambda^2}. \tag{12.26}$$

For a scanned array, the gain is reduced by the scan angle θ_o because the projected aperture is reduced in size

$$G(\theta_o) = \frac{4\pi A \cos\theta_o}{\lambda^2}. \tag{12.27}$$

12.4.Beam Steering

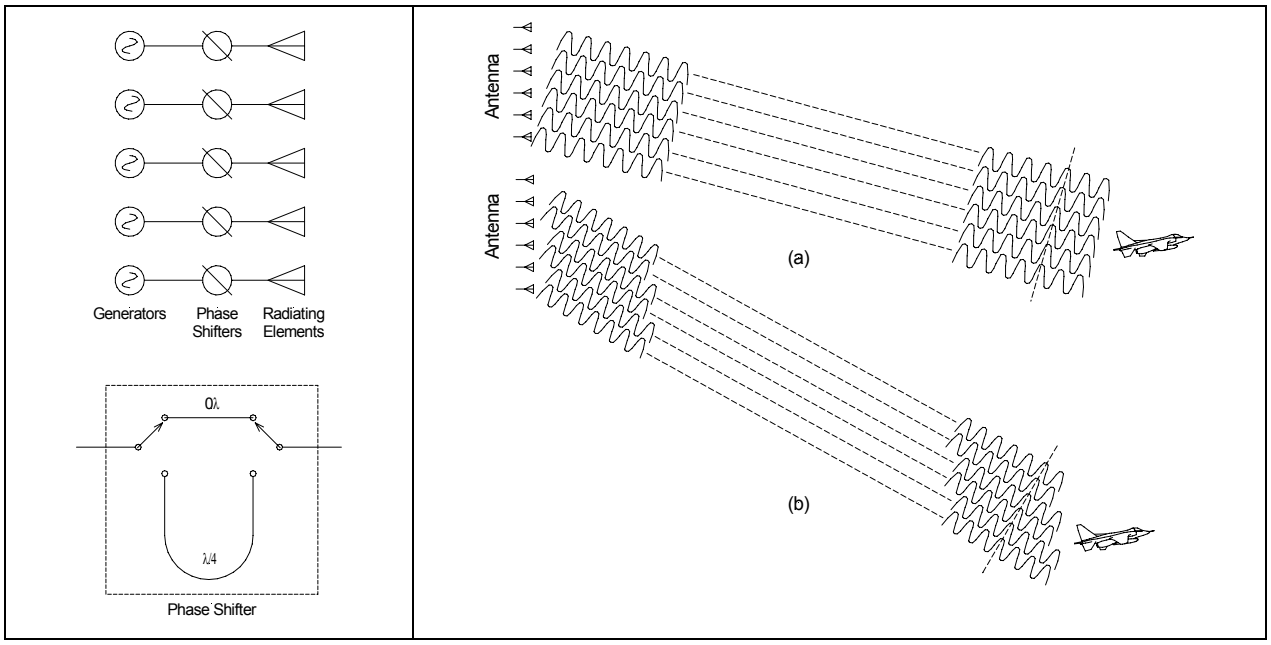


Figure 12.11: Beam steering uses phase shifts to shift the region of constructive interference in the direction required. Note from (b) that the phase shift need never exceed one wavelength

If the same phase is applied to all the elements of the array, then the main beam will be broadside to the array and $\theta = 0$

The direction of the main beam will be θ_o if the relative phase difference is

$$\varphi = \frac{2\pi d}{\lambda}\sin\theta_o . \tag{12.28}$$

The phase of each element will be $\varphi_c + m\varphi$, where m=1,2,3,… N-1 and φ_c is a constant phase applied to all of the elements

The radiation pattern is then

$$G_a(\theta) = \frac{\left|E_a\right|^2}{N^2} = \frac{\sin^2\left\{\frac{N\pi d}{\lambda}(\sin\theta - \sin\theta_o)\right\}}{N^2\sin^2\left\{\frac{\pi d}{\lambda}(\sin\theta - \sin\theta_o)\right\}} . \tag{12.29}$$

Grating lobes will occur at

$$\frac{\pi d}{\lambda}(\sin\theta_g - \sin\theta_o) = \pm n\pi . \tag{12.30}$$

For a scan over +/-90°, the element spacing should be $d = \lambda/2$, and for a practical array that can scan over +/-60°, the spacing $d > 0.54\lambda$.

If the element spacing is not reduced, then as soon as the beam scans, the grating lobes will appear as shown in the following figure.

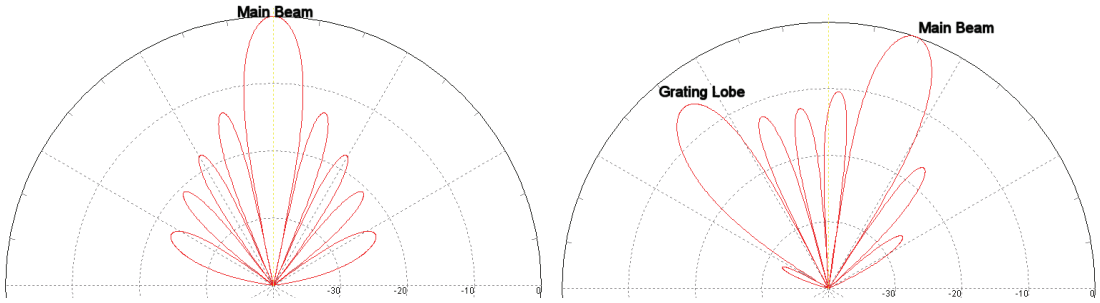


Figure 12.12: Beam scanning for an element spacing d = λ showing the appearance of a grating lobe

12.4.1. Active and Passive Arrays

With the advent of low-cost MMICs it is now practical to manufacture individual transceiver modules to produce active arrays. Such arrays have been manufactured at X-band for less than $15 per module.

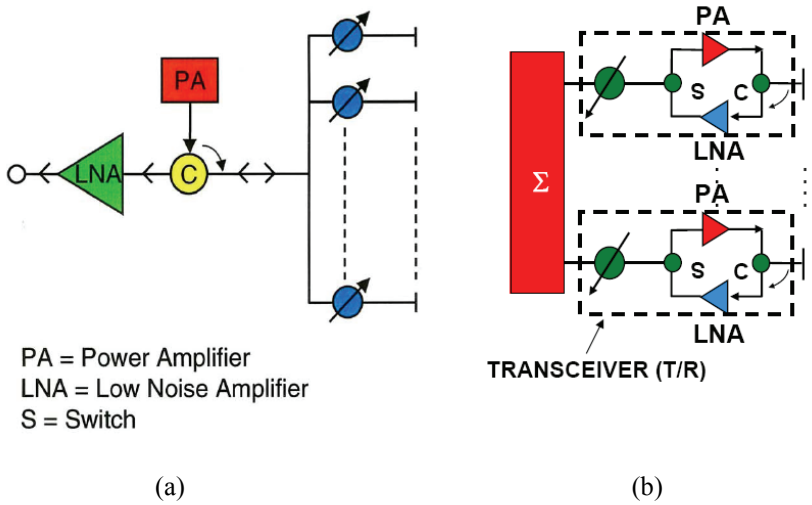


Figure 12.13: Schematic diagram showing the configuration for (a) passive and (b) active arrays

12.4.2. Corrections to Improve Range Resolution

Individual elements have to be fitted with phase delay circuitry to steer the beam. However, as the figure (a) below shows, if the squint angle is great, the pulse gets spread with a resultant decrease in the range resolution. To counter this, it is possible to use adjustable time delay circuits as shown in (b). However, in general, to minimise the amount of circuitry required in a system that might have thousands of elements, blocks of elements are delayed in time, before individual elements are steered using individual phase shifters.

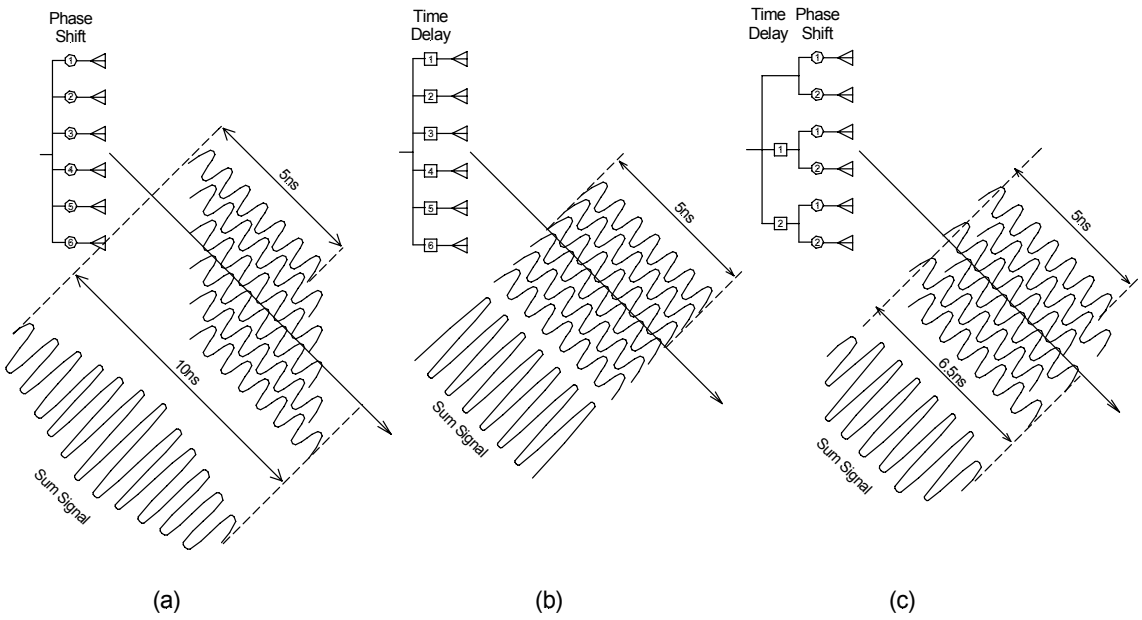


Figure 12.14: Difference between Phase Delay and Time Delay. In (a) which uses only phase delays, the 5ns pulse is stretched out to 10ns. In (b) only time delays are implemented to maintain the 5ns pulse, and in (c) a compromise is implemented to minimise the expense of using individual time delay circuits for each element

12.5.Array Characteristics

12.5.1. Half Power Beamwidth

The half power beamwidth of a scanned array can be approximated by the following formula (not valid near end-fire)

$$\theta_{3dB} \approx \frac{0.886\lambda}{NdCos\,\theta_o} \text{ (rad).} \tag{12.31}$$

Generally some form of taper is used to reduce the sidelobe levels. Using cosine on a pedestal $A_n = a_0 + 2a_1\cos(2\pi n/N)$ where $0 < 2a_1 < a_0$

$$\theta_{3dB} \approx \frac{0.886\lambda}{Nd\cos\theta_o}\left\{1 + 0.636\left[\frac{2a_1}{a_0}\right]^2\right\} \text{ (rad).} \tag{12.32}$$

12.5.2. Matching and Mutual Coupling

The impedance of the array elements varies with the scan angle and spurious lobes may appear due to the miss-match.

This is a difficult problem to solve analytically, and is often determined experimentally by exciting a single element and terminating all of the surrounding ones.

Coupling is proportional to $1/d$ for $d = \lambda/2$, so the pattern and impedance are drastically altered by surrounding elements. Generally the surrounding 5×5 or even 9×9 elements must be considered.

12.5.3. Thinned arrays

Reducing the number of elements leaves the mail lobe unaltered but degrades the sidelobe level. In this example the 4000 element array is thinned to 900 elements. The mean sidelobe level is 31.5dB which is still acceprable

Thinning to 10% reduces the main lobe level down to 10%, but leaves the sidelobe level at 90%

If the removed elements are replaced by matched dummy elements, then the pattern remains unchanged, only the gain is decreased

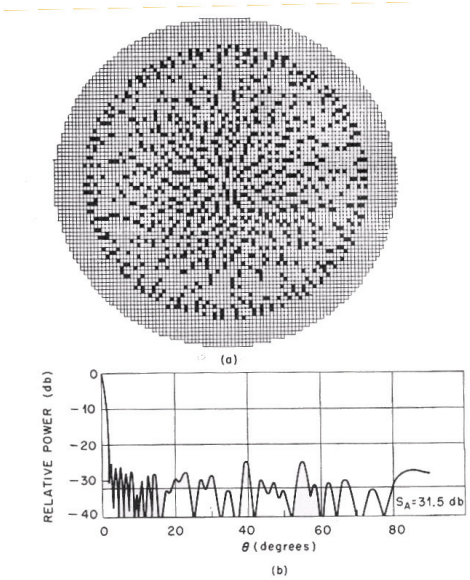


Figure 12.15: Thinned array

12.5.4. Conformal Arrays

It is theoretically possible to place array elements on any arbitrary surface and by applying the correct phase, amplitude and polarisation, to radiate a beam in the required direction

- It is difficult to control the beam-shape and to get low sidelobes
- Feeding mechanism is difficult
- Each element in the array is different (unless there is symmetry) so impedance and coupling will be different
- Computationally difficult as there are no element or array factors.

Most of the work has been done on truncated cones, cylinders or ogives

12.5.5. Advantages of using Phased Arrays

The following are the main advantages of using arrays

- Inertialess rapid bean steering
- Multiple, independent beams
- Potential for large peak and average powers
- Control of radiation pattern
- Graceful degradation
- Convenient aperture shape
- Electronic beam stabilisation

12.6.Applications

12.6.1. Acoustic Array

The imaging acoustic array shown below was developed by Paul Thompson as part of his honours thesis.

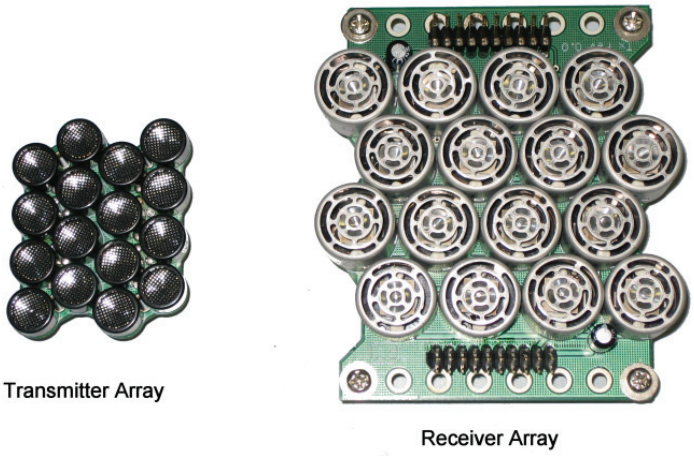
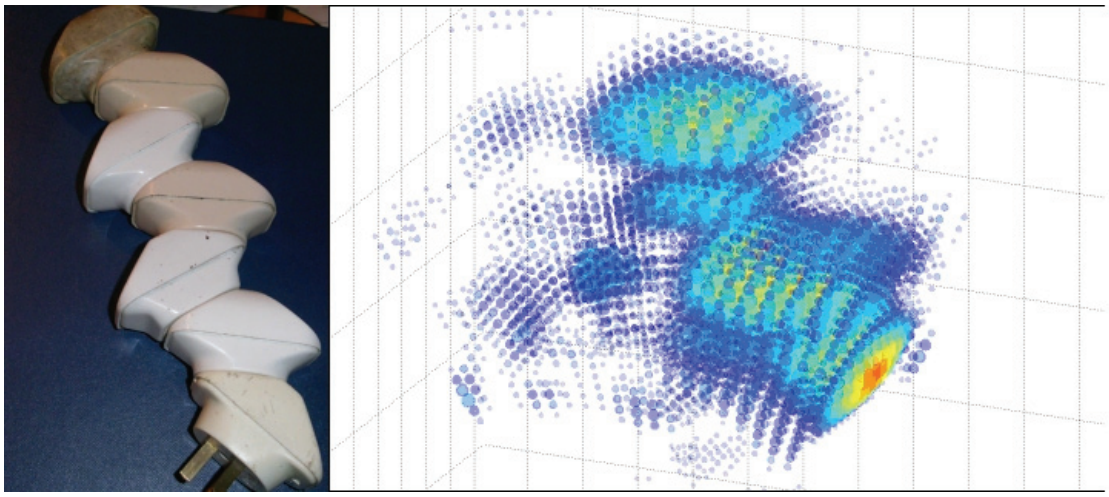


Figure 12.16: Acoustic array hardware

The following image shows a single sample synthesis made using an acoustic array made up from standard ultrasonic transducers.



(a) **(b)**

Figure 12.17: Imaging capability of acoustic phased array (a) photograph and (b) acoustic image (courtesy Paul Thompson)

12.6.2. Early Warning Radar Array

The figure below shows a photograph of Pave Paws, a long range early warning phased array radar that contains 1792 elements each radiating 322W peak power on a 30m wide face. It can detect targets with a $10m^2$ cross-section at a range of 6000km.

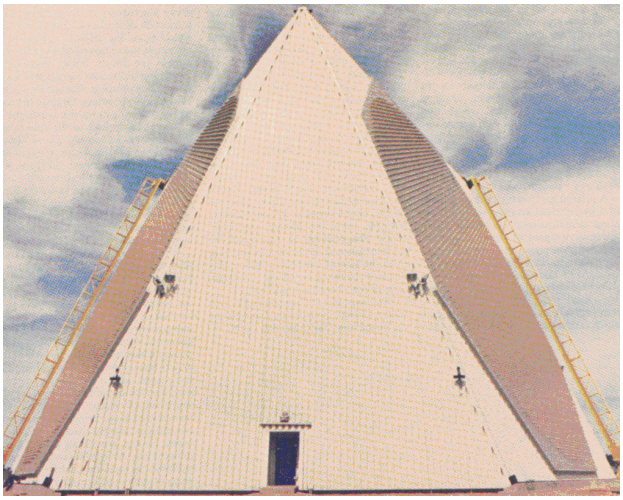


Figure 12.18: Photograph of the Pave Paws early warning radar

The following photograph shows the sea based X-band (SBX) radar which has been developed for ground based midcourse defence. Its specifications are as follows:

- Built on a semi-submersible oil-drilling platform
- Deck area 68×58m
- Draft 19m
- Displacement 50,000 tons
- Sea surface to radome top 65m (24 stories)
- Radome 26m high and 30m in diameter, 7000kg
- Radar, X-band phased array, 65% populated



Figure 12.19: Photograph of the SBX midcourse defence radar

A number of studies have been conducted to determine whether the high levels of microwave radiation emitted by this and other large phased array radars are a health risk.

12.7.Phased Array Application: Side-Scan Sonar

12.7.1. Operational Principle

Similar in principle to side-looking airborne radar (SLAR), the side-scan sonar antenna is a short ($\approx 50\lambda$) linear transducer array made of a piezo-electric material that is towed behind a ship in a straight line.

The transducer is excited by a short ($\tau \approx 3\mu s$) high voltage sinusoidal stimulus at a frequency close to the resonant frequency of the array which the array converts to vibrations and radiates into the water.

The operational frequency is generally between 50kHz and 500kHz with some short-range units operating up to 1MHz.

Because of its shape, the array produces a fan beam pattern with a narrow azimuth beamwidth (typ. 0.75 to 1.5°) determined by the length of the array and a wide elevation beamwidth (typ. 35 to 65°) determined by the vertical aperture of each element.

The beam propagates through the water until it hits the bottom. The roughness of the floor, and any projections reflect sound waves back in the direction of the sonar.

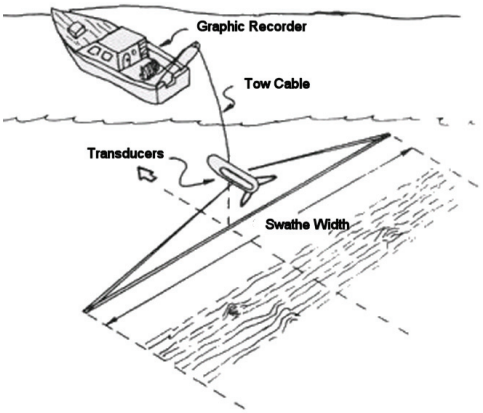


Figure 12.20: Sidescan sonar deployment

The array is generally mounted within a streamlined body called a “tow-fish” that looks like a torpedo. A set of fins keeps the tow body in line with the tow track. Arrays are placed on either side of the tow-fish and angled slightly downward to produce symmetrical beams as shown in the following figure.

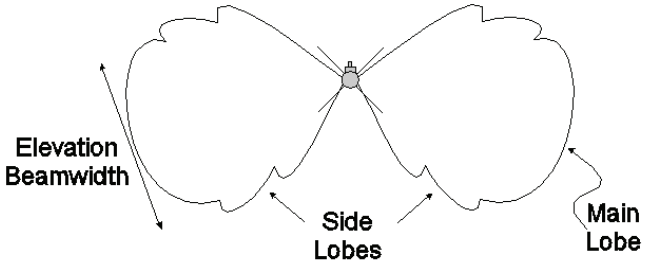


Figure 12.21: Array elevation beam pattern

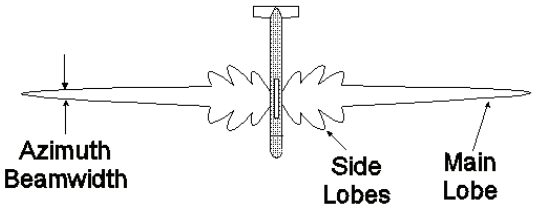


Figure 12.22: Array azimuth beam pattern

The same array is used to receive any echoes. These are then amplified and recorded to form an image. In modern systems, the signal is digitised in the tow-fish and transmitted to the surface for processing and display.

The operational range for low frequency units ($f \approx$ 100kHz) is about 500m, this decreases to 50m at a frequency of 1MHz due to the increasing attenuation of water with frequency (see Chapter 8).

To minimise the geometric distortion, operation is restricted to shallow water, typically 20% of the maximum operating range.

12.7.2. Hardware



Figure 12.23: Tow-Fish showing one transducer array

The specifications of a typical transducer that is mounted within the tow-fish is shown in the figure and table below

Table 12.1: Specifications of ITC-5202 side-scan array

Size of unit	68.5×4×3.8cm
Array dimensions	1.27x53cm shaded active array
Resonance frequency	117kHz
Usable frequency range	111-126kHz
Beam pattern 53cm line 1.27cm line	≈1.5° at 117kHz ≈60° at 117kHz
Efficiency	>40%
Input power <5% duty cycle	1500W
Operating depth	unlimited
Weight	4.3kg
Housing	aluminium



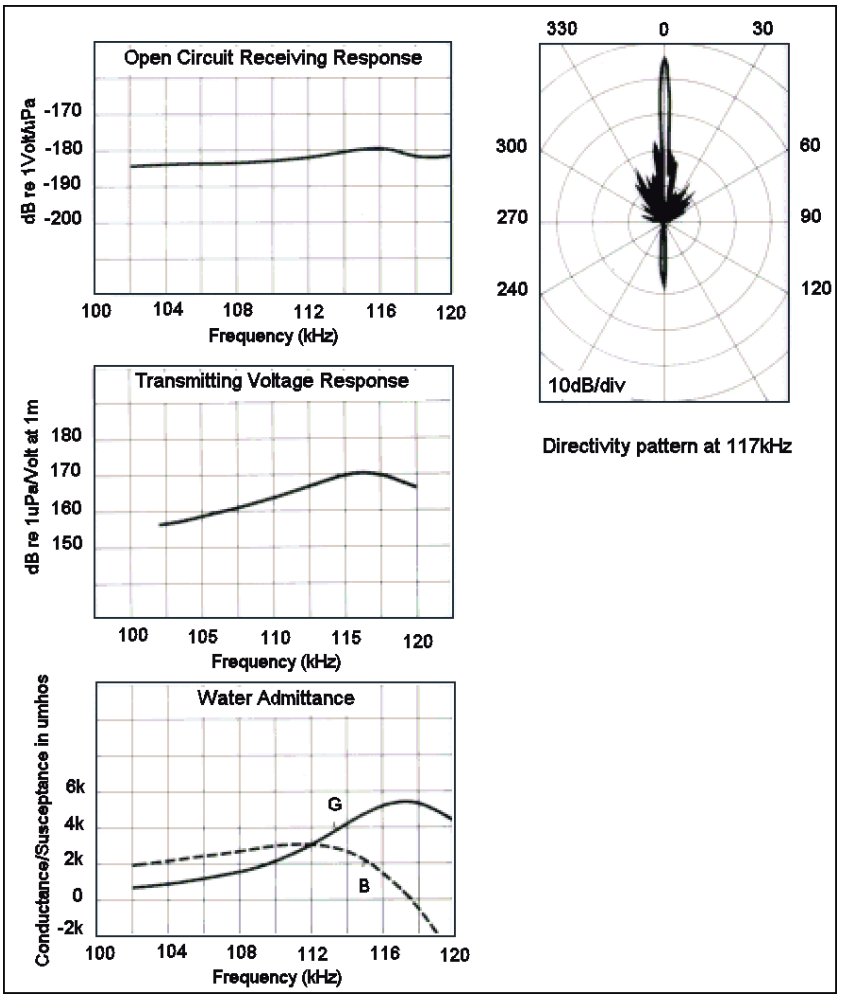


Figure 12.24: ITC sidescan transducer array specifications

12.7.3. Operation and Image Interpretation

The array is towed through the water at a speed between 5kt and 15kt depending on the maximum operational range and the azimuth beamwidth, and the returned echo intensity is plotted as a function of round trip time on the range axis, and synchronised to the tow ship velocity on the cross range axis.

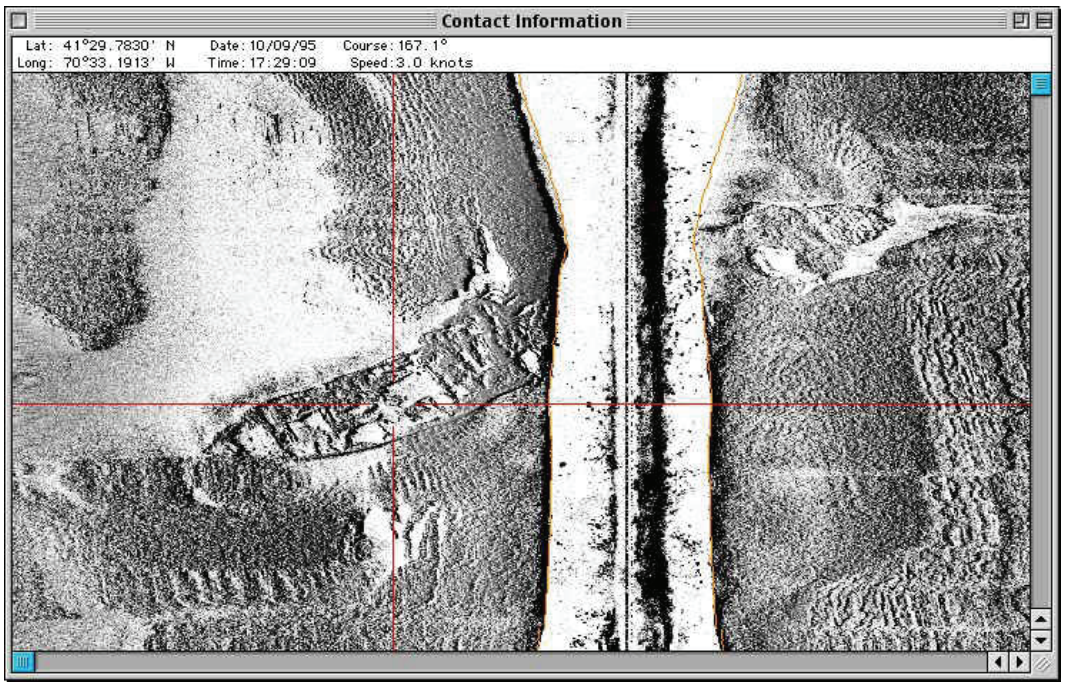


Figure 12.25: Side-scan sonar image of the wreck of the Port Hunter

The black line running vertically through the centre of the image is the bang pulse

The bare white region on either side of this is the slant range from the array to the intersection of the lower edge of the beam with the sea bottom with clutter in this region due to reflecting objects in the beam such as fish, krill etc.

The rest of the image is determined by the reflectivity properties of the sea bottom.



Figure 12.26: The Port Hunter

12.7.4. Shadowing

Because of the low grazing angles and large relative target heights, shadowing is often a pronounced feature of the images.

It can be used to generate an estimate of the height of the target above the sea floor.

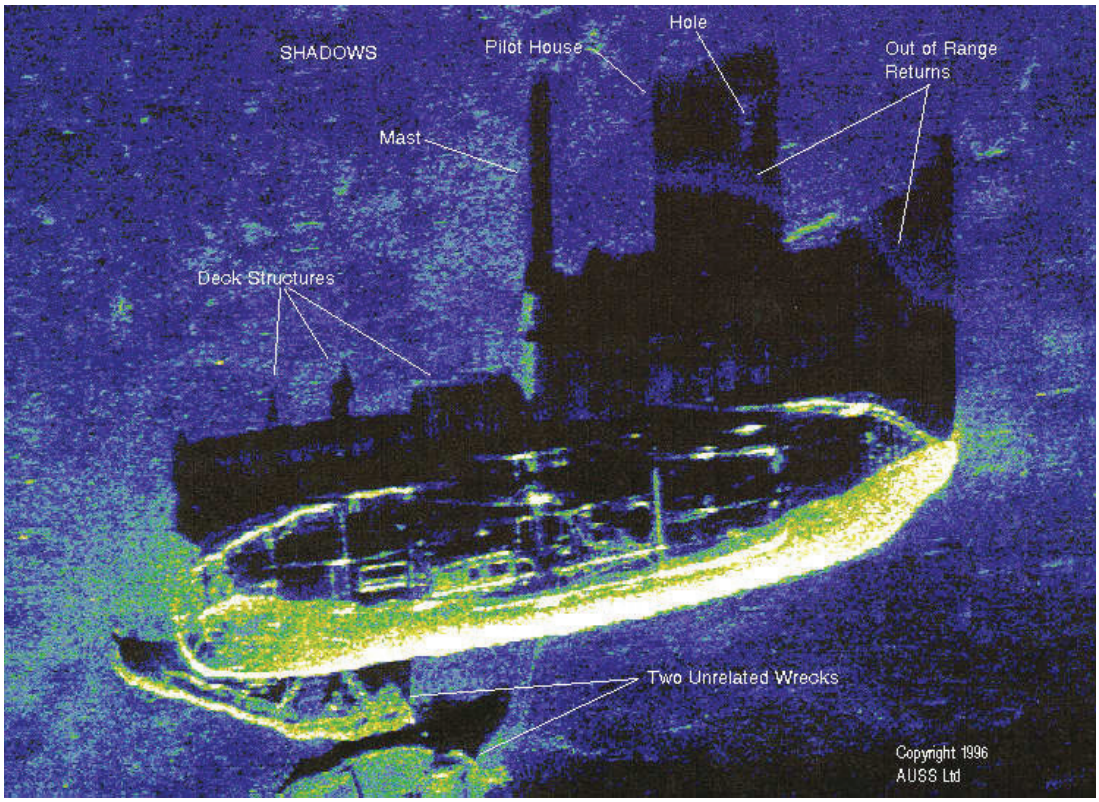


Figure 12.27: Shadowing may be useful ti identify the height if the target as can be seen in this sidescan image of a wreck

12.7.5. Signal Processing

The standard matched-filtering principles developed for radar are applied to sonar systems to ensure that the maximum SNR is achieved.

Most side-scan systems are real beam in that their cross range resolution is a function of range $x = R\theta_{az}$. However digital techniques can be applied to correct for phase front curvature. This is known as focussing, and it can be used to achieve a fairly constant linear beamwidth with range. The process is similar to that used for Synthetic Aperture Radar (SAR).

Beam scanning techniques using phase shifters in the arrays can be used to spotlight particular areas and pseudo 3D images can be created if views are made from more than one perspective.

Simultaneous multi-frequency operation eg. 100kHz and 600kHz is possible for high resolution short range operation and lower resolution long range operation.

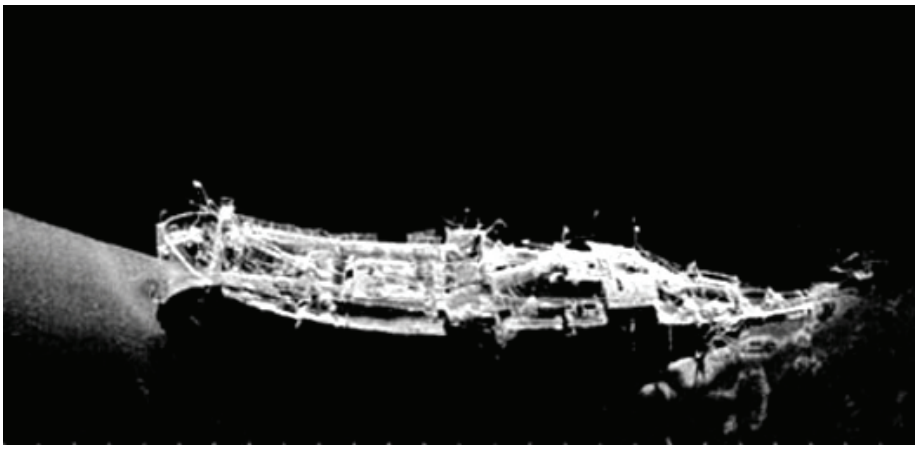


Figure 12.28: The Fritzen in 3D

Note the fine range and angular resolution that can be achieved due to the short operational wavelength

12.7.6. Performance Evaluation (ICT-5202 Transducer)

What is the smallest target that the sidescan sonar based on the ITC-5202 transducer can see at a range of 500m?

It operates at a resonant frequency *fr* = 117kHz which equates to a wavelength λ = 13mm if the velocity of sound is taken to be *c* = 1522m/s

For a useable frequency range (Bandwidth *B*) from 111 to 126kHz, the quality factor *Q* of the transducer is calculated to be

$$Q = \frac{f_r}{f_u - f_l} = \frac{117}{126 - 111} = 7.8. \tag{12.33}$$

The rise time of any pulse generated by the transducer is related to the resonant frequency and the quality factor

$$\tau_{rise} = \frac{Q}{f_r} = \frac{7.8}{117 \times 10^3} = 66.7 \mu s. \tag{12.34}$$

The minimum pulse-width must be at least twice the rise time if the pulse it to reach its peak value. As a rule of thumb, for a "rectangular" pulse, the total pulse-width is $5\tau_{rise}$ = 333μs. This equates to a pulse width, and an effective range resolution of

$$\delta R = \frac{c\tau}{2} = \frac{1522 \times 333 \times 10^{-6}}{2} = 0.25m. \tag{12.35}$$

The cross range resolution at 500m (for the given beamwidth of 1.5°) is the product of the beamwidth in radians and the range

$$\delta XR = R\theta = 500 \times \frac{1.5}{57.3} = 13.1m. \tag{12.36}$$

To operate out to a maximum unambiguous range of 500m, the maximum pulse repetition frequency PRF is

$$PRF_{\max} = \frac{c}{2R_{\max}} = \frac{1522}{2 \times 500} = 1.52Hz\,. \tag{12.37}$$

The transmitter power is limited to a maximum of 1500W for a duty cycle, *DC,* of less than 5%. The duty cycle in this case is

$$DC = 100PRF\tau = 100 \times 1.52 \times 333 \times 10^{-6} = 0.05\%\,, \tag{12.38}$$

which is much smaller than the limit, so the maximum power can be applied to the transmitter.

If the transducer was omnidirectional, then the power density at a range of 1m would be the product of the electrical power, P_{elec}, and the conversion efficiency, η, divided by the surface area of a sphere with a radius of 1m

$$I_{iso} = \frac{P_{elec}\eta}{4\pi} = \frac{1500 \times 0.4}{4\pi} = 47.7W/m^2\,. \tag{12.39}$$

The antenna gain, known as the Directivity Index, *DI*, which is defined in terms of the power with respect to an isotropic radiator can be calculated from the elevation and azimuth beam widths (in radians)

$$G = \frac{4\pi}{\theta\phi} = \frac{4\pi \times 57.3^2}{1.5 \times 60} = 458.4\,. \tag{12.40}$$

The actual power density in the direction of the peak gain is the product of the gain and the isotropic value

$$I = I_{iso}G = 47.7 \times 458.4 = 21864W/m^2\,. \tag{12.41}$$

The sound pressure level (SPL) or *S* is generally given in dB relative to 1μPa at a range of 1m. This can be calculated from the power density and the acoustic impedance *Z* of the water.

The acoustic impedance of water is the product of the density and the velocity

$$Z = \rho_o c = 1026.4 \times 1522 = 1.56 \times 10^6 kg/m^2 \sec\,. \tag{12.42}$$

The relationship between the acoustic pressure, *P,* in Pascals, the power density, *I,* in W/m^2 and the impedance, *Z,* is

$$P^2 = IZ\,. \tag{12.43}$$

This can be re written for the acoustic pressure in μPa for the sound pressure level S as follows

$$S = (10^6 P)^2 = 10^{12} IZ \,. \tag{12.44}$$

This is generally written in dB form

$$\begin{aligned} 10\log_{10} S &= 20\log_{10}(10^6 P) = 10\log_{10} 10^{12} + 10\log_{10}(IZ) \\ S_{dB} &= 120 + 10\log_{10}(21864 \times 1.56 \times 10^6) = 120 + 105.3 = 225.3dB \end{aligned} \tag{12.45}$$

If the transmitting voltage response (TVR) shown in the transducer specification table is examined. It can be seen that at 117kHz, the response is 170dB rel to 1μPa/Volt at 1meter.

The electrical power input P_{elec} is related to the RMS voltage V and the transducer conductance G

$$P_{elec} = V^2 G \,. \tag{12.46}$$

For a conductance G = 5.5k μmho from the transducer specification table and a power P_{elec} = 1500W

$$V = \sqrt{\frac{P_{elec}}{G}} = \sqrt{\frac{1500}{5.5 \times 10^{-3}}} = 522Vrms \,. \tag{12.47}$$

The sound pressure level S_{dB} for a RMS voltage of 522V applied to the transducer is

$$S_{dB} = 170 + 20\log_{10}(V) = 170 + 20\log_{10}(522) = 224.3dB \,. \tag{12.48}$$

This is as expected, because it should equal the sound pressure level calculated from the maximum input power. The 1dB difference could be due to some of the assumptions made, and difficulties with reading the graphs accurately.

As the signal propagates through the water, the sound pressure level reduces because the wave is expanding on a spherical wave-front and due to attenuation. The transmission loss in dB is H and is determined as

$$H = 20\log_{10}\frac{r_2}{r_1} + \alpha_{dB}(r_2 - r_1) \,. \tag{12.49}$$

Because the sound pressure is determined relative to the level existing at one meter from the effective centre of the sound source, the equation can be rewritten for this reference distance as follows

$$H = 20\log_{10} r + \alpha_{dB} r \,. \tag{12.50}$$

The attenuation in dB/m is given by the following formula where f is the frequency of the sound in kHz.

$$\alpha_{dB} = \frac{0.036 f^2}{f^2 + 3600} + 3.2 \times 10^{-7} f^2 = \frac{0.036 \times 117^2}{117^2 + 3600} + 3.2 \times 10^{-7} \times 117^2 = 0.0329 dB/m$$

The target strength, T, in dB is defined by ratio of the reflected sound pressure scattered by the target at a distance of one meter from the effective centre of the scattered sound to the incident sound pressure on the target as follows

$$T = 20 \log_{10} \frac{P_r}{P_i}. \tag{12.51}$$

This target strength is determined by its size, shape and the fraction of sound that is re-radiated.

If the scattering cross section is σ square meters, then T in dB is given by

$$T = 10 \log_{10} \frac{\sigma}{4\pi}. \tag{12.52}$$

A with the radar case, a sphere with a radius, a much larger than the wavelength will have a cross section equal to the projected area and T can be rewritten to take that into account

$$T = 10 \log_{10} \frac{\pi a^2}{4\pi} = 10 \log_{10} \left(\frac{a}{2} \right)^2 = 20 \log_{10} \frac{a}{2} \tag{12.53}$$

The target strength of World War II submarines varied between about 10dB at the bow to 25dB broadside and 15dB at the stern.

For a spherical target, the echo sound pressure level E relative to 1μPa at a range of 1m from the receiver is easily calculated as follows

$$E_{dB} = S_{dB} - 2H + T, \tag{12.54}$$

$$E_{dB} = S_{dB} - 40 \log_{10} r - 2\alpha_{dB} r + 20 \log_{10} \frac{a}{2}. \tag{12.55}$$

If the target is the sea floor, then the cross section will be the product of the range resolution and the cross range resolution modified by a scaling factor to take into account the reflectivity σ^o of the surface.

$$T = 10 \log_{10} \frac{\sigma^o A}{4\pi} = 10 \log_{10} \frac{\sigma^o . \delta R . \delta XR}{4\pi} = 10 \log_{10} \frac{\sigma^o \delta R . r . \theta_{az}}{4\pi}, \tag{12.56}$$

and the echo sound pressure level will be

$$E_{dB} = S_{dB} - 40\log_{10} r - 2\alpha_{dB} r + 10\log_{10} \frac{\sigma^o \delta R.\theta_{az}}{4\pi} r \,. \tag{12.57}$$

Substituting for S_{dB} –225.3dB,
α_{dB} – 0.0328 dB/m,
σ^o – 0.1 m^2/m^2,
θ_{az} – 1.5° (0.0262 rad),
δR – 0.25m,

$$E_{dB} = 225.3 - 40\log_{10} r - 0.0656r - 42.8 + 10\log_{10} r \,. \tag{12.57}$$

As with the radar range equation, detection can only take place if the signal exceeds the noise level by a specific margin. It will be assumed that the noise is white and the signal is sinusoidal so that the same graphs can be used.

The noise level at sea is mostly generated by wind and wave action on the surface. It is proportional to sea-state and inversely proportional to frequency. All the sources of noise are shown in Figure 12.29.

Table 12.2: Noise pressure as a function of sea state

Sea State	Wind Speed (knots)	Noise Pressure N_1 (dB rel 1μPa) Isotropic, 1Hz Bandwidth at 1kHz
6	35	70
3	15	65
2	10	60
1	5	55
0.5	2	50
0	0	45

The frequency relationship to map the noise pressure level at 1kHz to the transducer frequency is

$$N_f = N_1 - 17\log_{10} f_{kHz} \tag{12.58}$$

For a sea state 3 and the transducer frequency of 117kHz

$$N_f = 65 - 17\log_{10} 117 = 30dB \tag{12.59}$$

The total noise pressure level in dB relative to 1μPa must take into account the bandwidth B of the transducer (in Hz) and its directivity or gain G because the sensor is only looking at a small part of the hemisphere.

$$L_N = N_f + 10\log_{10} B - 10\log_{10} G \tag{12.60}$$

$$L_N = 30 + 10\log_{10}(15\times10^3) - 10\log_{10}(458.4) = 45dB$$

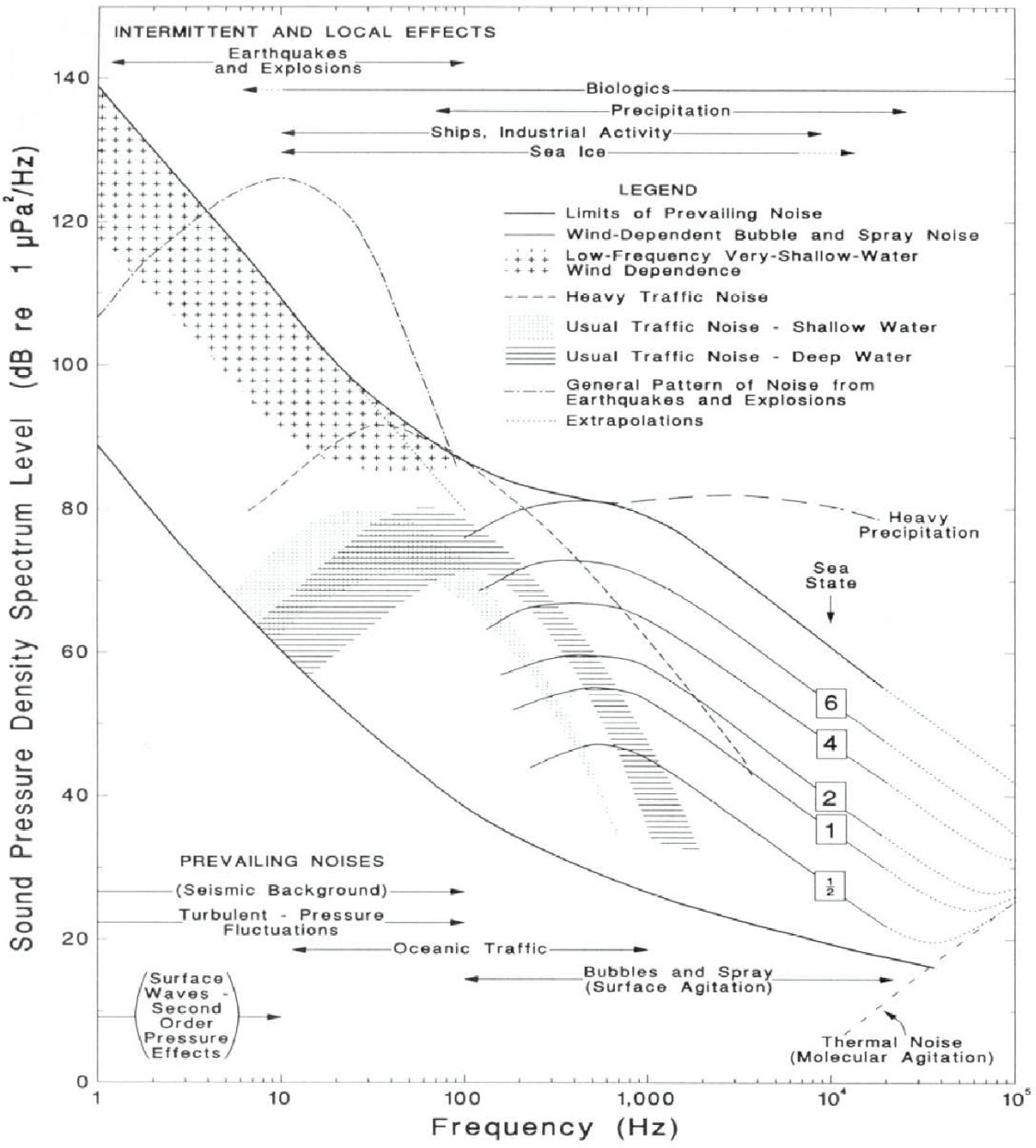


Figure 12.29: Undersea acoustic noise

A second source of noise called reverberation noise is equivalent to volumetric clutter in radar, and will not be considered here, neither will all the other sources of noise shown in the figure above, be considered.

The echo sound level for various perfectly reflecting spheres and the ground and the noise pressure are plotted in the figure below

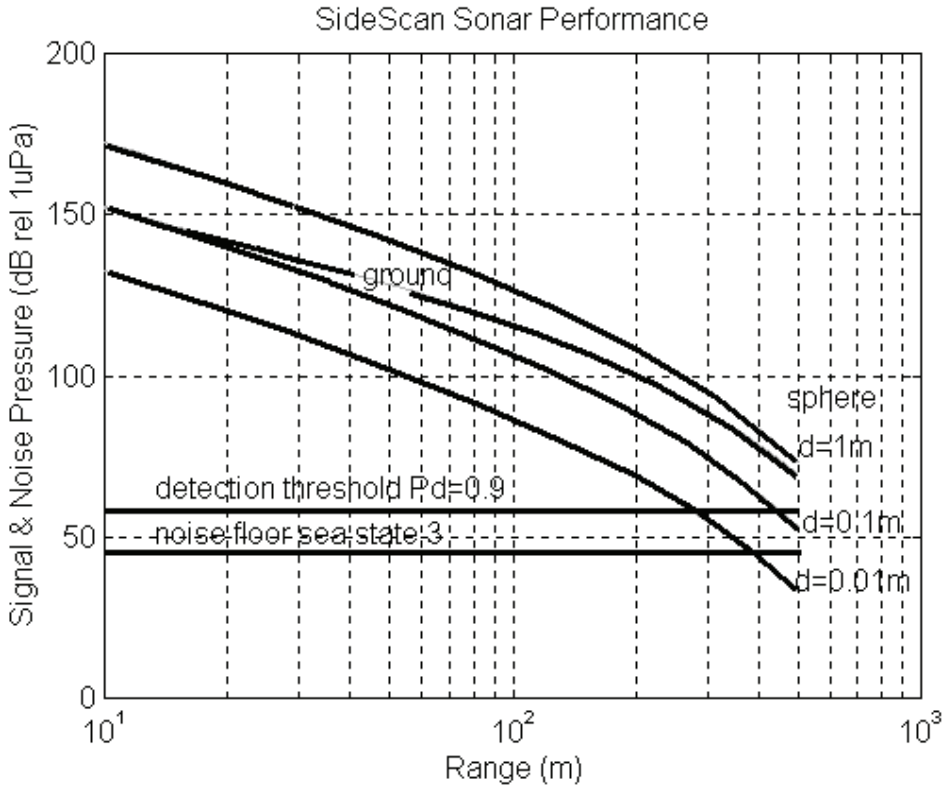


Figure 12.30: Sidescan sonar performance

The sea bottom should be detectable at 500m as would a sphere with a diameter of 1m.

Detection of the 0.1m sphere would be marginal as though the SNR is positive, the 13dB margin required for a $P_d = 0.9$ and a $P_{fa} = 10^{-6}$ is not there.

The actual voltage output by the transducer is determined from the transducer specifications. The open circuit receiving response at 117kHz is –180dB rel 1V/μPa. For a signal pressure of 70dB (the sea-floor return at 500m), the output is

$$20\log_{10}(V) = 70 - 180 = -110dB \tag{12.61}$$

This can be converted to a voltage

$$V = 10^{-110/20} = 3.16\mu V \tag{12.62}$$

Which, in this case, is very small, and so receiver noise would be a consideration when the actual detection characteristics of the system were being considered.

12.8.Other Applications

Sonar technology has also become a valuable resource in vulcanology studies. In 2003, NOAA began a survey of submarine volcanoes along the Mariana Arc north of Guam. Seafloor mapping technology has allowed the group to successfully monitor and track the activity of these underwater volcanoes. The figure below shows the Nikko Volcano, imaged using a multibeam sonar system. The history of the volcano

can be deduced from the seafloor map and current activity, such as the thermal plumes and lava flows, can be monitored by geologists.

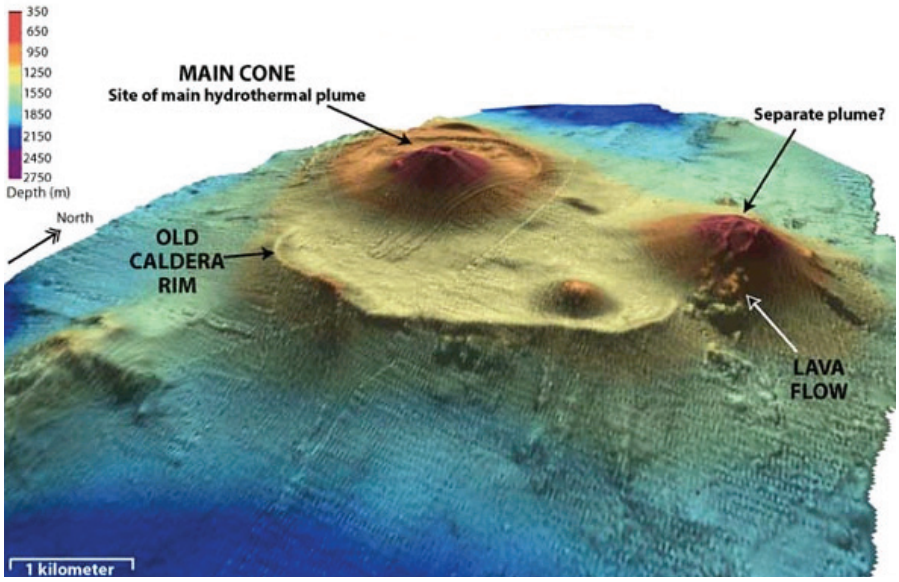


Figure 12.31: Sonar image of the Nikko submarine volcano

12.9.Phased Array Image Gallery

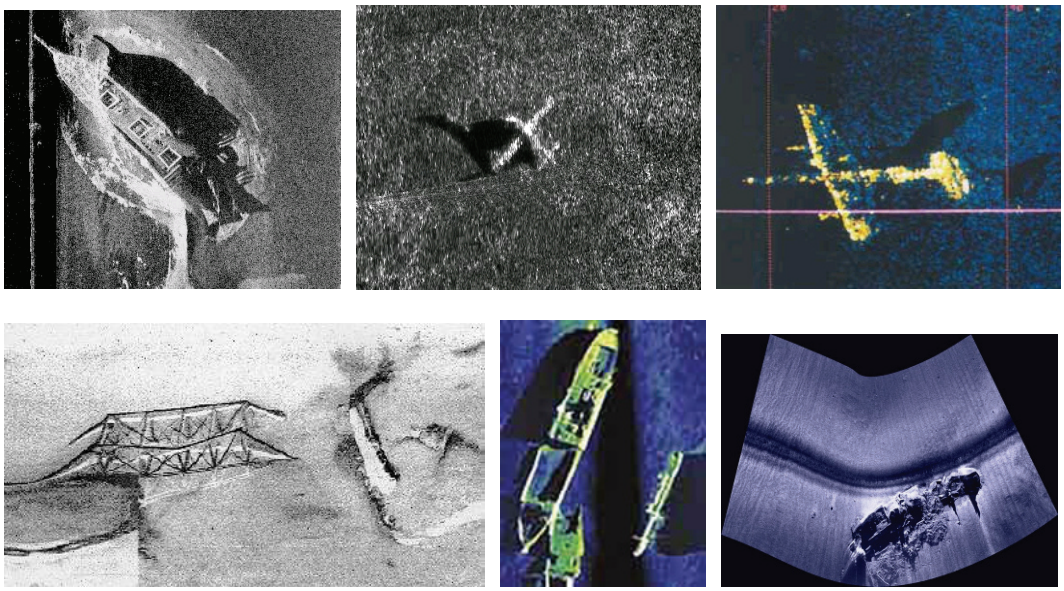


Figure 12.32: A few interesting sidescan sonar images

12.10. Doppler Beam-Sharpening

As an introduction to synthetic aperture processing, it is useful to consider a similar, but more basic signal processing technique called Doppler beam-sharpening. It uses the different Doppler shifts across the antenna beam footprint to narrow the effective beamwidth.

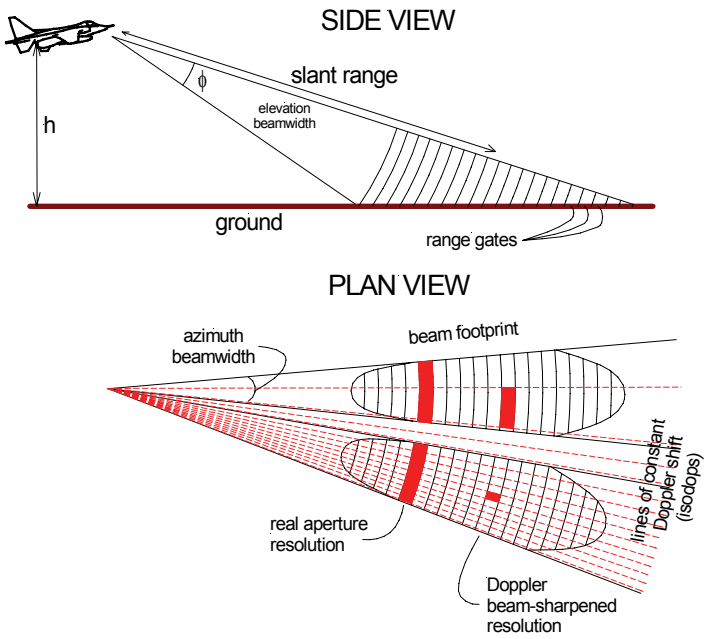


Figure 12.33: Process of Doppler Beam Sharpening

The resolution of a real aperture imaging radar is determined by the range resolution and the azimuth beamwidth, as shown in the figure. If the radar is moving at a constant velocity, then the dotted lines in the figure show lines of constant Doppler shift (isodop lines).

If the beam is offset from the direction of travel, then the isodop lines get closer together, and it becomes easier to isolate smaller sections of the beam by their relative Doppler shifts.

For lines of constant velocity (isovel lines) separated by ΔV m/s, the angular spacing will be

$$\theta_N = \arccos\left(1 - \frac{N.\Delta V}{V_t}\right) \tag{12.63}$$

For V_t = 250m/s and ΔV = 1.25m/s the isovel lines will be at the following angles which correspond to those in the figure.

Table 12.3: Isovel Contours

Velocity (m/s)	Azimuth Angle (deg)
1.25	5.73
2.5	8.11
3.75	9.93
5	11.48
6.25	12.84
7.5	14.07

The Doppler shift of a target viewed from an aircraft travelling at V_t (m/s), at an angle θ (deg) to the direction of travel is

$$f_d = \frac{2V_t}{\lambda}\cos\theta \tag{12.64}$$

If the radar operates at a frequency of 10GHz, then these isovel lines become isodop lines with a separation of 83Hz

One of the limitations to the amount of sharpening that can occur is the observation time. To obtain a resolution of 83Hz in the received Doppler requires that the target be observed for $\tau = 1/\beta = 1/83 = 12$ms. However, at a speed of 250m/s, the target would have moved 3m during this period, and that might be bigger than the range resolution. This is referred to as range walk, and is accommodated in both unfocussed and focussed SAR

For a fixed transmission frequency, there as therefore a trade off that must be made between the range resolution, and the amount of sharpening that occurs.

An alternative is to increase the transmit frequency as this will increase the Doppler shift. For example, if the radar operates at 94GHz, then the separation in the isodop lines increases to 800Hz, and the observation time is reduced to 1.2ms during which time the aircraft only moves 0.3m.

A reflectivity image is constructed by physically scanning the antenna beam to the one side (or both) of the direction of travel, with the pixel size determined by the range resolution and the sharpened angular resolution.

The following figure shows a comparison between a real beam millimetre wave image, and one that has been Doppler beam-sharpened. An examination of point targets in the two images show that even though the cross range resolution has only been improved by a factor of 2, the image quality is much improved.



Figure 12.34: Comparison between a real beam millimetre wave radar image, and one that has been Doppler beam sharpened

12.11. Operational Principles of Synthetic Aperture

The term "synthetic aperture" refers to the distance that the sensor travels during the time that the reflectivity data are collected from a single point. Energy from each scatterer is made to arrive in phase at the output of the processor for all of the samples to realise the narrow beamwidth. The process of synthesis relies on the fact that the wide antenna beamwidth can be subdivided into narrow slivers using the differences in the Doppler shifts across the beam as shown in the figure below.

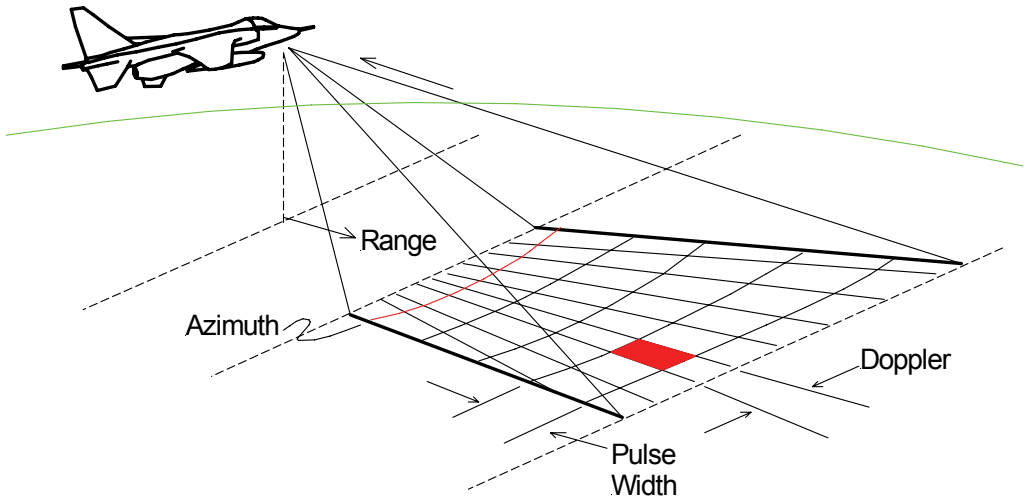


Figure 12.35: Synthetic aperture generation

The process to determine the radiation pattern is similar to that used for the fixed array, with the primary difference being that the signal received by each element is due only to the received power scattered by target P from **one** transmitter element. This results in a slightly different radiation pattern for SAR. The beamwidth is narrower, but the sidelobes are higher than that for the equivalent phased array.

For, $L_{e,}$ the synthetic array length,

$$G_{SAR}(\theta) = \frac{\sin^2\left\{\frac{\pi L_e}{\lambda}\sin\theta\right\}}{\left\{\frac{\pi L_e}{\lambda}\sin\theta\right\}^2}. \tag{12.65}$$

The half power beamwidth can be found by solving for $G_{SAR}(\theta) = 0.5$ and solving graphically (or using Newton) to show that

$$\frac{\pi L_e}{\lambda}\sin\theta \approx \frac{0.886\pi}{2} = 1.39. \tag{12.66}$$

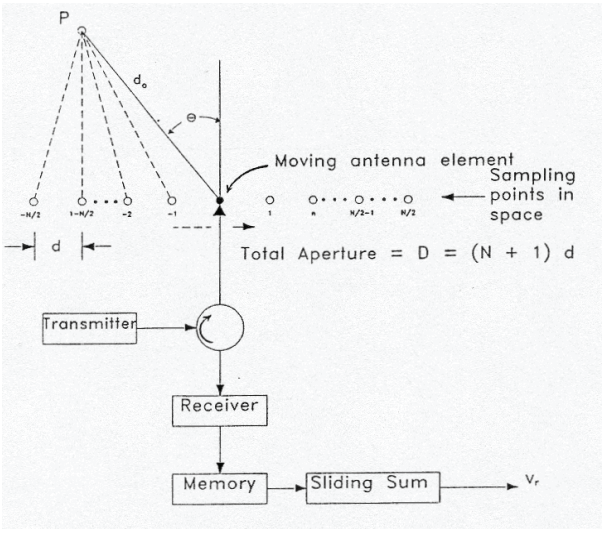


Figure 12.36: Generation of a synthetic array

12.12. Range and Cross Range Resolution

In most SAR applications, the good range resolution is achieved by pulse compression and the cross-range resolution is obtained using synthetic aperture processing. It is useful to plot the radar response to a point scatterer on the range-azimuth plane as the various processes which determine the overall resolution can be illustrated graphically.

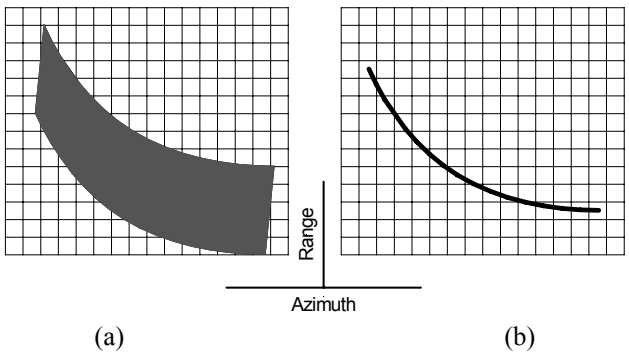


Figure 12.37: Range-azimuth response to a point scatterer shows range walk and range curvature (a) before pulse compression and (b) after pulse compression

12.12.1. Unfocussed SAR

In unfocussed SAR, all aircraft motion that deviates from a straight line is compensated for. In addition, for a forward squinted radar, the reduction in range induced by the radar velocity, called range-walk, is also removed prior to processing.

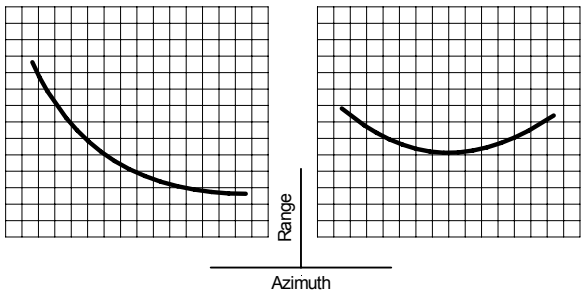


Figure 12.38: Effect of range-walk removal on a range-compressed point scatterer signature shows (a) signature before range walk removal and (b) signature after range walk correction

However, unfocussed SAR makes no attempt to correct for the relative phase shift to a point target across the aperture which is known as range curvature. The limiting condition is, therefore, the point where the phase error reaches λ/4 as determined in the following construction

$$\frac{\lambda}{4} = \frac{L_{\max}}{2}\sin(\theta/2) \tag{12.67}$$

for $\sin(\theta/2) \approx \frac{L_{\max}}{2R}$

$$\frac{\lambda}{4} \approx \frac{{L_{\max}}^2}{4R} \tag{12.68}$$

$$L_{\max} = \sqrt{R\lambda} \tag{12.69}$$

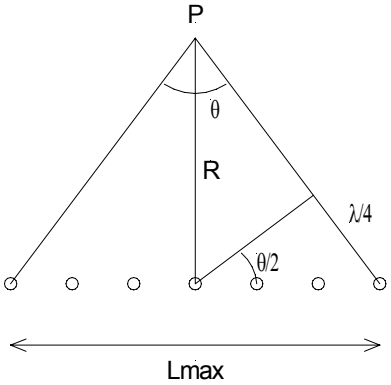


Figure 12.39: Largest possible synthetic aperture

A second limiting condition is that the beamwidth is sufficiently wide to illuminate the target at point P. $L_{\max} \le R\theta_{3dB}$

The beamwidth is obtained by equating $G_{SAR}(\theta) = 0.5$ as before

$$\frac{\pi L_e}{\lambda}\sin\theta \approx \frac{0.886\pi}{2} = 1.39 \tag{12.70}$$

$$\sin\theta = \frac{0.886\pi\lambda}{2\pi L_e} \tag{12.71}$$

The cross-range resolution $\delta_{cr} = R\theta = R.\sin\theta$ for small angles

$$\delta_{cr} = \frac{R0.886\pi\lambda}{2\pi L_e} \tag{12.72}$$

Substituting $L_e = L_{\max} = \sqrt{R\lambda}$ and simplifying

$$\delta_{cr} \approx 0.5\sqrt{R\lambda} \tag{12.73}$$

12.12.2. Focussed SAR

In the focussed SAR case both range-walk and range curvature are removed so that all of the returns from a point scatterer are concentrated at a single range. This allows the signal to be correlated across the azimuth plane to produce the required cross-range resolution.

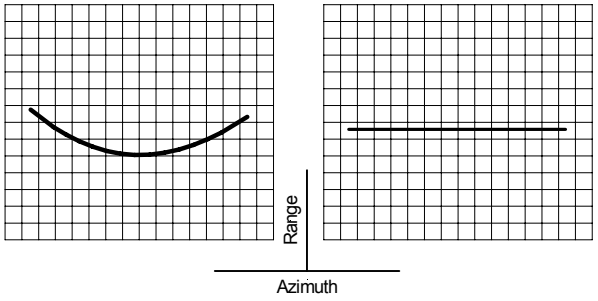


Figure 12.40: Effect of range-curvature removal on a range-compressed point scatterer signature shows (a) signature before range curvature removal and (b) signature after range curvature correction

This cross-range resolution can be determined from a Doppler perspective by considering the following:

A point scatterer enters the forward edge of the beam as shown in the figure below. It will have Doppler frequency

$$f_d = \frac{2v_r}{\lambda} = \frac{2v}{\lambda}Cos(\theta_{3dB}/2)\,. \tag{12.74}$$

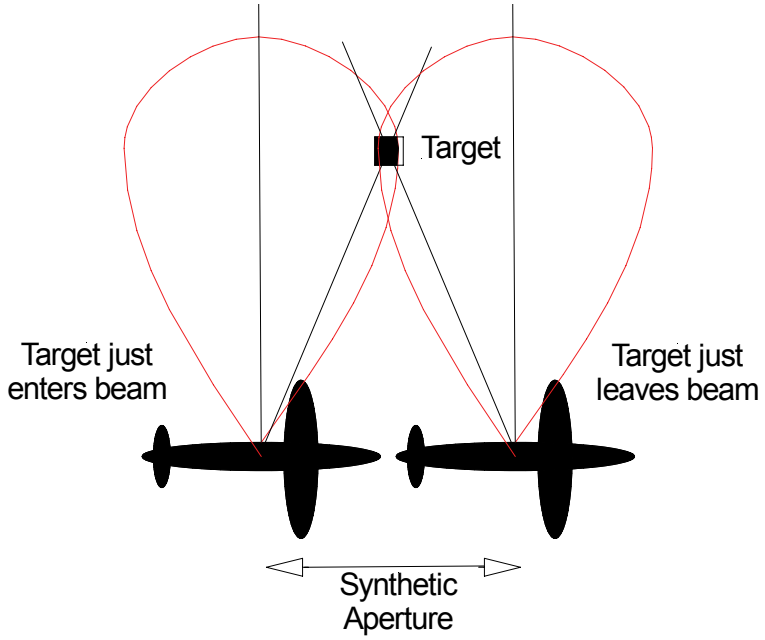


Figure 12.41: Antenna beamwidth limitations to maximum aperture

For small beamwidths, the Doppler frequency decreases linearly to 0 and then increases again and the angle to the target $\theta \approx \frac{vt}{R}$ as a function of time

The Doppler frequency as a function of time will then be

$$f_d(t) = \frac{2v_r}{\lambda} = \frac{2v}{\lambda}\cos\left[\frac{vt}{R}\right]. \tag{12.75}$$

Taking the derivative to obtain the rate of change of Doppler frequency, or the Doppler slope

$$\frac{df_d}{dt} = \frac{2v}{\lambda}.\frac{v}{R}\sin\left[\frac{vt}{R}\right], \tag{12.76}$$

equating at $t = 0$,

$$\frac{df_d}{dt} = \frac{2v^2}{R\lambda}. \tag{12.77}$$

The total Doppler shift over time T_d = time within the beam for $\theta = -\theta_{3dB}/2$ to $+\theta_{3dB}/2$, assuming a linear change in frequency

$$\Delta f_d = \frac{2v^2}{R\lambda}.T_d. \tag{12.78}$$

By analogy to the linear FM range resolution, the signal can be passed through a matched filter to give a spectral resolution $\delta f = 1/T_d$.

The cross range resolution is then the optimised cross range resolution of the real beam $\delta_b = L_e$ scaled by the ratio of the spectral resolution to the whole Doppler shift

$$\delta_{cr} = \delta_b . \frac{\delta f}{\Delta f_d} = L_e \frac{\delta f}{\Delta f_d} . \tag{12.79}$$

Substituting (12.78) into (12.79)

$$\delta_{cr} = L_e \frac{R\lambda}{2v^2 T_d} \frac{1}{T_d} = L_e \frac{R\lambda}{2v^2 T_d{}^2} . \tag{12.80}$$

But $L_e = v.T_d = R.\theta_b = \frac{R\lambda}{D}$ where D = antenna aperture, so the cross range resolution becomes

$$\delta_{cr} = \frac{R\lambda}{2L_e} = \frac{R\lambda}{2} . \frac{D}{R\lambda} = \frac{D}{2} . \tag{12.81}$$

The cross range resolution for focussed SAR is independent of the range R

12.12.3. Resolution Comparison

For a radar with a real aperture of 150mm and a frequency of 77GHz the following figure shows the relationship between the cross-range resolution of a real aperture antenna, unfocussed SAR and focussed SAR.

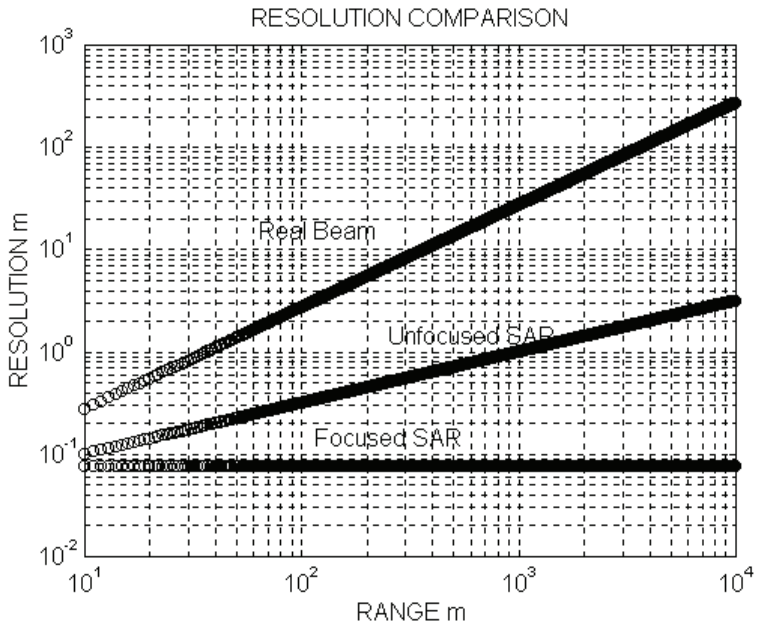


Figure 12.42: Resolution comparison: real aperture, unfocussed SAR and focussed SAR

12.13. Example: Synthetic Aperture Sonar

Three targets are placed with cross range separations of 20mm and 35mm at a range of 2m. Coordinates [20,2000], [40,2000], [75,2000] mm

The antenna aperture d = 10mm produces a 3dB beamwidth of 49°

As the sonar moves past the targets the received phase of the echoes is determined and shown in the following figure.

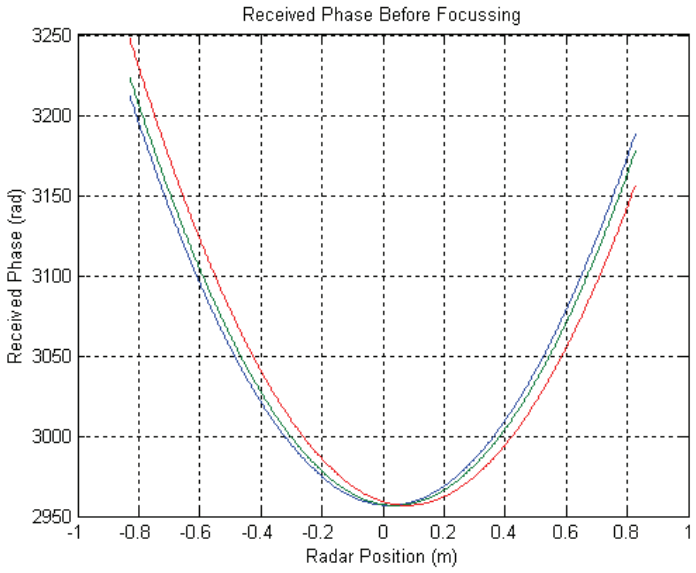


Figure 12.43: Received phase from three targets

The focussed returns are generated by subtracting a reference phase profile generated by a fixed target at [0,2000] making the phase shift with radar position is now almost linear.

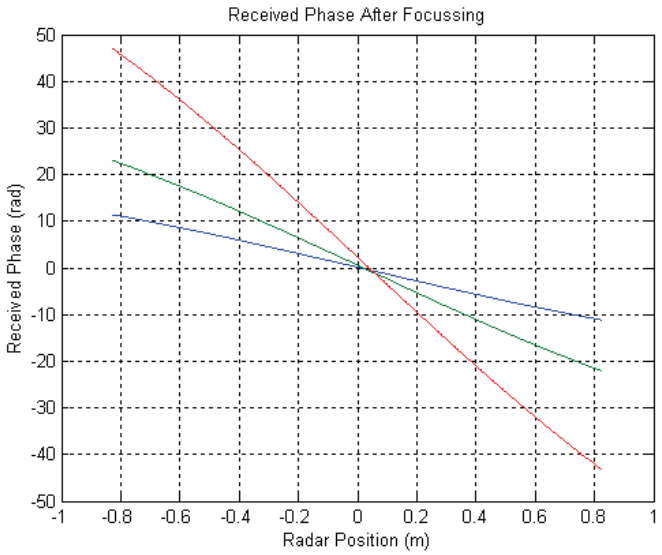


Figure 12.44: Received phase from three targets after focussing

The received signal is reconstructed by taking the cosine of the individual phases and summing the results

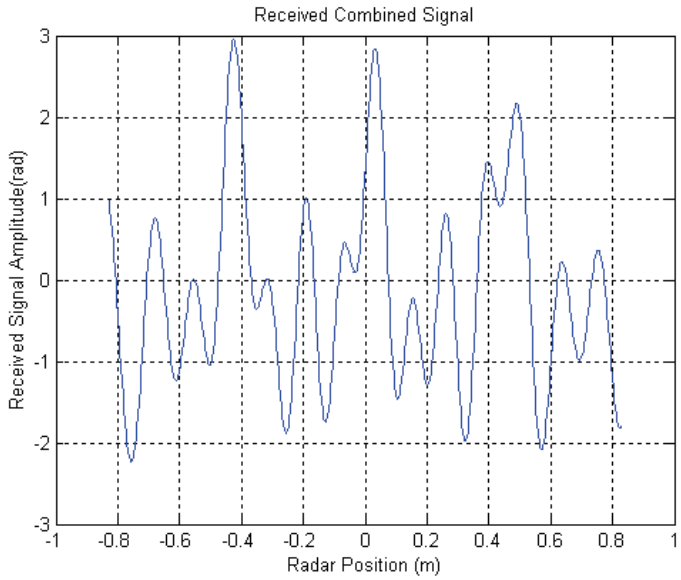


Figure 12.45: Received signal from the three targets

The individual targets are resolved by examining the spectrum of the received signal

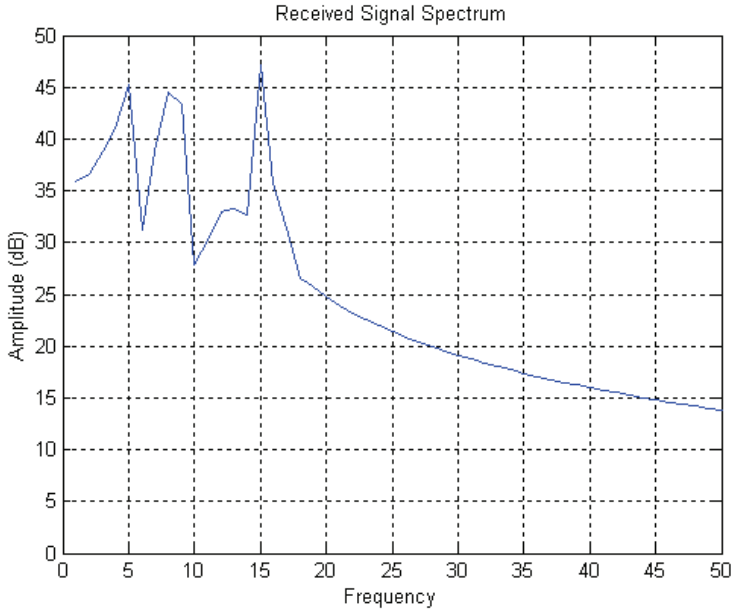


Figure 12.46: Resolved target cross range positions for d = 10mm

Increasing the antenna aperture to d = 30mm decreases the cross range resolution, and the two closest targets are no longer resolvable

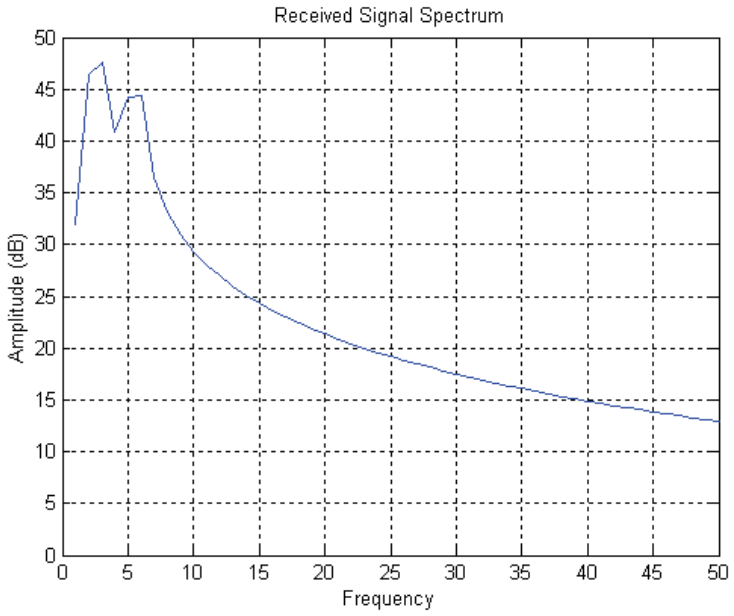


Figure 12.47: Resolved target cross range positions for *d* = 30mm

```
% Synthetic Aperture Sonar Model
% sar01.m

degrad = pi/180;          % degrees to radians
c = 340;                  % speed of sound (m/s)
f = 40e+03;               % operational frequency (Hz)
d = 10.0e-03;             % antenna aperture (m)
lam = c/f;                % wavelength (m)

% generate three targets at the same range separated by
% a fraction of the beamwidth
x0 = 0;
y0 = 2.0;                    % reference target to determine phase shift for
focus

x1 = 20.0e-03;
y1 = 2.0;
x2 = 40.0e-03;
y2 = 2.0;
x3 = 75.0e-03;
y3 = 2.0;

% assume a rectangular aperture and determine the 3dB beamwidth
thet3 = asin(0.887*lam/d);

% generate a moving radar system along the x axis
lmax=2*y1*sin(thet3/2);
dx = lmax/511;
xrad = -lmax/2:dx:lmax/2;
yrad = zeros(size(xrad));

% calculate the range to each of the targets
dx1 = xrad-x1;
dy1 = yrad-y1;
dx2 = xrad-x2;
dy2 = yrad-y2;
dx3 = xrad-x3;
dy3 = yrad-y3;

[az1,r1]=cart2pol(dx1,dy1);
[az2,r2]=cart2pol(dx2,dy2);
```

```
[az3,r3]=cart2pol(dx3,dy3);

% calculate the range to the reference target for focussing
dx0 = xrad-x0;
dy0 = yrad-y0;
[az0,r0]=cart2pol(dx0,dy0);

% calculate the phase due to the round trip distance to each target

ph1 = 4*pi*f.*r1./c;
ph2 = 4*pi*f.*r2./c;
ph3 = 4*pi*f.*r3./c;

ph0 = 4*pi*f.*r0./c;            % reference phase

plot(xrad,ph1,xrad,ph2,xrad,ph3)
grid
title('Received Phase Before Focussing')
xlabel('Radar Position (m)')
ylabel('Received Phase (rad)')
pause

% subtract the reference phase to focus

dph1 = ph1-ph0;
dph2 = ph2-ph0;
dph3 = ph3-ph0;
plot(xrad,dph1,xrad,dph2,xrad,dph3);
grid
title('Received Phase After Focussing')
xlabel('Radar Position (m)')
ylabel('Received Phase (rad)')
pause

% generate the time domain signal from the received phases
sig1 = cos(dph1);
sig2 = cos(dph2);
sig3 = cos(dph3);
sig = sig1+sig2+sig3;
plot(xrad,sig1,xrad,sig2,xrad,sig3);
grid
title('Received Signal')
xlabel('Radar Position (m)')
ylabel('Received Signal Amplitude(rad)')
pause

% extract the target returns by looking at the received spectrum
sigf = fft(sig);
db = 20*log10(abs(sigf(1:255)));
plot(db);
grid
title('Received Signal Spectrum')
xlabel('Frequency')
ylabel('Amplitude (dB)')
axis([0,50,0,50])
```

12.14. Radar Image Quality Issues

12.14.1. Perspective of a Radar Image

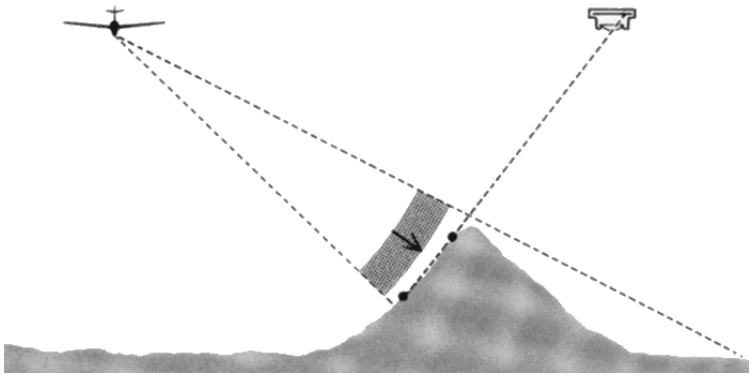


Figure 12.48: Perspective in a radar is different to that from an aerial photograph because in a radar two objects coincide if they are at the same range, whereas in an optical system, they appear to coincide if they re at the same angle

12.14.2. Image Distortion

The most common forms of distortion that are suffered by radar images made from the air include the following:

- layover, when the range to the top of an object is less than the distance to its base
- foreshortening, when the near side of elevated objects appears steeper than it actually is
- stretching due to the differences between the ground and slant ranges
- shadowing, when a tall opaque object blocks the signal path behind it, and no returns are received

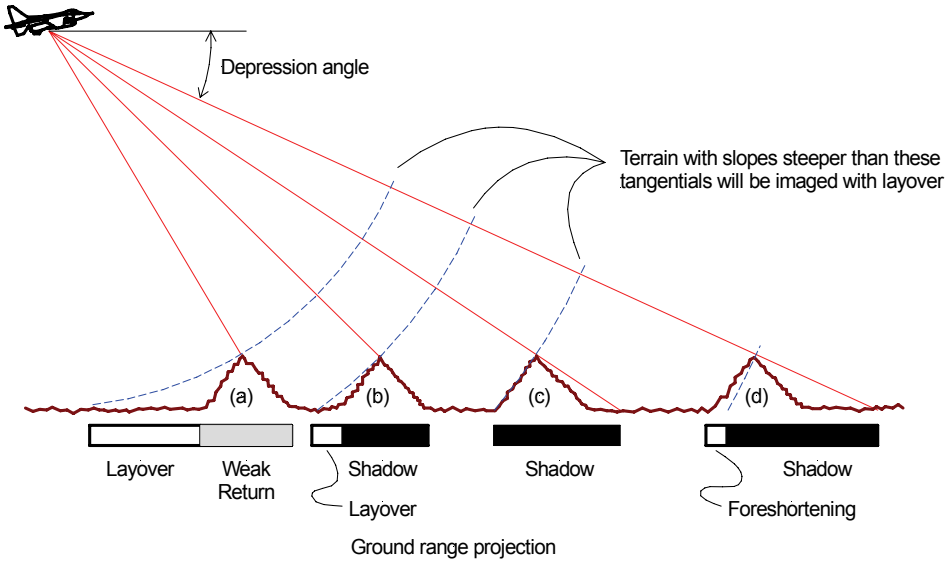


Figure 12.49: Some image distortion effects

Stretching

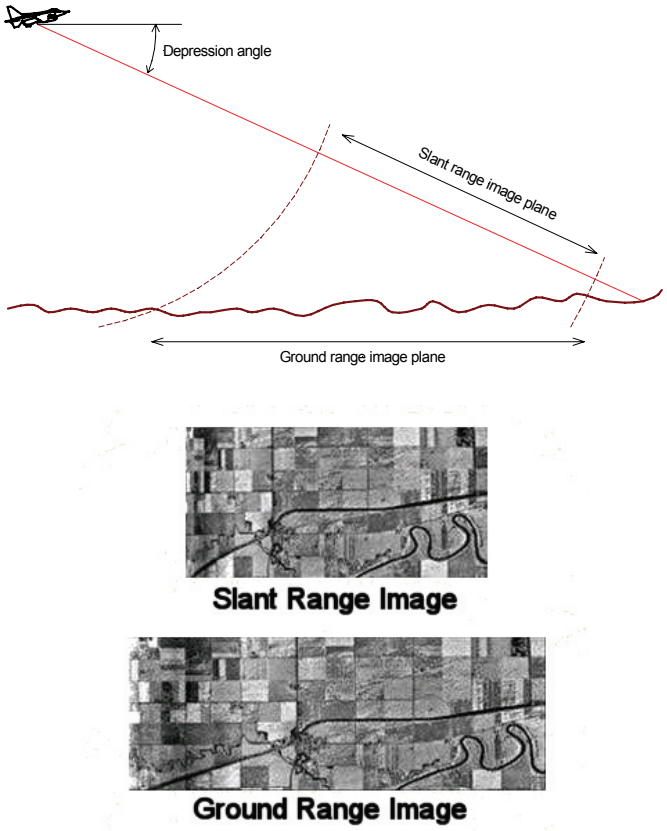


Figure 12.50: The radar image is distorted and compressed by the non-linear mapping between the ground range and the slant range

Shadowing

As illustrated in an earlier figure, shadowing occurs mostly at low grazing angles when an object, or objects block the signal path to cast shadows which cannot be imaged by the radar.

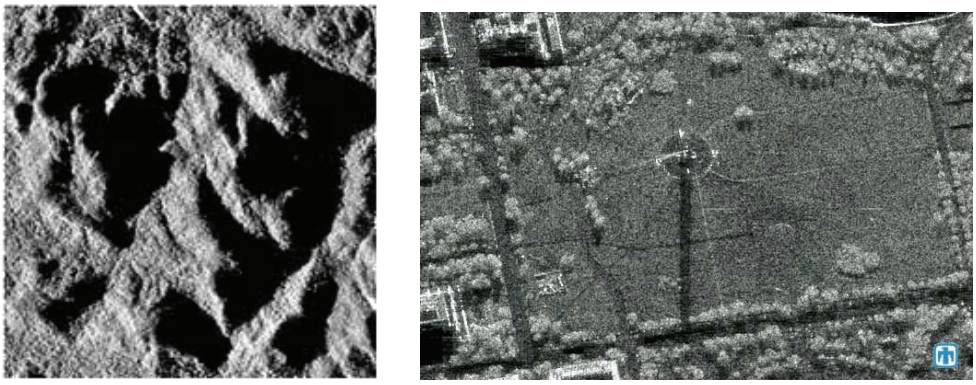
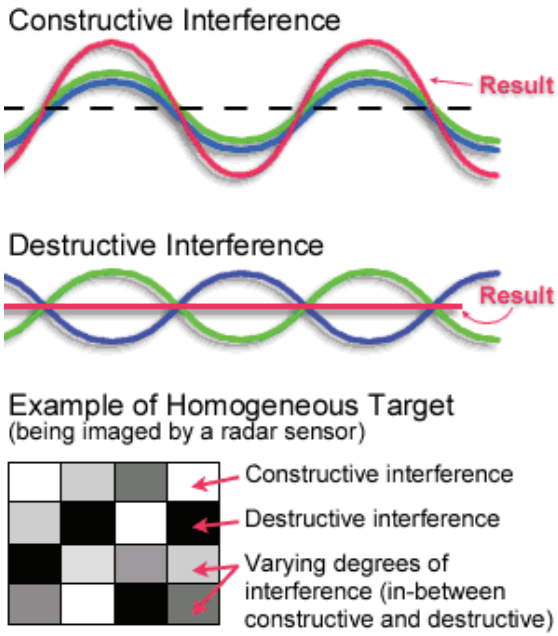


Figure 12.51: For low grazing angle applications, buildings and mountains cast shadows which are not imaged by the radar. These appear as black regions in SAR images

12.14.3. Speckle

Because the measurement process is coherent, in some of the image pixels, the scatterers will add constructively to produce high intensity returns, and in some the summation will be destructive to produce dark returns.

Various filtering and averaging techniques are used to reduce these effects and to make the images easier to interpret.



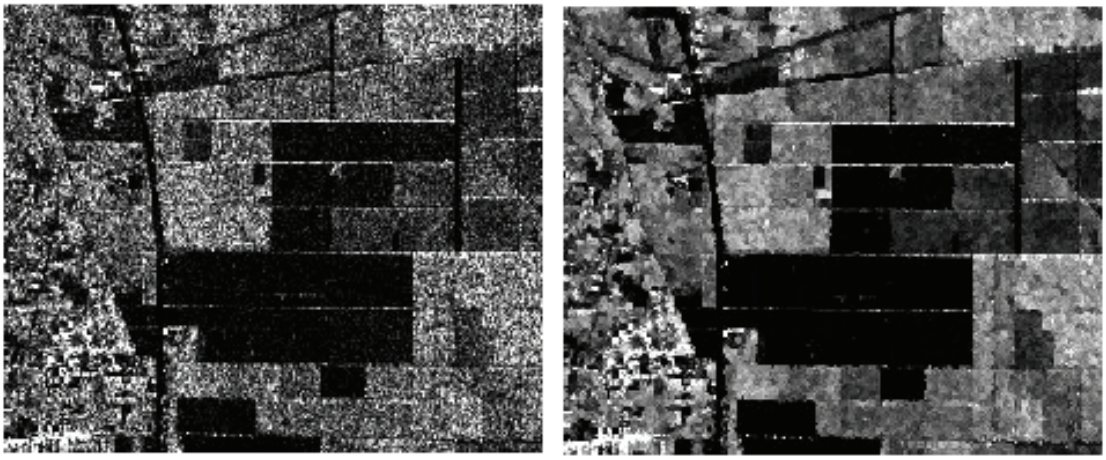


Figure 12.52: SAR Speckle and examples of some filter processes used to reduce its effect

12.15. Airborne SAR

Probably the best known of all the SAR systems designed for use in UAVs is the xxx installed in the Predator.

TESAR (Tactical Endurance Synthetic Aperture Radar) (also called the MAE UAV SAR) is a strip mapping SAR providing continuous 0.3 meter (1 foot) imagery. The focused imagery is formed on-board the Predator aircraft, compressed and sent to the

Predator Ground Control Station over a Ku band data link. The imagery is reformed and displayed in a scrolling manner on the SAR workstation displays. As the imagery is scrolling by, the operator has the ability to select 1k by 1k image patches (approximately 800m x 800m box at 15,000 AGL) for exploitation at the Predator DEMPC workstation. The imagery is also recorded continuously on DLT tapes.

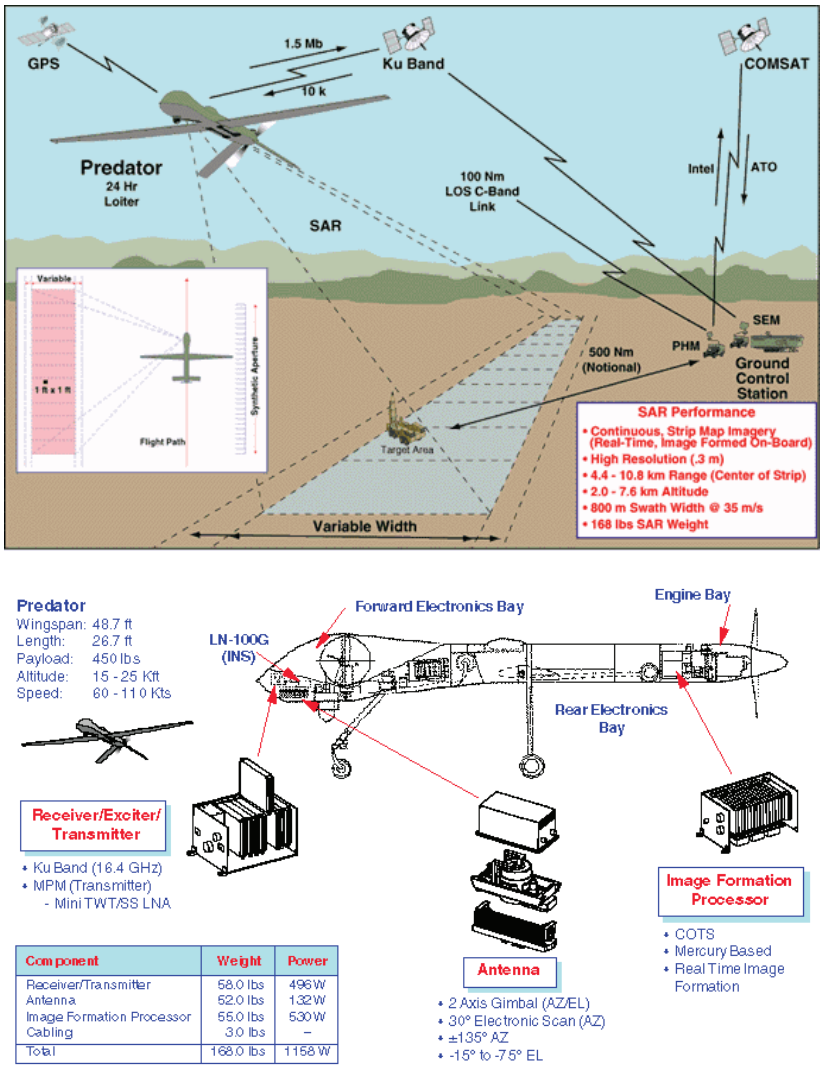


Figure 12.53: TESAR in the Predator UAV

There are 2 modes of operation. Mode 1 provides a non-centered strip map. That is the map center moves with respect to the aircraft motion. Mode 2 is the classic strip map mode. Mapping occurs over a predetermined scene center line, irrelevant of the aircraft motion.

The radar is designed to map while squinting up to ±45 degrees off the velocity vector. At ground speeds from 25-35 m/sec, the swath width is 800 meters. At speeds beyond 35 m/sec, the swath width decreases proportionally with the increase in ground speed

Other SAR developments for both UAV and manned aircraft include those by Dornier and Sandia.

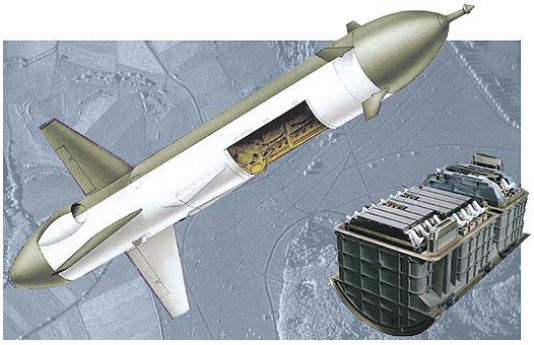


Figure 12.54: Dornier SAR for UAV applications

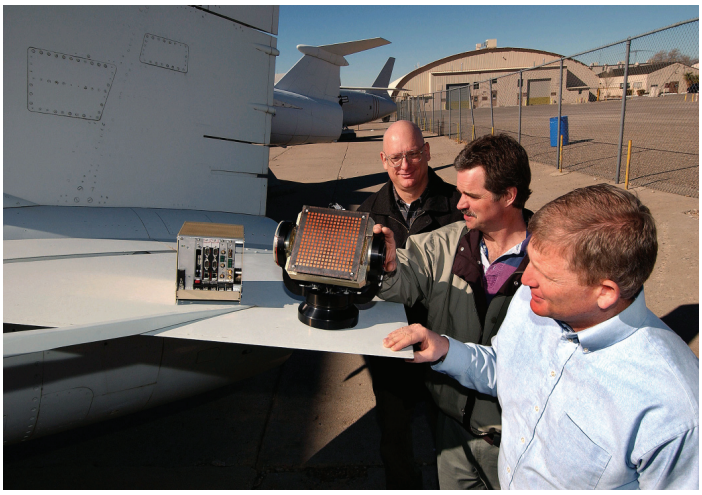


Figure 12.55: Photograph of the Sandia miniSAR developed for UAVs

By the end of 2006 Sandia National Laboratories was flying the smallest SAR ever to be used for reconnaissance on near-model-airplane-sized unmanned aerial vehicles (UAVs). Weighing less than 10kg, the miniSAR is one-fourth the weight and one-tenth the volume of its predecessors currently flying on larger UAVs such as the General Atomics' Predator. It is the latest design produced by Sandia based on more than 20 years of related research and development.

The new miniSAR will be able to take high-resolution (four-inch) images through bad weather, at night, and in dust storms out to a range of about 15km.



Figure 12.56: X-band SAR image of airport from Sandia labs

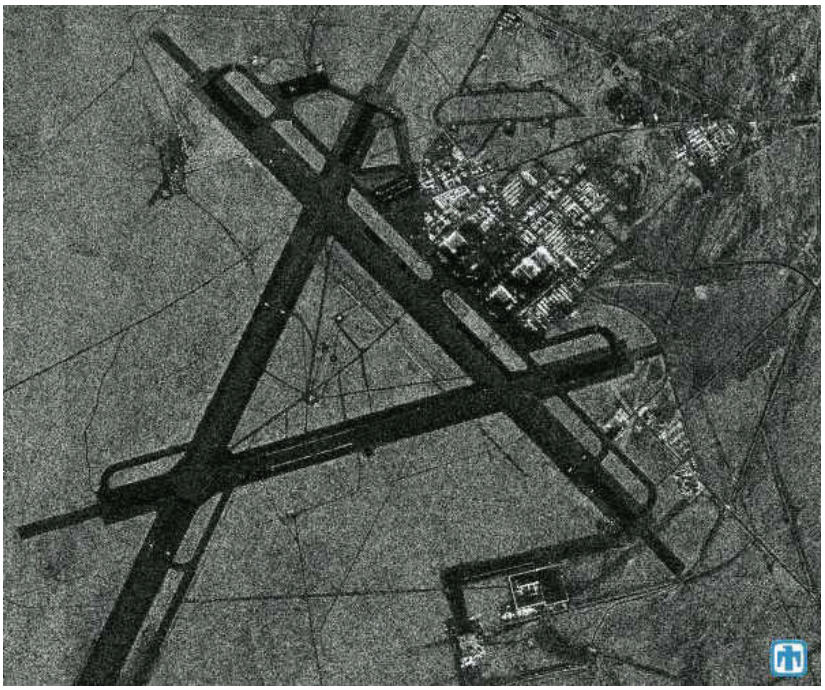


Figure 12.57: X-band SAR image of China Lake airfield (3m resolution)



Figure 12.58: SAR images of tanks and a ring of corner reflectors

12.16. Space Based SAR

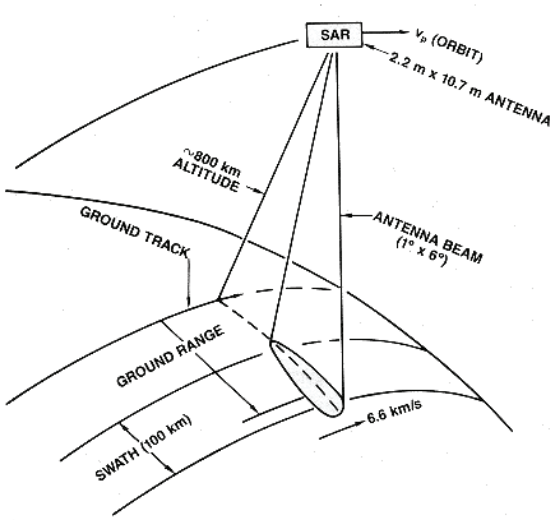


Figure 12.59: Space Based SAR System

To achieve good angular resolutions from real-aperture space-borne radars is impossible at lower frequencies because the size of the antenna becomes prohibitively large. As derived previously, the large synthetic aperture results in a cross range resolution independent of range, $\delta_{cr} = D/2$, where D is the antenna aperture.

The good range resolution with $\delta_r = c/2\Delta f$ is achieved by transmitting a wide bandwidth chirp.

The primary advantage of space-borne SAR is because the trajectory of the satellite or shuttle is precisely known and stable, motion compensation is not required and exceptionally high quality images can be produced. However, an additional distortion due to the curvature of the earth that must be compensated for is shown in the figure below.

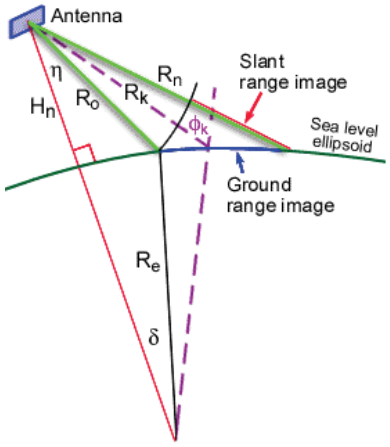


Figure 12.60: Geometric distortion due to the curvature of the earth

12.16.1. Interferometry

Because SAR is concerned with the phase relationships between scatterers on the ground. If two similar images are produced using offset antennas, or on subsequent passes over the same area, the interference patterns can be used to determine the true height of the objects on the ground.

In addition to being useful for mapping ground features, this technology has a number of other uses:

- Local deformation of the earth's crust as an early warning of earthquakes or volcanoes.
- Ground subsidence due to mining activities or excessive use of ground-water



This image of an area near White Sands, New Mexico, is the first to be returned from the Endeavour mission

Finally, earth in three dimensions

CAPE CANAVERAL: The Earth-mapping mission of the US shuttle Endeavour introduces a new use of technology never before employed in outer space: radar interferometry.

The use of radar to bounce signals off a foreign object and analyse them upon their return is nothing new.

But the Endeavour mission for the first time brings to outer space a technique called radar interferometry, which allows mapping in three dimensions through the simultaneous use of two radar receptors placed close to each other.

Endeavour's main antenna, installed in its payload bay, sends very high-frequency signals toward Earth. When they bounce back, however, they are received not only by the main antenna but also by a second one located at the end of a 60m-long mast.

This second antenna, located at a distance from the first, is the main new element of the Endeavour's radar topography mission.

It allows the production of a second, separate picture of the Earth, differing just slightly from that produced by the first antenna by virtue of its slightly different viewpoint.

The resulting "stereo" image is then converted by computer into three-dimensional form, much the way our own two eyes, spaced slightly apart, allow the human brain the capacity for depth perception.

Radar has two advantages over traditional optical instruments: it can collect information at night and it is not disrupted by clouds, which usually cover 40 per cent of the Earth's surface.

Figure 12.61: A Mission to Map the Earth in 3D using Interferometric SAR

12.16.2. SAR Image Gallery

C-band and X-band SAR systems have been flown by the shuttle over the last couple of years, and have produced the following incredible images.

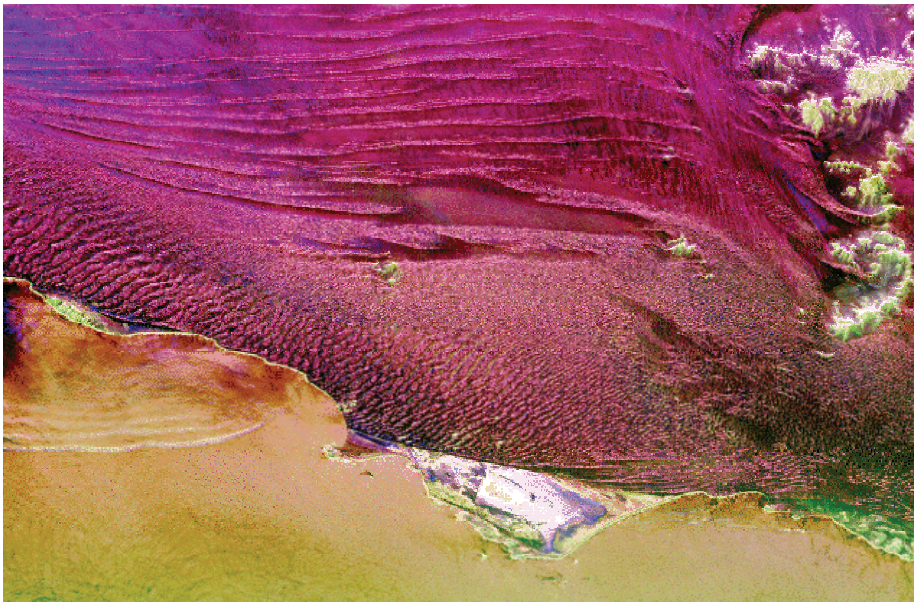


Figure 12.62: The Namibian sand dunes (54×82km).In ultra arid regions such as this, the radar signal can penetrate the sand and produce images of sub-surface features such as abandoned stream courses.

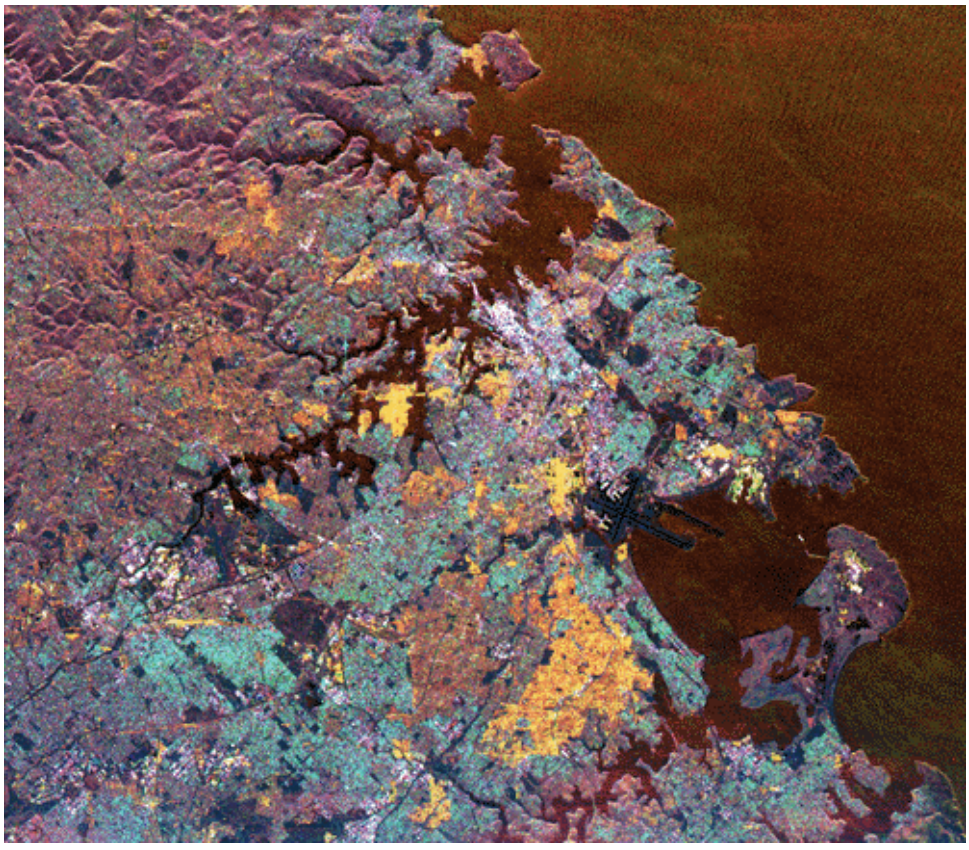


Figure 12.63: Sydney

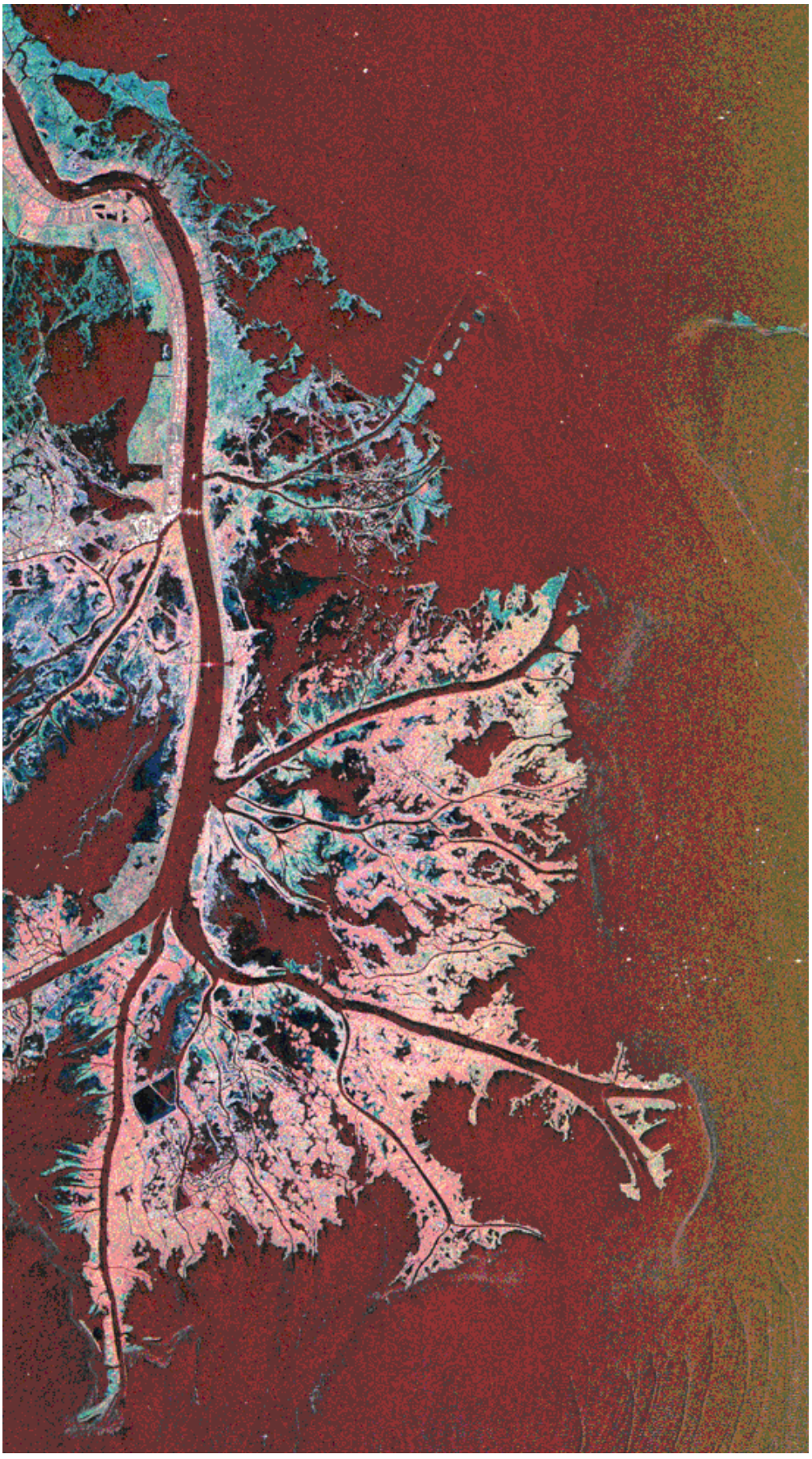


Figure 12.64: The Mississippi Delta (63×43km).The bright spots out at sea are oil rigs.

12.17. Magellan SAR Map of Venus

The radar sensor, Magellan's sole scientific instrument, produced high resolution SAR images of more than 80% of the planets surface. It was also used as a radar altimeter and the high gain antenna was used to communicate with the earth.

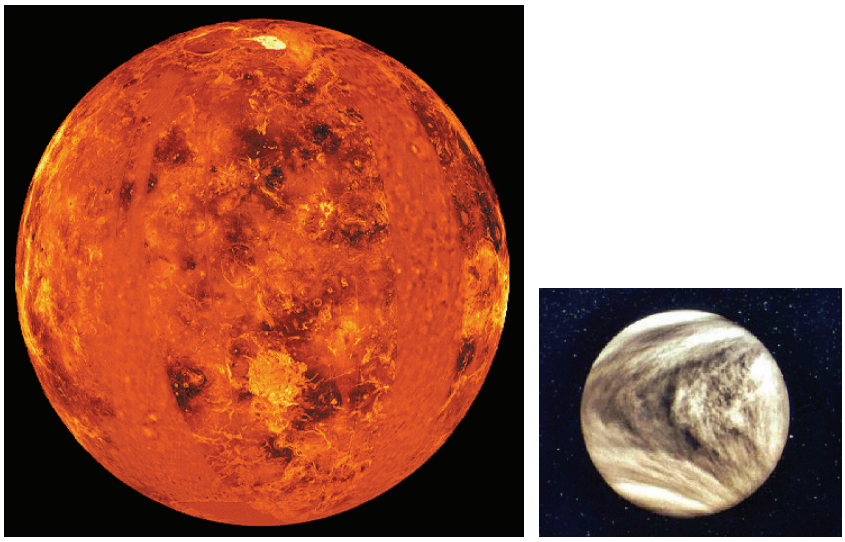


Figure 12.65: Magellan SAR map of Venus and visible image (inset)

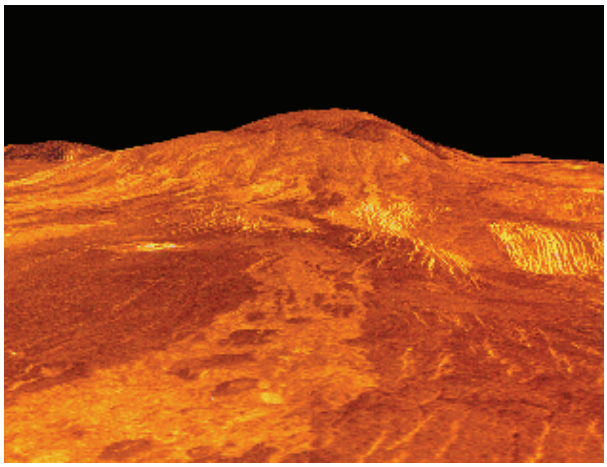


Figure 12.66: Sif Mons 2km high and 300km in diameter, 3D image produced by combining SAR and altimetry data

12.18. References

[1] M.Skolnik, *Introduction to Radar Systems,* McGraw-Hill, 1980

[2] E.Brookner, Phased-Array Radars, *Scientific American,* february, 1985

[3] E.Knutstal and W.Bunker, Guidelines for Specifying Underwater Electroacoustic Transducers, DT-92 Conference, London, June 1992.

[4] ITC Sidescan transducer, http://www.itc-transducers.com, 01/04/2001.

[5] C. Johnson, Sidescan Sonar- It's Not Just for Professionals, Treasure Diving, http:// onlinether. com/dive.html, 31/03/2001.

[6] Sonar Systems, http://www.eee.hw.ac.uk/oceans/projects/sonar/genson/, 31/03/2001.

[7] Acustics and Sonar, Marine Group, http://www.marine-group.com/acoustic.html, 31/03/2001.

[8] C. Danworth, Acoustic Applications of Phased Array Technology, http://fuse.pha.jhu.edu/ ~danworth/ sonar/sonar1.html, 31/03/2001.

[9] Kleinsonar, http://www.kleinsonar.com, 23/02/2000.
[10] EdgeTech, http://edgetech.com, 23/02/2000.
[11] GeoAcoustics Ltd, http://geoacoustics.co.uk, 23/02/2000.
[12] C-Max, http://www.c-max.co.uk, 23/02/2000.
[13] Scientific American
[14] http://www.jpl.nasa.gov/magellan
[15] http://www.jpl.nasa.gov/radar/sircxsar
[16] http://southport.jpl.nasa/gov/html/sirc_desc.html
[17] "X-band Synthetic Aperture Imagery, Sandia Labs," http://www.sandia.gov/RADAR/imageryx.html, accessed 26/11/04

Chapter 13.

Range and Angle Tracking

13.1.Introduction

Estimation of the position of a moving target is generally referred to as “tracking”. For most time-of-flight sensors that operate in polar space, this process involves following the target independently in both range and angles to obtain good estimates of its position in 3D. This chapter is concerned with the mechanics of this tracking process.

13.2.Range Estimation and Tracking

Range Tracking is an extension to the independent measurement of range on a pulse by pulse or block (integrated) by block basis. It involves using the previous estimate of the range and range rate to predict the where the target will be in time for the next range measurement. The most common configuration uses a type-2 loop (2 cascaded integrators or their equivalent) to give zero lag for constant range-rate targets.

In the radar context, standard estimation techniques were introduced first to automate the process of plot extraction for individual aircraft targets observed by surveillance radars. This is a process called “track while scan” and is discussed in Chapter 15.

13.2.1. Range gating

Range gating process consists of sampling the received video signal at a specified time after the transmit pulse has been radiated. This involves measuring the amplitude of this signal over a short period using some form of electronic switch that charges a capacitor, or a sample-and-hold (S&H) IC which performs a similar function

The sample period should be about equal to the length of the transmitted pulse so that the maximum amount of pulse energy and the minimum amount of noise is incorporated into each sample.

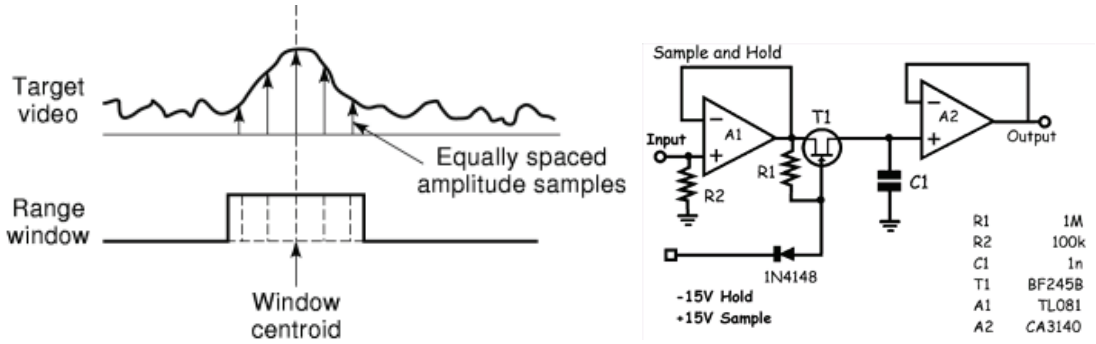


Figure 13.1: Sampling an echo pulse is known as range gating

13.2.2. Principles of a Split-Gate Tracker

A split gate tracker consists of two S&H circuits separated in time (range) by about one pulse width. They are known as the Early and Late gates. In early analog systems, these were in fact FET switches that allowed charge to flow into an integrator for the gate period τ_e and τ_l as shown in the figure below.

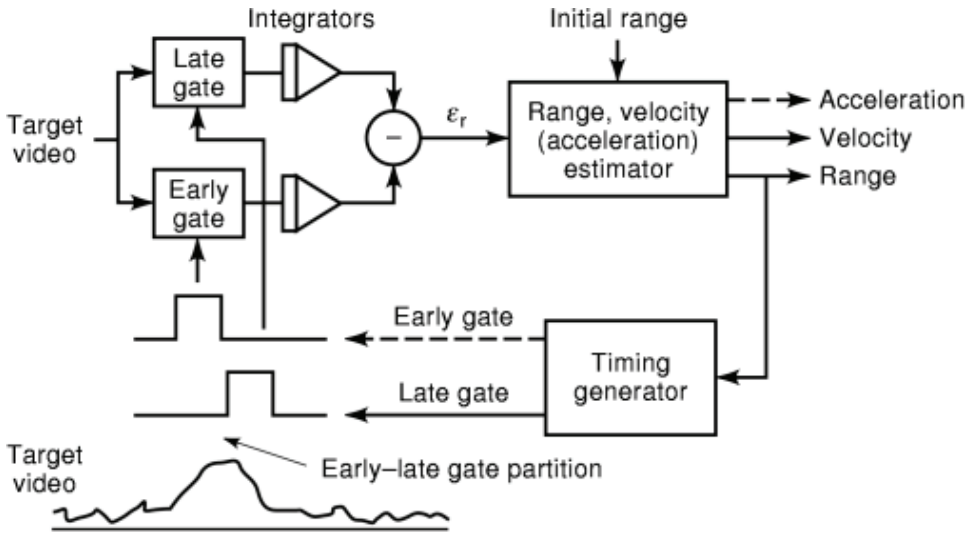


Figure 13.2: Analog split gate tracker

13.2.3. Range Transfer Function

At the time corresponding to the estimated target range, these S&Hs are triggered, one, one half pulsewidth prior to the estimated range delay, and one the same period after. If the range estimate is accurate, they sample equal amplitudes of the target echo pulse on either side of the peak as shown in the diagram, and the difference between the two gate voltages V_L-V_E = 0. However, if there is a small range error then the contents of the one gate will be larger than the contents of the other and the difference will not be zero.

To measure the transfer function, it is possible to keep the gates still and move the target through the gates, or visa versa. The following figure shows the measured range transfer function for an ultrasonic radar simulation

Note that the sign and magnitude of the error will drive the tracking error toward zero only over a limited range (in this case, about 20 samples) and the "linear" region over which the magnitude of the range error can be used to estimate the actual error is more limited (about 8 samples).

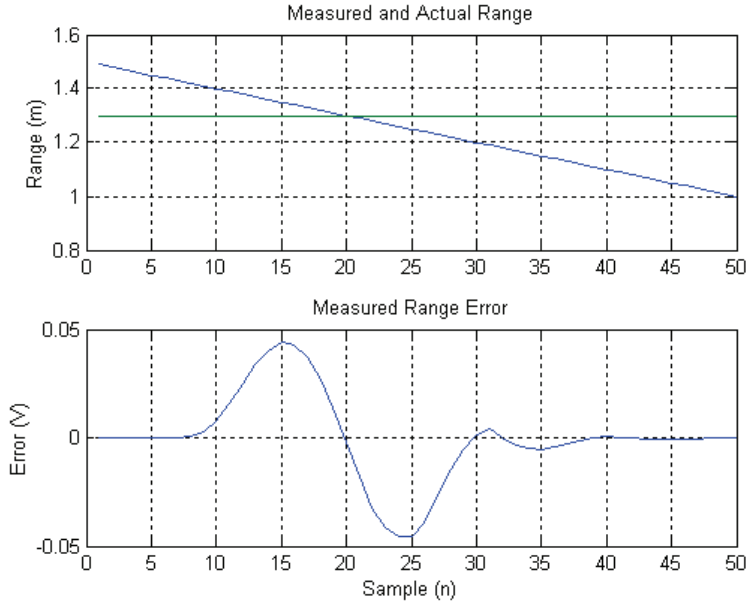


Figure 13.3: Measured transfer function for an ultrasonic range tracker

The range tracking loop uses this error to drive the smoothed estimate of the target range until the amplitudes of the early and late gate samples are equal. If, however, the error falls outside the bounds shown by the transfer function, then the error voltage falls to zero (in a noise free environment) and there is no driving function.

13.2.4. Noise on Split Gate Trackers

According to Barton, the best noise performance is achieved when the matched filter is matched to the pulse spectrum ($1<B.\tau<2$) and then the range gate width is matched to the pulse width ($\tau<\tau_g<2\tau$). This leads to an error slope, k_r = 2.5, and a normalised RMS tracking error given by

$$\sigma_r = \frac{\tau}{2.5\sqrt{2S/N}} \tag{13.1}$$

Note that the tracking error will take on the dimensions of τ, so it can be a range error in metres or a delay error in seconds.

13.2.5. Tracking and Estimation

The process applies simple estimation theory to predict where the target will be and by placing the gates at the correct range prior to the measurement taking place it improves the measurement accuracy. Placing the range gates as accurately as possible has two primary functions. In the first instance it ensures that any residual errors are within the linear region of the range transfer function and are close to the true error.

Secondly, by sampling the echo signal close to its peak ensures that the measurement SNR is as high as possible.

Typically a range tracker will include three gates, the early and late gates discussed earlier, and a sum gate which straddles the other two gates and samples the whole received pulse. This gate can be a physical realisation as shown in the figure below, or it can be the sum of the voltage signals from the early and late gates.

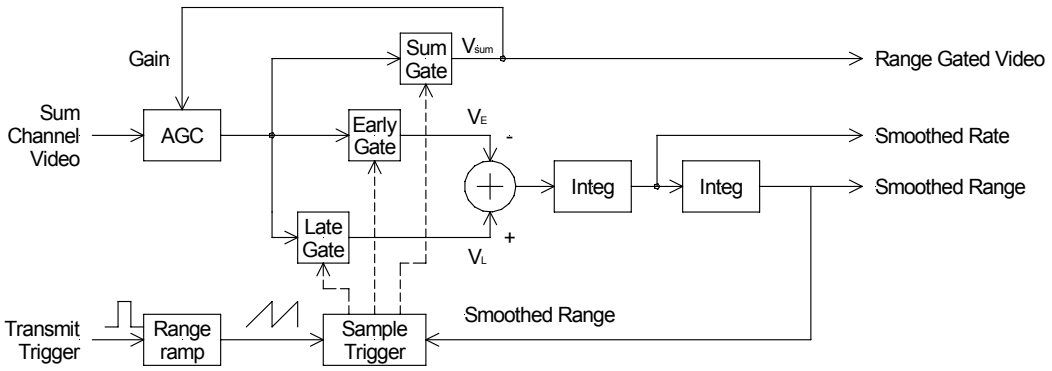


Figure 13.4: Analog tracking loop implementation using cascaded integraters

To maintain a constant loop bandwidth, the error function V_L-V_E must be normalised otherwise targets with a large RCS, or closer to the radar will produce larger errors than one with a small RCS or further away even if the true range error is the same. Normalisation is achieved using an automatic gain control (ACG) driven by the sum channel range gated video signal to maintain a constant target amplitude in the receiver irrespective of the range or the target RCS

The normalised range error drives a second order tracker, represented in the figure by a pair of cascaded integrators. In most modern trackers digital implementations of the tracking loop are used as they are easy to implement and can be optimised to suit specific conditions. Problems with pure digital implementations such as the one shown in the figure below is that the S&H and ADC must have sufficient dynamic range to digitise the full range of target amplitude variations. This can exceed 80dB.

The primary object of such tracking filters is to minimise the output noise variance under specific conditions of input. (usually constant velocity).

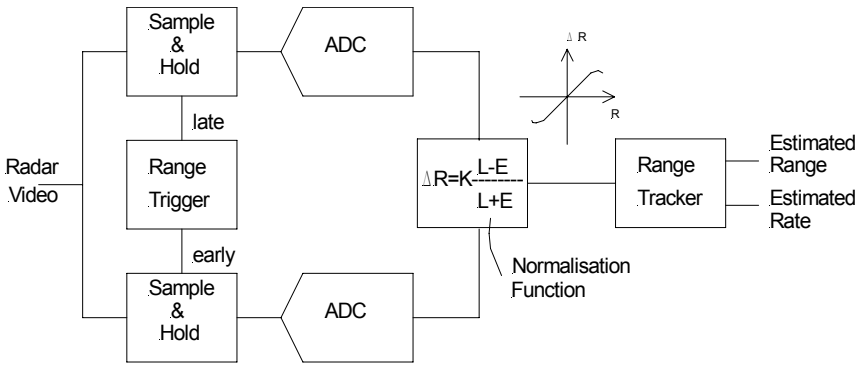


Figure 13.5: Digital tracking loop implementation

Kalman filters and α-β trackers are widely used for track-while-scan and single target tracking applications.

The α-β Filter

The α-β tracker is a fixed gain formulation of the Kalman filter and is still widely used because it is easy to implement performs well in general.

The α-β Tracker equations for range tracking are defined below

Smoothing

$$\hat{R}_n = \hat{R}_{pn} + \alpha\left(R_n - \hat{R}_{pn}\right) \tag{13.2}$$

$$\hat{V}_n = \hat{V}_{pn} + \frac{\beta}{T_s}\left(R_n - \hat{R}_{pn}\right) \tag{13.3}$$

Prediction

$$\hat{R}_{p(n+1)} = \hat{R}_n + \hat{V}_n.T_s \tag{13.4}$$

$$\hat{V}_{p(n+1)} = \hat{V}_n \tag{13.5}$$

where: $\hat{R}_n$ = Smoothed Estimate of Range

$\hat{V}_n$ = Smoothed estimate of Range Rate

R_n = Measured Range

$\hat{R}_{p(n+1)}$ = Predicted range after T seconds

$\hat{V}_{p(n+1)}$ = Predicted Range Rate after T seconds

$\hat{R}_{pn}$ = Predicted range at the Measurement Time

$\hat{V}_{pn}$ = Predicted Velocity at the Measurement Time

T_s = Sample Time

α, β = Smoothing Constants

It has been suggested (Benedict and Bordner) that to minimise the output noise variance at steady state, and the transient response to a manoeuvring target as modelled by a ramp function, then the gain coefficients are related as

$$\beta = \frac{\alpha^2}{2-\alpha}. \tag{13.6}$$

Other criteria can also be used. For example, it has been suggested that the filter should have the fastest possible step response (critical damping), in which case the coefficients are related as

$$\alpha = 2\sqrt{\beta} - \beta. \tag{13.7}$$

The actual values of the gains depend on the sample period, the predicted target dynamics and the required loop bandwidth.

The main disadvantage of the fixed gain α-β tracker is that it estimates the position of an accelerating target with a constant lag. The magnitude of this lag can easily be estimated and the filter can use adaptive gains to improve the RMS tracking accuracy under these circumstances. The cost function that is minimised is $[lag^2 + \sigma_r^2]$.

For estimates that include accelerations, the α-β-γ tracker can be substituted. However, its performance if the target is moving with constant velocity is poorer than that for the simple α-β tracker because of the additional noise introduced by the acceleration estimate.

The Kalman Filter

A simplified view of the Kalman filter when compared to one of the sub-optimal filters is that it uses optimum weighting coefficients (somewhat analogous to α and β) that are dynamically computed each update cycle. In addition a more precise model of the target dynamics can be used.

The benefits of the additional complexity include

- Improvement in tracking accuracy
- A running measure of the accuracy
- Method of handling measurements of variable accuracy, non uniform sample rate or missing samples
- Higher order systems are easier to handle

13.2.6. Ultrasonic Range Tracker Example

An ultrasonic range tracker based on the matched filter simulation described earlier in these notes has been developed to evaluate the effects of various gate and tracking filter parameters. In the following figure an approaching target is shown starting at a range of 1.5m and moving towards the sensor. The tracking gates are placed at a range of 1m.

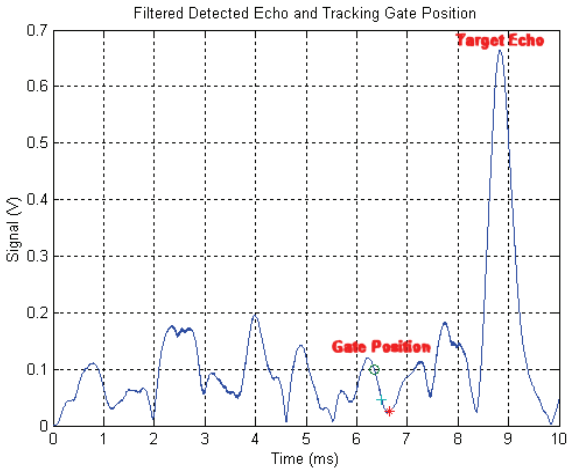


Figure 13.6: The target echo and the split gate position at the start of the simulation

To perform the split gate tracking function, three gates are shown, the early and the late gates straddling the sum (central) gate (o,+,*). Due to receiver noise, the tracking gates will move in a random fashion until they are reached by the target echo, at which time they will lock onto it.

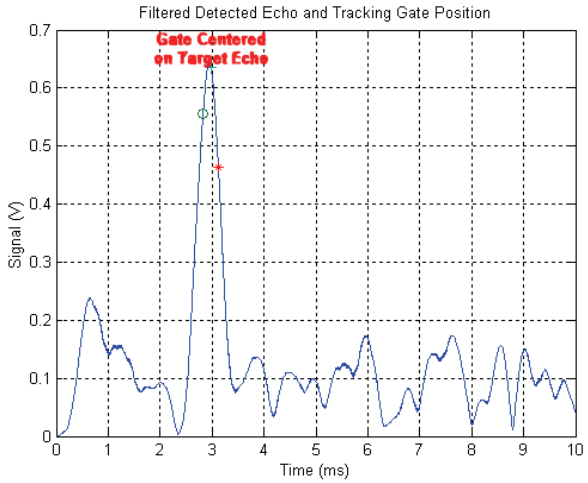
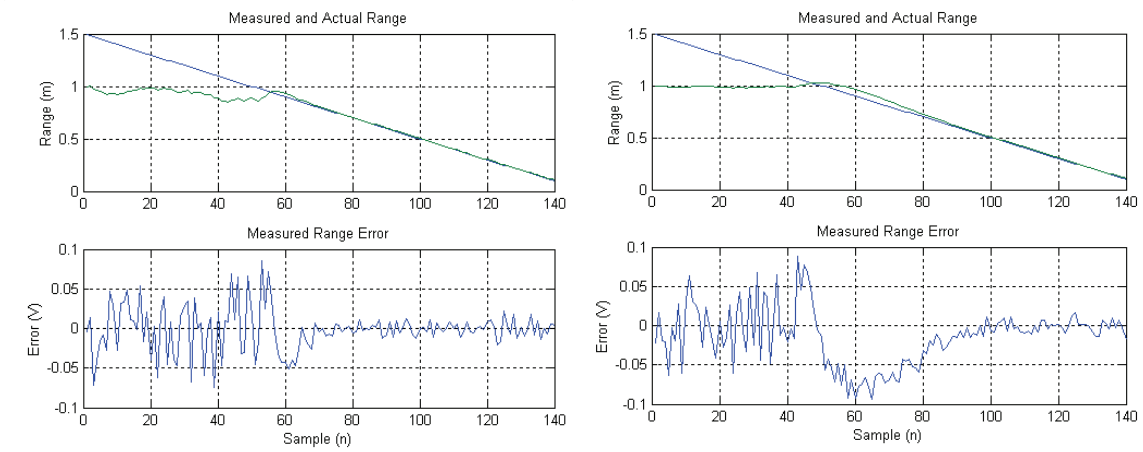


Figure 13.7: Target echo and split gate position during tracking

Note that the early and late gates are sitting astride the target echo peak which would produce a small tracking error voltage that drives the tracking loop to maintain its position straddling the target echo perfectly, even though the latter is moving.



(a) **(b)**

Figure 13.8: Target range, measured range and track error with time for (a) α = 0.3 (b) α = 0.1

13.2.7. Tracking Noise after Filtering

An α-β filter (with Benedict-Bordner gains) takes the error signal produced by the split gate tracker and produces estimates of the position and the rate. The one-sample-ahead position estimate is fed back into the range gate timing circuitry to trigger the early and late gate samples.

Note that for large α, the measured noise has a larger effect on the gate position than the for small α but that the filter settles onto the target much more quickly once the

two coincide. In general this settling period can be reduced by injecting course estimates of the target rate into the track filter, or by starting out with a filter with a wider bandwidth and progressively reducing it as the tracking error decreases

The tracking noise, being the difference between the measured and the true target position (for constant vel target) is inversely proportional to the filter bandwidth.

The filter performance is often characterised by the ratio in the variances of the output and input noise. This is known as the Variance Reduction Ratio (VRR). Measurements made over n samples combine in the filter to provide an output whose RMS noise is reduced by $1/\sqrt{n}$ compared to a single pulse measurement.

In terms of the equivalent noise bandwidth β_n of the range filter,

$$n = \frac{f_r}{2\beta_n} = \frac{1}{VRR}, \tag{13.8}$$

where VRR – Variance reduction ratio,
f_r – Pulse repetition frequency (Hz),
β_n - Bandwidth of tracking filter (Hz).

For the split gate tracker, the filtered thermal noise output will therefore be

$$\sigma_r = \frac{\tau}{2.5\sqrt{(S/N)(f_r/\beta_n)}}. \tag{13.9}$$

The following figure shows the filter transfer function (for the position estimate) for various values of α and β (using the Benedict Bordner relationship).

The filter half power bandwidth is measured at |H(z)| = 0.707.

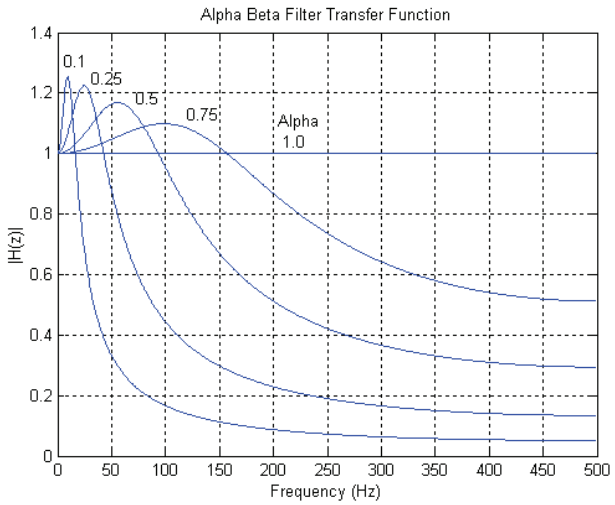


Figure 13.9: The α-β filter transfer function for a sample rate of 1kHz

It can be seen that the filter bandwidth increases as α increases, but that the relationship is not completely linear. As stated earlier in these notes, a filter can be

characterised completely by its impulse response. This allows the VRR to be determined as a ratio of the power in the impulse response to the power in the impulse. This is valid because the frequency spectrum of an impulse is flat whereas the response will be low-passed as shown in the figure.Alternative methods of calculating the VRR is to inject white noise of known variance into the filter and to measure the variance on the output, or by integrating the area under the $|H(z)|^2$ curve in the frequency domain.In general, the impulse response method converges very quickly and so has been used to determine the VRR plotted in the following figure.

Using the VRR formulation, the filtered tracking noise can be rewritten as

$$\sigma_r = \frac{\tau}{2.5\sqrt{2S/N}}\sqrt{VRR}\,. \tag{13.10}$$

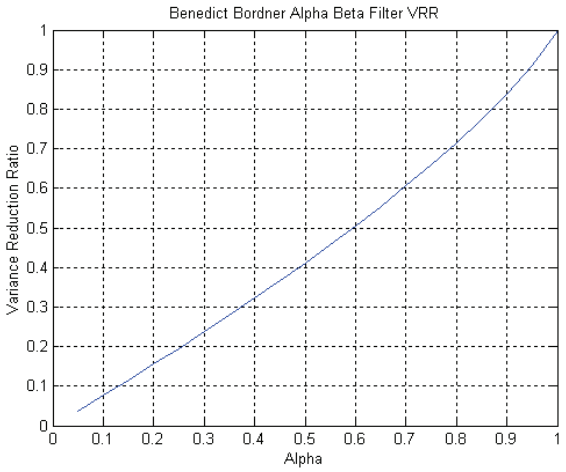


Figure 13.10: The α-β filter variance reduction ratio as a function of α

13.2.8. Tracking Lag for an Accelerating Target

The lag in the position estimate of an accelerating target is a function of the sample time T_s, the gains and the magnitude of the acceleration.

For the α-β filter, the lag as a function of the two gains is

$$L = \ddot{r}\frac{T_s^2}{\beta}(1-\alpha)\,, \tag{13.11}$$

where L – Lag (m),
$\ddot{r}$ - Range acceleration (m/s^2),
T_s – Sample interval (sec),
$\alpha\beta$ - Filter gains.

For a sample rate of 1kHz (T_s=1ms) and an acceleration of 1m/s^2, the following figure shows the lag as a function of α for a Benedict Bordner α-β filter. Because of the fast sample rate, this lag is insignificant unless the target acceleration is very high

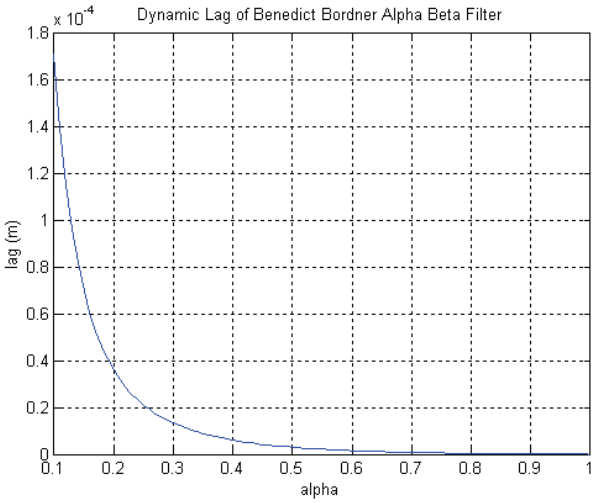


Figure 13.11: The α-β filter dynamic lag for a sample time of 1ms and an acceleration of 1m/s^2 as a function of α

13.2.9. Range Tracker Bandwidth Optimisation Example

As discussed earlier in this chapter, as the tracking filter bandwidth increases, the dynamic lag decreases. To minimise the mean squared tracking error the cost function that is minimised is [$lag^2 + \sigma_r^2$]. In this example, the α-β tracker must be optimised to track a missile with the following parameters:

- Target acceleration of 1g
- Split gate tracker with a gate size of 3m
- Signal to Noise Ratio assumed to be 10dB (thermal noise only)
- Sample rate 50Hz

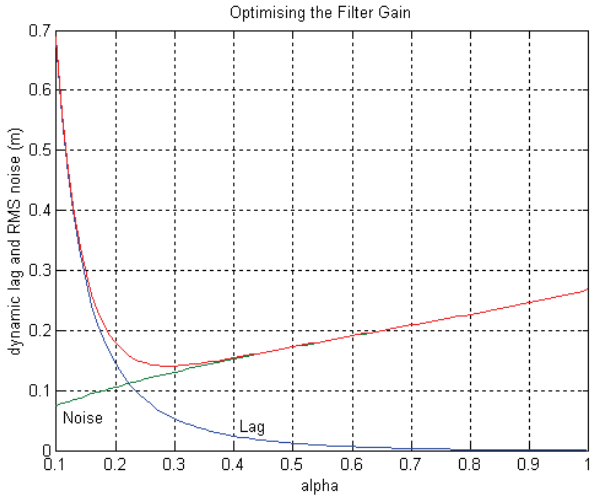


Figure 13.12: RMS noise and dynamic lag and the root mean squared sum determines the optimum gain

Using the equations developed to determine the RMS noise and the dynamic lag, use Matlab to calculate the root mean squared sum of the two parameters. In this case the optimum position estimation gain will be α = 0.29 from the graph. Using the Benedict Bordner relationship, the velocity estimation gain, β = 0.049.

For a 50Hz sample period, the filter bandwidth can be obtained from Figure 13.9 by interpolation and scaling to be about 3.5Hz

```
% determine the optimum tracking gains for an
% ablag.m
acc = 10;               % target acceleration
fs = 50;                % sample frequency
snrdb = 10;             % SNR
tau = 3;                % Range gate size(m)
snr = 10.^(snrdb/10);
ts = 1/fs;
% Calculate the RMS range tracking noise out of the split gate
% from the pulse width and the signal to noise ratio
sigma = tau/(2.5*sqrt(2*snr));
k=0
sigmaout=zeros(1,90);
alp=zeros(1,90);
lag=zeros(1,90);
for alpha=0.1:0.01:1,
    k=k+1;
    beta = alpha.^2/(2-alpha);
% position transfer function for the alpha beta filter
    b=([alpha, beta-alpha]);
    a=([1,beta+alpha-2,1-alpha]);
% measure the power in the impulse response to determine the VRR
    sig=[zeros(1,500),ones(1),zeros(1,500)];
    out=filter(b,a,sig);
    alp(k)=alpha;
    vrr=sum(out.^2);
% calculate the noise after the track filter
    sigmaout(k) = sigma*sqrt(vrr);
% calculate the dynamic lag from the formula
    lag(k) = acc*ts.^2*(1-alpha)/beta;
end
comb = sqrt(lag.^2+sigmaout.^2);
plot(alp,lag,alp,sigmaout,alp,comb);
grid
xlabel('alpha')
ylabel('dynamic lag and RMS noise (m)')
title('Optimising the Filter Gain')
```

13.3.Range Tracking Systems

Most range tracking systems are associated with angle trackers to produce a full 3D tracking capability as discussed in Chapter 15. However, there are a number of applications in which the angular pointing function is either non existent or performed manually. Good examples of the latter are combined optical and ranging systems like close-in weapon systems or even some radar or lidar speed traps.

13.3.1. Lidar Speed Trap

As discussed in Chapter 7, laser range finders provide the most cost-effective method to measure long range in benign environments because very narrow beamwidth can be achieved using low cost optics.

This allows for high angular resolution measurements to be made.

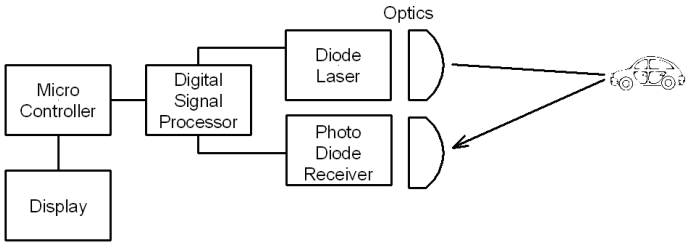


Figure 13.13: Laser radar operational principle

In the typical configuration shown in the figure above, the standard pulsed time of flight measurement technique is followed by a second micro-controller that processes the measured data to estimate the position and speed of the target.

Estimates the velocity can be made using a tracking filter of the kind discussed above, or by measuring the change in range as determined by the varying time of flight between successive pulses. From the figure below it can be seen that once the filter has settled, the estimates are smoother than those produced by the difference method. The amount of smoothing is determined by the value of the gains.

In practical applications, the filter is seeded with a reasonable estimate of the rate based on the position difference over a few samples. This speeds convergence considerably.

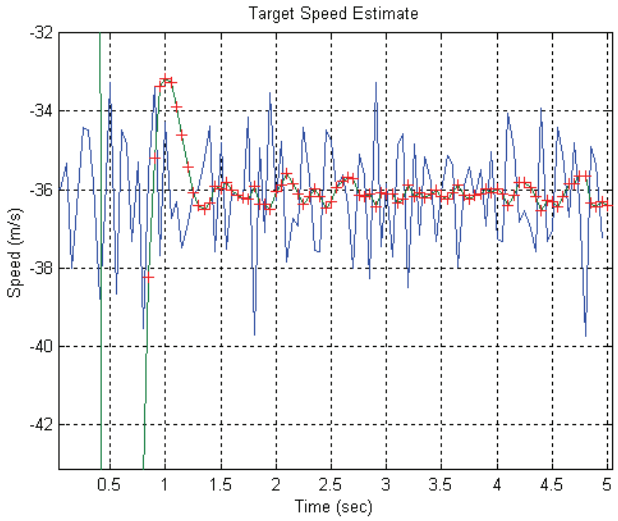


Figure 13.14: Tracker performance comparison between an αβ filter estimate of the speed and a simple difference shows the superiority of filtering

```
% laser radar range tracking simulation
% uniform distribution of range error +/-10cm
% sample rate 20Hz
xtar = 400;                    % start range of target (m)
vtar = -130.0;                 % speed km/h
v = vtar/3.6;                  % speed m/s
fs = 20.0;                     % sample rate (Hz)
ts = 1/fs;                     % sample time (sec)
% determine the actual position of the car with time
t=(0:ts:5);                    % allowed time to get a fix 5s
x=xtar+v*t;
% uniformly distributed range errors between -0.1 and +0.1m
xnoise = 0.2*(rand(size(t))-0.5);
xmeas=x+xnoise;
% measure rate by subtracting pairs of pulses
rate=diff(xmeas)./ts;
len=length(rate);
t1=t(1:len);
%plot(t1,rate);
% measure the rate by filtering using an anpha beta filter
alpha = 0.5;
beta = alpha.^2/(2-alpha);
b1=([alpha,beta-alpha]);       % position transfer function
a1=([1,beta+alpha-2,1-alpha]);
b2=beta*([1,-1,0])/ts;
a2=a1;
xdfil=filter(b2,a2,xmeas);     % rate transfer function
plot(t1,rate,t,xdfil,t,xdfil,'+')
grid
title('Target Speed Estimate');
xlabel('Time (sec)')
ylabel('Speed (m/s)')
```

Riegl FG21-P



- Operating principle pulsed time of flight
- Beamwidth 2.5mrad
- Measuring time 0.4 to 1s
- Speed 0-250km/h
- Accuracy +/-3km/h to 100km/h +/-3% of reading above 100km/h
- Range 30-1000m
- Accuracy +/-10cm typical
- Target marker, circular reticle matched to beam diameter

Figure 13.15: Riegl speed scope

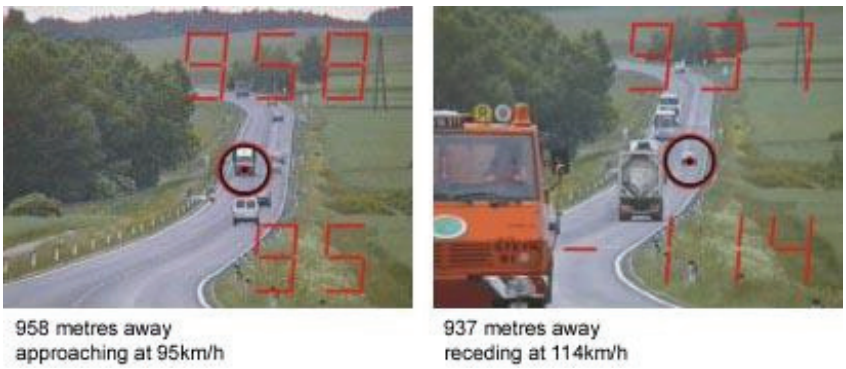


Figure 13.16: Effectiveness of narrow beam for measuring speed on congested roads

13.4.Angle Measurement

13.4.1. Amplitude Thresholding

At its most simple, a received echo amplitude threshold can be used to determine that a target is within the beam. This gives a very rough measure of the target direction.

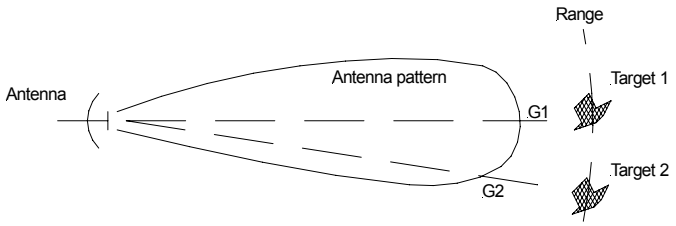


Figure 13.17: Antenna gain as a function of angle

Note that because the antenna gain drops off sharply, the angular position uncertainty is usually constrained to within the 10dB (one way) beamwidth of the antenna.

Both pulsed and CW techniques can be used for angle measurement over a broad range of frequencies from microwave to the infrared band. In general complex modulation schemes are generally used for the following reasons:

- Discriminate against ambient solar radiation,
- Eliminate interference from fluorescent lights,
- Reduce the probability of interference from other sensors.

One of the disadvantages of this technique is that the cross-range resolution degrades with range because sensors operate with a constant angular resolution, as shown in the figure below.

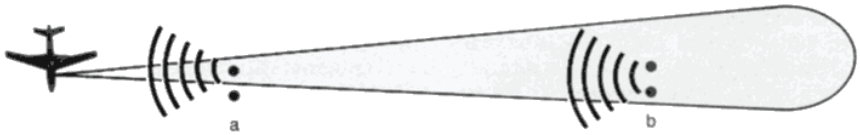


Figure 13.18: The cross range resolution degrades with range as shown. At position (a) the targets are resolved but at position (b) the targets, with the same spacing, are not resolved

The main applications are limited to CW or modulated IR proximity detectors for industrial & robotic applications.

Proximity Detector Example

IR proximity detectors are commonly used for short range robotic applications but not normally used to measure range, just the presence of a target within the beam. They operate in the near IR (at a wavelength just longer than visible light, typically 880nm) and are visible to CCD's so can be observed using a video camera (which can be useful)

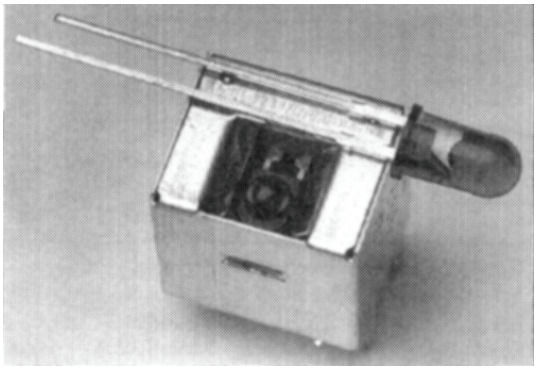


Figure 13.19: Infrared proximity detector

A typical receiver made by Sharp Electronics (GP1U52X) is shown here with a near infrared LED transmitter.

This receiver package includes a photodiode, amplifiers, filters and a limiter. The receiver responds to a burst modulated 40kHz signal with an on-off period of 600+600μs. The digital output goes low is a target is detected.

13.5. Angle Tracking Principles

13.5.1. Scanning Across the Target

Using the antenna beam pattern, it is possible to get a more accurate bearing on the target by sweeping the beam across it and noting changes in the signal amplitude. This process can increase the angular resolution with only a marginal effect on the range accuracy, however it relies on a good SNR for best results.

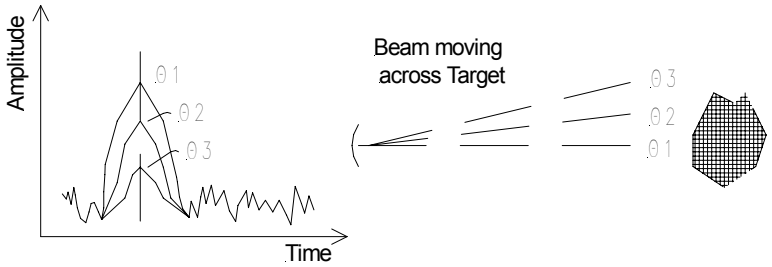


Figure 13.20: Using the beam pattern to estimate angle

13.5.2. Null Steering

Null steering involves the subtraction of the returns from a pair of overlapping beams to produce a "null' when the target is aligned with the beam axis of symmetry. Extremely accurate angle measurements can be achieved. For a point target, the theoretical improvement in angle measurement accuracy in thermal-noise is limited only by the signal to noise ratio of the measurement

$$\sigma_t = \frac{\theta_{3dB}}{k\sqrt{2(S/N)}} \text{ deg} \tag{13.12}$$

where: θ_{3dB} – Antenna Beamwidth (deg)
k – Constant dependant on the tracking type
S/N – Signal to noise ratio

The following are the most common null steering techniques:

- Lobe Switching
- Conical Scan
- Monopulse

13.6.Lobe Switching (Sequential Lobing)

This technique involves sequential transmission from two antennas with overlapping but offset beam patterns.

Sequential returns will be amplitude modulated if the target is not on boresight. This AM can be used to generate an angle error estimate, or used to drive the antenna to null the error.

Two additional switching positions are needed to obtain the angular error in the orthogonal axis, so 4 pulses are required to control the antenna in 2D.

This technique can be used for EM or acoustic tracking systems and most non scanning collision avoidance radars use this technique with either 2 or 3 lobes to determine the angular offset of cars. In addition a passive version of lobe switching is often used for direction finding applications as discussed in the example at the end of this chapter.

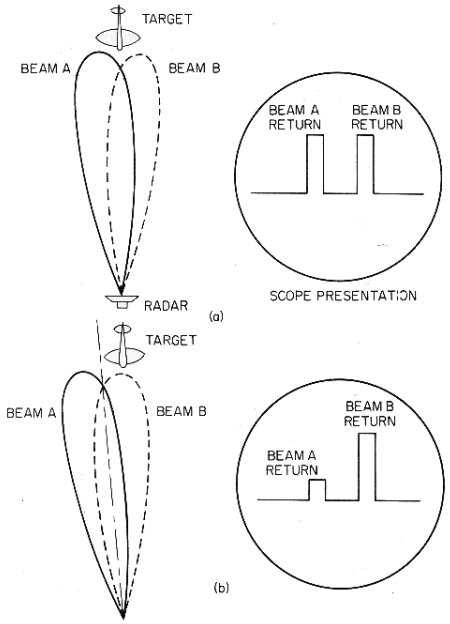


Figure 13.21: Principle of sequential lobbing to determine angle error

13.6.1. Main Disadvantages of Lobe Switching

This technique results in a reduced bandwidth because 4 pulses are required to resolve the target in 2D.

Fluctuations in the signal level due to variations in the echo strength on a pulse-by-pulse basis reduce the tracking accuracy so it is susceptible to modulation by the target, either natural (propellers, wing beats etc.) or as part of the electronic countermeasures.

The antenna gain in the on boresight direction is less than the peak gain, so the peak gain is reduced with a consequent reduction in the maximum operational range.

13.7.Conical Scan

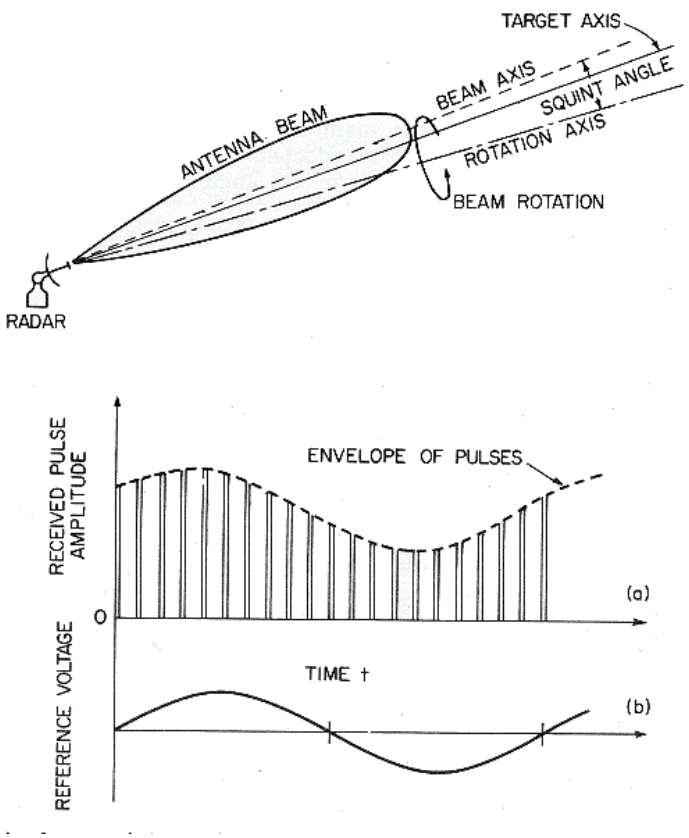


Figure 13.22: Principle of conical scanning to determine angle error

A single beam displaced in angle by less than the antenna beamwidth is nutated on its axis. (Nutated means spinning without rotating the polarisation). Beam displacement is often achieved by incorporating a rotating sub-reflector in a Cassegrain antenna as shown in the figure below.

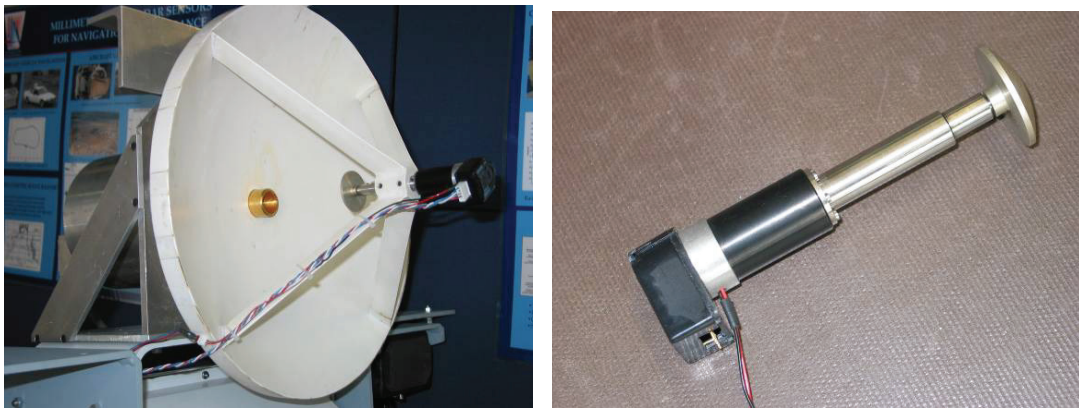


Figure 13.23: Cassegrain antenna with a canted spinning sub-reflector to implement conical scan

The scan rate is generally limited to between 5 and 25 pulses per revolution for long range operation, but at short range where the PRF is higher, many more pulses are

received. Amplitude modulation of target returns will be a function of the position of the boresight with respect to the target and phase detectors (multipliers) using quadrature phase references for the two orthogonal axes demodulate the received AM signal out of the range tracking gate to generate the angle tracking error as described in the following section.

There are a number of different methods that can be used to perform this demodulation function as shown in the following figure.

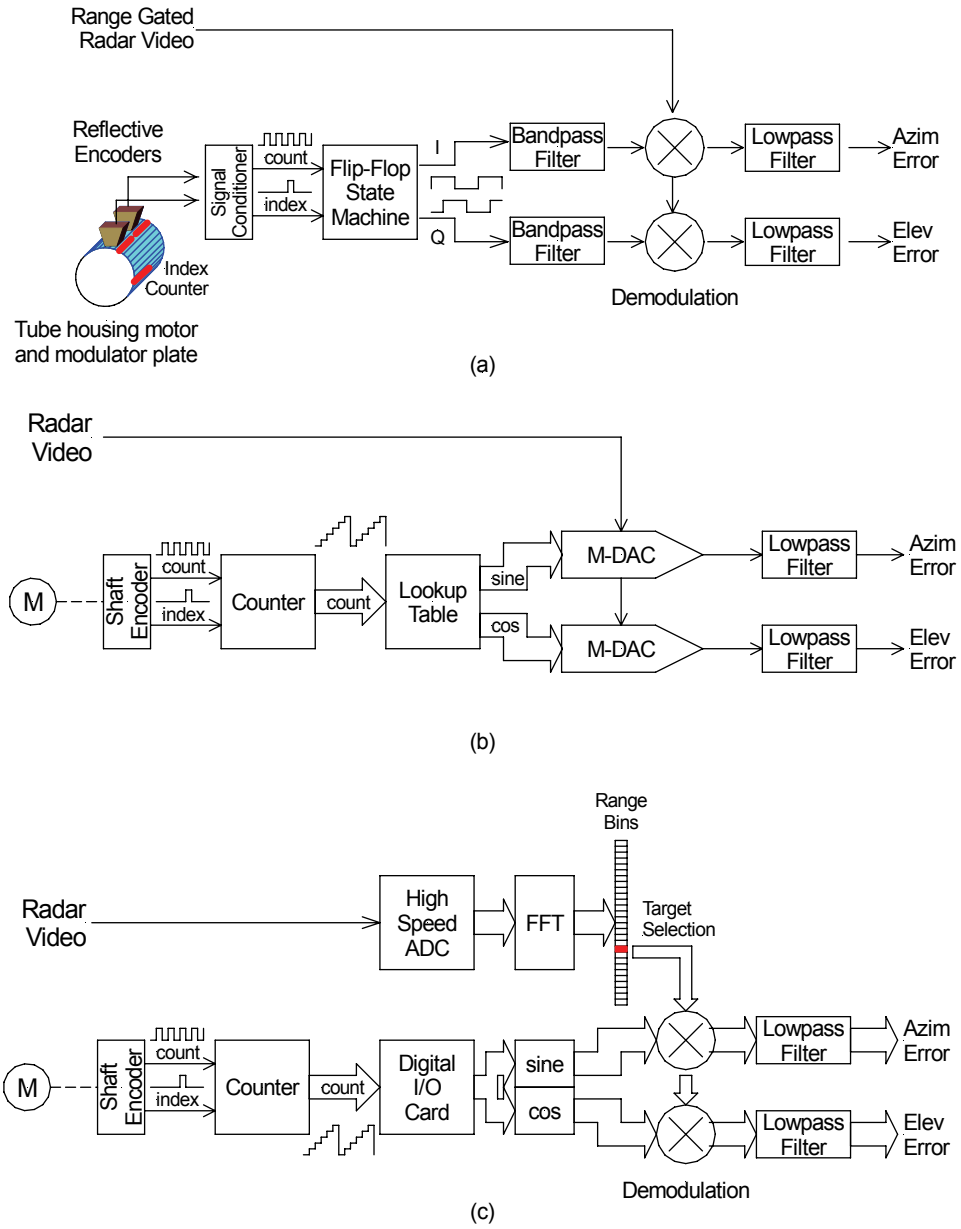
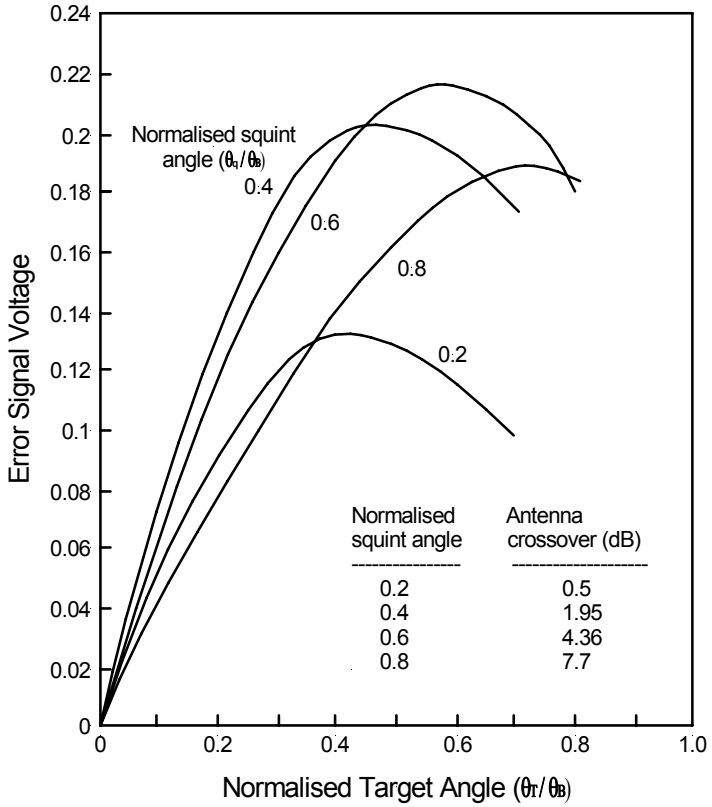


Figure 13.24: Some conscan angle error demodulation techniques (a) analog, (b) digital in hardware and (c) digital in software

An implementation of the digital solution in software is discussed in detail in the following section.

13.7.1. The Squint Angle Optimisation Process

It can be seen in the figure below that the error voltage is a function of the squint angle and the target angular error each as a function of the antenna beamwidth.



θ_q – Squint angle θ_T – Target angle θ_B – Antenna 3dB beamwidth

Figure 13.25: Error voltage as a function of target angle with squint angle as a parameter

The greater the slope of the error signal, the more accurate will be the angle tracking. This occurs at θ_q/θ_B just greater than 0.4

The gain on boresight is about 2dB down on the peak

The range tracking accuracy is determined by S/N, and so requires the maximum on boresight gain $\theta_q/\theta_B = 0$ which is not feasible.

A compromise is used with $\theta_q/\theta_B = 0.28$, which corresponds to an antenna gain about 1dB below the peak.

13.7.2. The Transfer Function of a Conically Scanned Antenna

In the same way that a split gate range tracker can be characterised by its range transfer function, so too can an angle tracking system be characterised by its angle error transfer function.

To perform this measurement, a conscan antenna mounted on a pan-tilt unit, as illustrated in the following figure, is swept past a point source of radiation (in the far field of the antenna). The received signal level at the output of the antenna is logged as a function of angle (or time if the angular rate is constant).

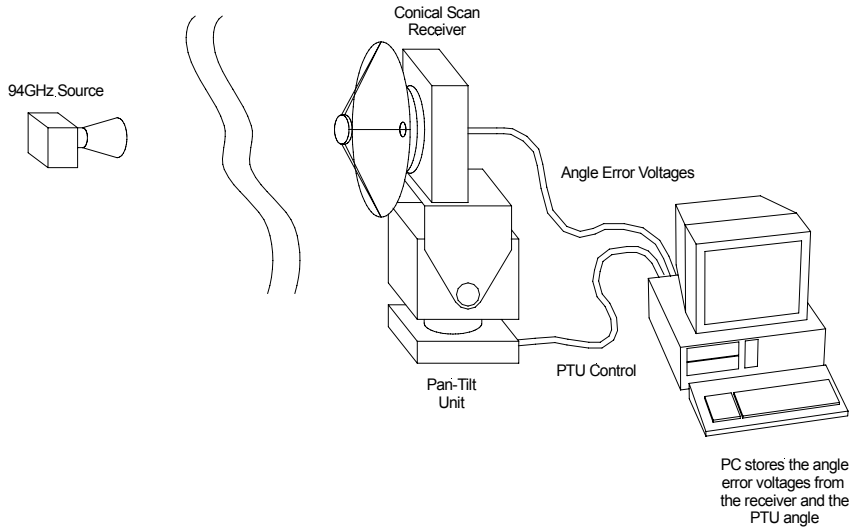


Figure 13.26: Mechanism for measuring conscan error transfer function

The received signal are as shown in the figure (a) below:

- The AC component of the signal level starts out low, when the offset is large
- It then increases to a peak when the beam squint angle equals the target offset on the one side
- On axis, the beam offset is symmetrical so there is no modulation, and so the AC component reduces to zero
- At the cross-over point the phase of the modulation is reversed but the shape of the modulation is the mirror image due to the symmetry of the process

In (b), both the modulation signal and the reference signal are shown. Note the 180° phase shift at crossover which is an indication that the sign of the error is reversed.

In (c), the synchronously demodulated signal is shown. Note that there is a DC component and an AC component at twice the modulation frequency.

Figure (d) shows the filtered demodulated signal showing only the DC which is the conscan angular transfer function for the Cassegrain antenna shown.

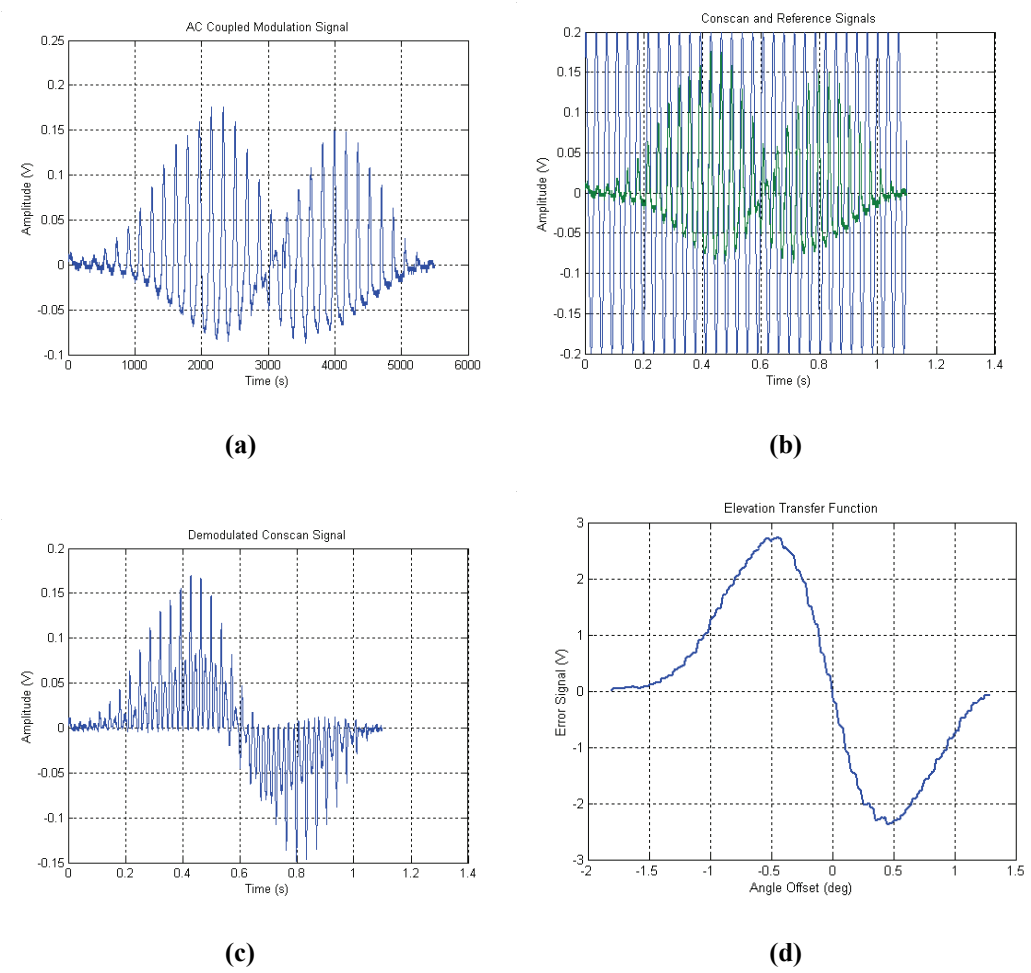


Figure 13.27: Conscan demodulation process showing the following: (a) modulated signal as antenna sweeps past source (b) modulated signal and reference signal showing phase reversal (c) the demodulation product of the signal and the reference (d) filtered demodulated error signal mapped to angle

13.7.3. Application

Conscan systems are typically used in tracking radars as shown in the figure below. They are simple to implement, using a single beam and a single receiver and transmitter, and beacon tracking can be implemented without the transmitter and without the range gating circuitry.



Figure 13.28: Conscan radar tracker on pedestal

As shown in the schematic diagram of the tracker, below, the AM signal out of the range gate is demodulated by the azimuth and elevation reference signals to produce the two angle error signals.

These angle errors drive the angle servos which in turn control the position of the antenna, and drive it to minimise the error (a null tracker).

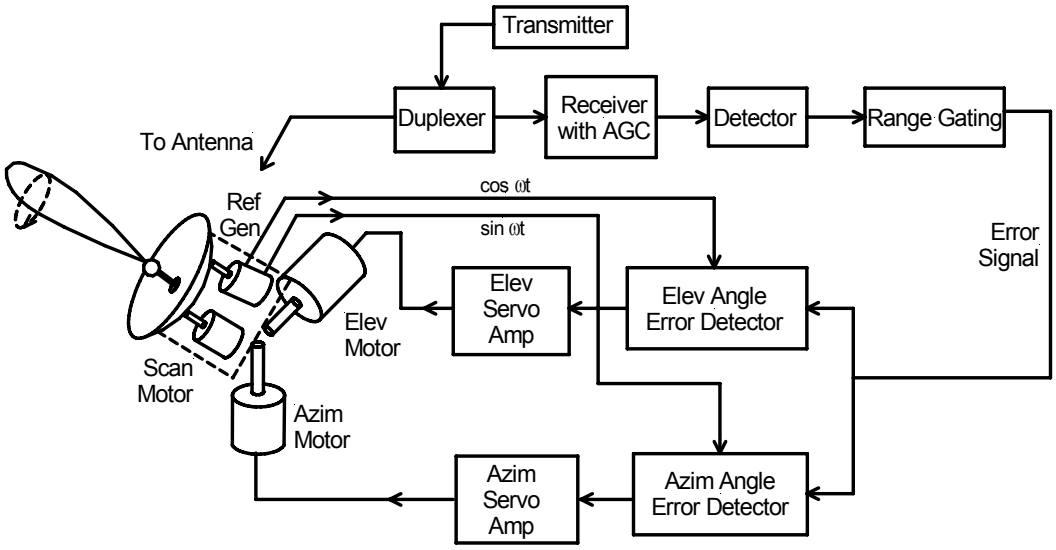


Figure 13.29: Block diagram of a conical scan radar

13.7.4. Main Disadvantages

The tracking bandwidth is reduced because of the number of pulses required to produce each error estimate.

Fluctuations in the echo signal amplitude induce tracking errors and so these systems are sensitive to target modulation. However modified conscan techniques have been developed to eliminate the modulation problem:

- Dual Conscan (Russian Design)
- COSRO (Conscan on receive only)

The antenna gain is reduced on boresight due to the squint angle. A trade-off is required to optimise the tracking accuracy.

13.7.5. Other considerations

Automatic Gain Control (AGC) that is required to normalise the pulse amplitude for range tracking must be carefully designed not to interfere with the conscan modulation.

13.8.Amplitude Comparison Monopulse

In essence this technique uses two overlapping antenna beams for each of the two orthogonal axes that are generated from a single reflector illuminated by 4 adjacent feed horns. The difference in the amplitude (or phase) of the signals output by these beams is used to derive the angle errors in both elevation and azimuth.

13.8.1. Antenna Patterns

The sum pattern, Σ, of the 4 horns shown in the figure below is used on transmit and for range measurement on receive.

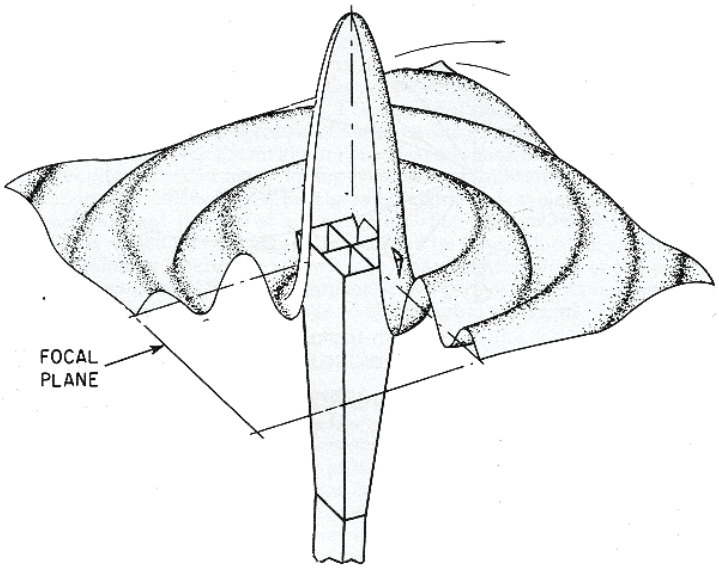


Figure 13.30: Monopulse sum-channel beam pattern

On receive the four horns produce 3D beam patterns as shown in the figure below.

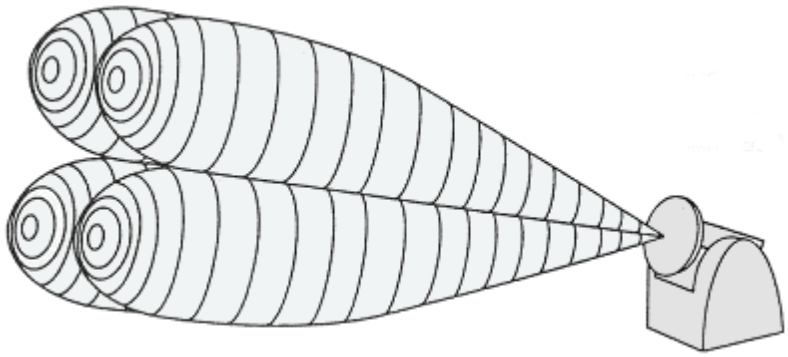


Figure 13.31: Monopulse difference-channel beam patterns

When viewed from the front, the beam patterns are as shown in the cross section shown below.

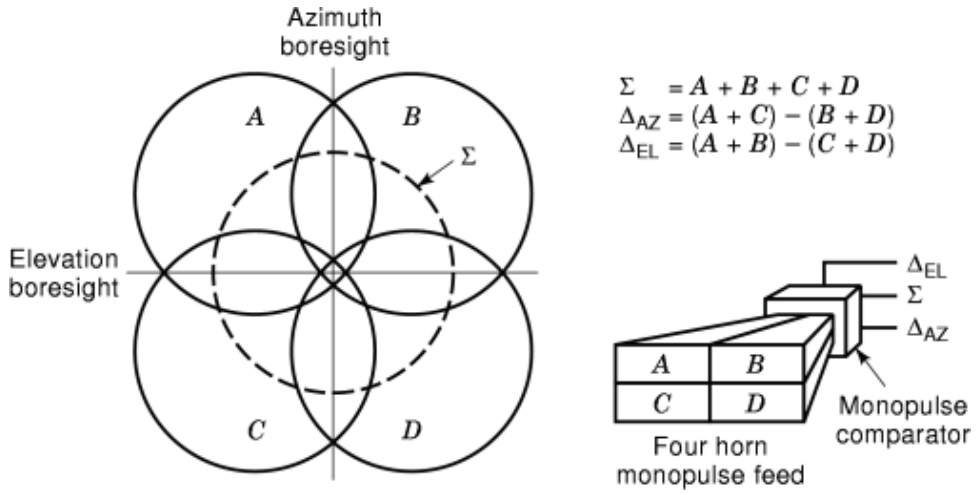


Figure 13.32: Combining beams using a microwave hybrid circuit (monopulse comparator) to produce sum and difference channel signals

13.8.2. Generation of Error Signals

The difference patterns ΔAz and ΔEl are produced on receive using a microwave hybrid circuit called a monopulse comparator. The hybrids perform phasor additions and subtractions of the RF signals to produce output signals shown in the following figure.

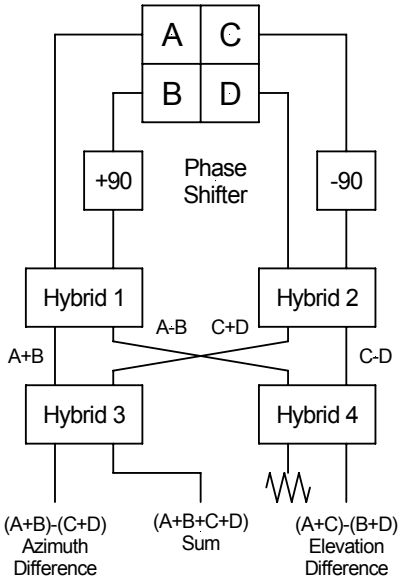


Figure 13.33: Hybrid configuration to produce sum and difference channel signals at the transmit frequency

It can be seen that if a target is on boresight, then the amplitudes of the signals received in the four channels (A,B,C,D) will be equal, and so the difference signals will be zero. However, as the target moves off boresight, the amplitude of the signals received will differ, and the difference signal will take on the sign and magnitude proportional to the error.

The difference channel signals are normalised with respect to the sum channel to produce an error signal that is independent of the echo amplitude as shown in the figure below.

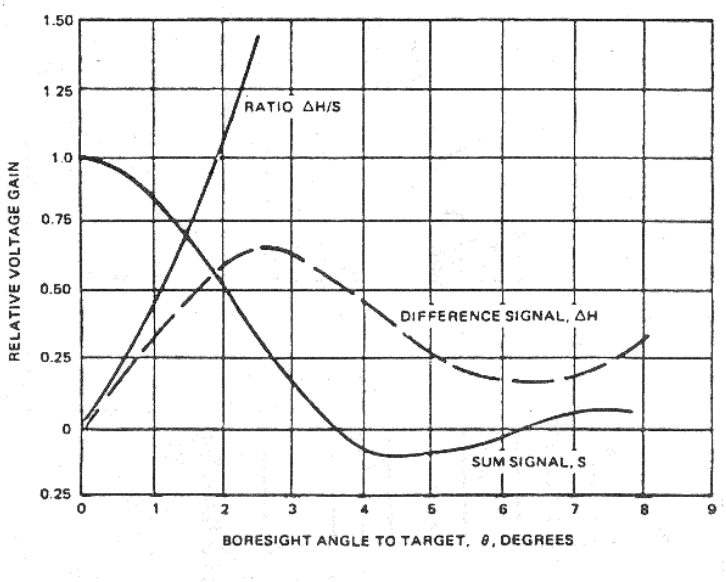


Figure 13.34: Normalised sum and difference channel gains as a function of angle off boresight

This ratio can be obtained using an AGC circuit that operates on the two difference channels and is driven by the detected sum-channel output of the tracking gate, or by division in a digital tracker.

Phase detectors demodulate the azimuth and elevation error signals using the sum channel IF signal as a reference to produce the two error voltages. These must also be range-gated so that their magnitudes represent the error signals from the correct target.

The sum and difference channel voltage signals can be modelled quite accurately using the following

$$\begin{aligned} E_{sum} &= \cos^2(1.14\Delta) \\ E_{dif} &= 0.707\sin(2.28\Delta) \end{aligned} \tag{13.13}$$

where E_{sum} – Normalised sum channel output voltage,
E_{dif} – Difference channel output voltage (normalised to sum channel),
Δ - Angle from the beam axis normalised wrt to the half power sum channel beamwidth.

New error measurements are produced with every pulse, hence MONOPULSE

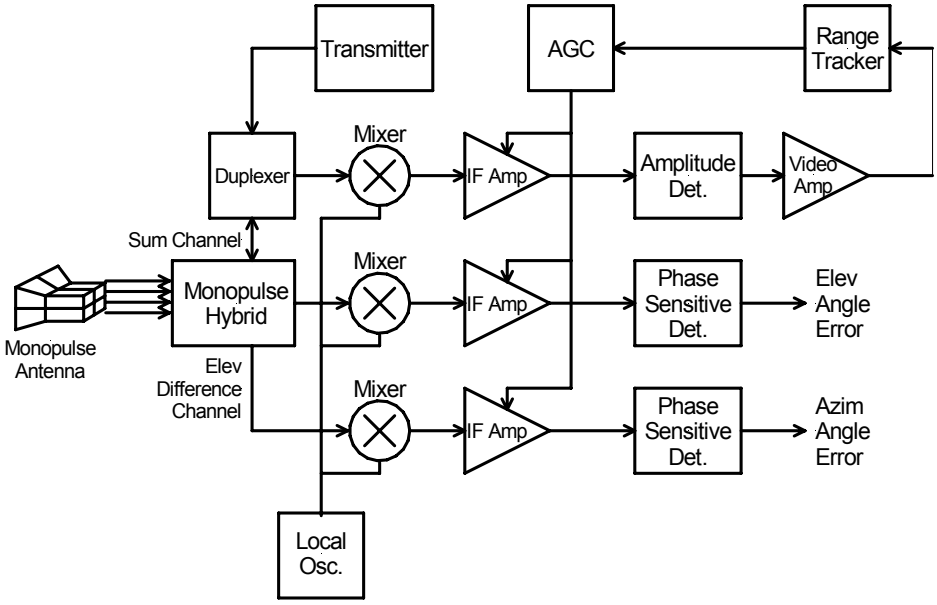


Figure 13.35: Schematic diagram of a monopulse front end

13.9.Comparison between Conscan and Monopulse

The monopulse option gives greater SNR for the same size target due to the higher on-boresight antenna gain, only the gains of the difference channel signals are reduced by the beam squint angle. In addition the steeper error slope near the origin results in superior tracking accuracy, and because new tracking information is generated with each new pulse, tracking is not degraded by fluctuations in echo amplitude.

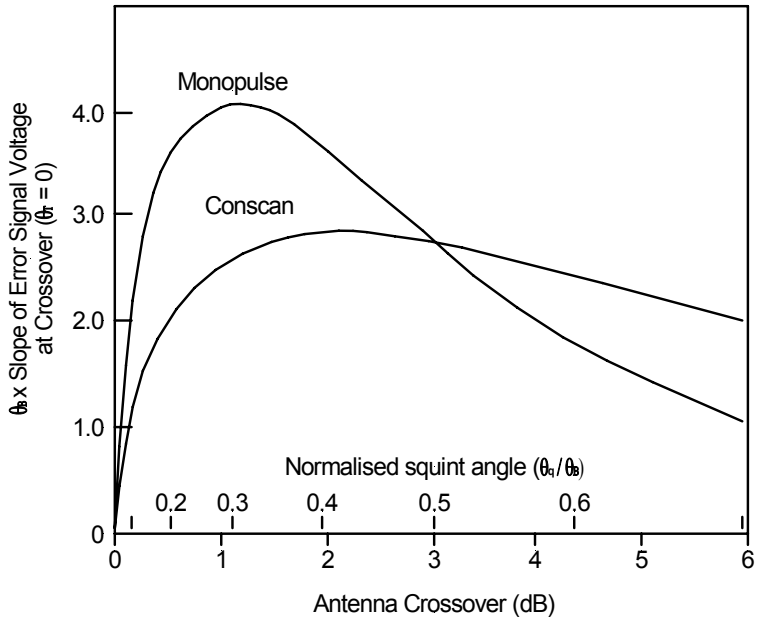


Figure 13.36: Error signal slope as a function of squint angle

13.10. Angle Tracking Loops

Unlike the range tracking loop in which the gate position is controlled electronically, in an angle tracker the beam must be physically displaced in angle to minimise the azimuth and elevation tracking errors.

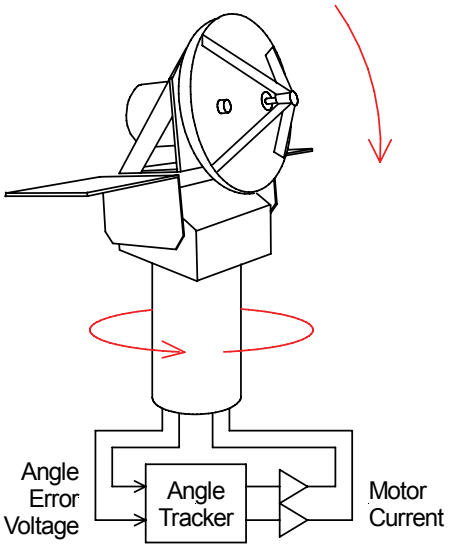


Figure 13.37: Angle tracking loop rotates the antenna to minimise tracking error

This physical displacement is achieved by supplying the angle servos with the error signals from the conscan or monopulse modules. These servos power the motors which rotate the antenna in the correct direction.

The RMS tracking error due to thermal noise is now

$$\sigma_t = \frac{\theta_{3dB}}{k\sqrt{2(S/N)f_r/\beta_n}} \text{ deg,} \qquad (13.14)$$

where: θ_{3dB} – Antenna Beamwidth (deg),
k – Constant dependant on the tracking type,
S/N – Signal to noise ratio,
f_r – Radar pulse repetition frequency (Hz),
β_n – Angle servo bandwidth (Hz).

13.11. Angle Estimation and Tracking Applications

In combination with range trackers as shown in the schematic above, monopulse techniques are generally used for modern tracking radar systems. The technique can be used for direction finding (DF) systems or beacon tracking in passive receivers. The principle can be used by EM or acoustic systems, and quadrant detectors that apply the same basic principles are generally used by laser trackers to estimate angle errors.

13.11.1. Example of an Acoustic Tracker

Two receiver transducers are mounted with a slight outward squint so that if the acoustic signal from the beacon arrives at an angle, the signal amplitude received by one of the receivers will be higher than that received by the other. Only when the illumination is symmetrical will the two signals be equal as shown in the figure below.

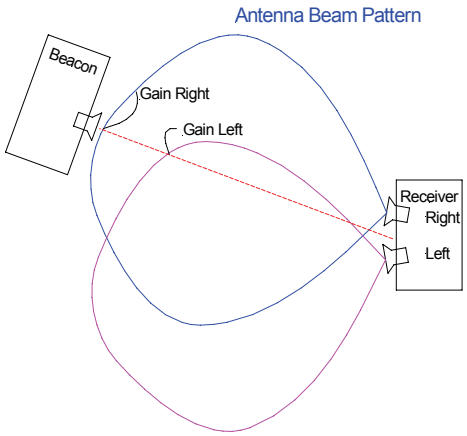


Figure 13.38: Conceptual diagram showing analog angle measurement

A simulation model was built using the ultrasonic transducer normalised gain pattern supplied by the manufacturer as shown in the polar plot below.

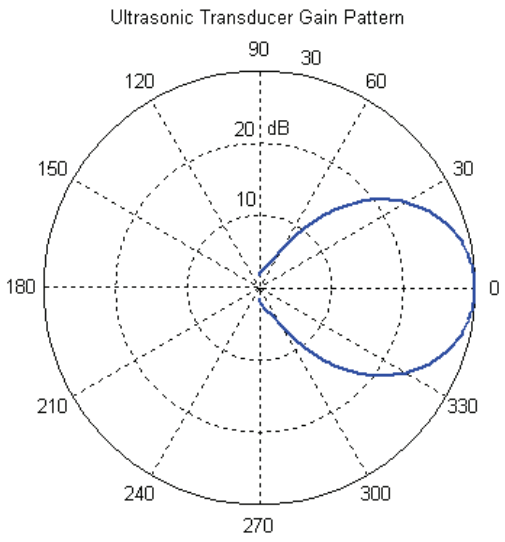


Figure 13.39: Polar pattern of ultrasound sensor normalised gain

The difference between the signal voltages received from the left and the right transducer is calculated to produce the receiver transfer function as shown in the following figure. It can be seen that the peak magnitude of the error signal is dependent on the squint angle. It can also be seen that there is a region of about +/-20° over which the relationship between the voltage and the beacon angle is almost linear.

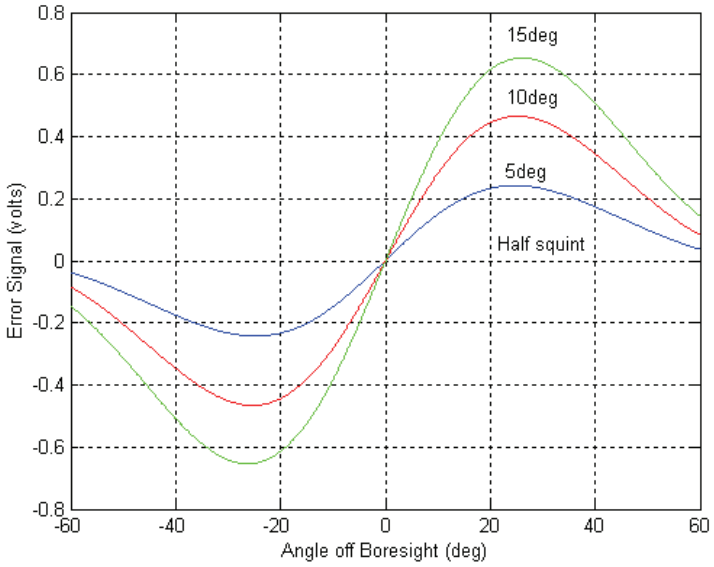


Figure 13.40: Angle error transfer function with squint angle as a parameter

13.11.2. Instrument Landing System (ILS)

An inverted application of the monopulse, dual beam concept

The ILS facilities are a highly accurate and dependable means of navigating an aircraft to the runway in low visibility

It consists of

- A localiser transmitter
- A glide path transmitter
- An outer marker (can be replaced by a non directional beacon or other fix)
- The approach lighting system

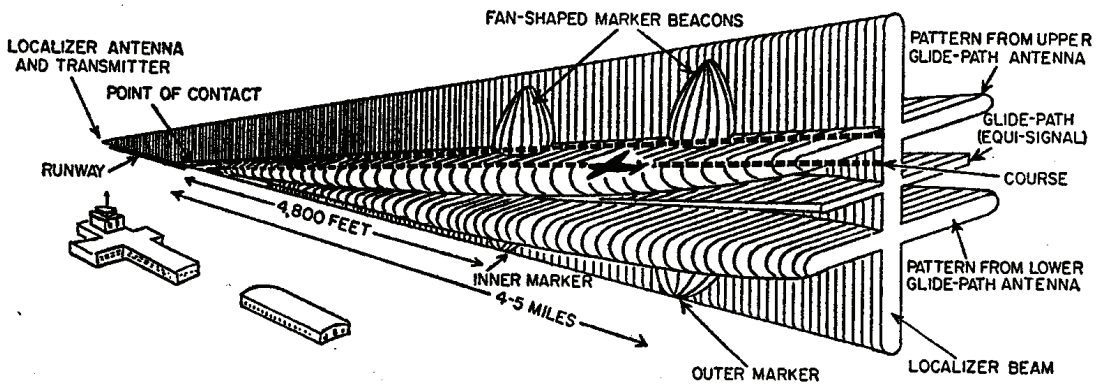


Figure 13.41: ILS signal patterns

A category I ILS provides guidance information down to a decision height (DH) of 200ft and with good equipment ILS can even be used for Category II approaches, 100ft on the radar altimeter. The ILS provides both lateral and vertical guidance necessary to fly a precision approach if glide slope information is provided.

Localiser Transmitter

The transmitter provides lateral guidance. It operates at VHF in the range 108.1MHz to 111.95MHz. The transmitter and antenna are situated on the centreline at the opposite end of the runway from the approach threshold. The antenna radiates two vertical fan shaped beams that overlap at the extended centre line of the runway. To differentiate between the two overlapping beams that are radiating at the same frequency, the right hand side of the beam (as seen by an approaching aircraft) is modulated at 150Hz and the left hand beam at 90Hz. The total width of the beam pair can be varied between 3° and 6°. It is adjusted to provide a track signal 700ft wide at the runway threshold increasing to 1nm wide at a range of 10nm.

Localiser Receiver

The localiser receiver activates the needle of one of the cockpit instruments as illustrated above. For an aircraft to the right of the beam, in the 150Hz region only, the needle will be deflected to the left, and visa versa in the 90Hz region. In the overlap region, both signals apply a deflection force to the needle causing a deflection in the direction of the strongest signal so that when the aircraft is precisely aligned, there will be zero net deflection, and the needle will point straight down..

Glide Slope Equipment

The glide slope equipment consists of a transmitter and antenna operating in the UHF at a frequency between 329.30MHz and 335.00MHz. It is situated between 750 and 1250 ft down the runway from the threshold, offset 400 to 600ft to the one side of the centreline and it is monitored to a tolerance of +/-0.5°. It consists of two overlapping horizontal fan beams modulated at 90 and 150Hz respectively and the thickness of the overlap is 1.4°, or 0.7° above and below the optimum glide slope. The glide slope may be adjusted between 2 and 4.5° above the horizontal plane depending on any obstructions along the approach path. Because of the antenna construction, no false signals can be obtained at angles below the selected glide slope, but are generated at multiples of the glide slope angle. The first is at about 6°. It can be identified because the instrument response is the reciprocal of the correct response.

The glide slope signal activates the glide slope needle in a manner analogous to that of the localiser. If there is sufficient signal, the needle will show full deflection until the aircraft reaches the point of signal overlap, at this time the needle will show partial deflection in the direction of the strongest signal. When both signals are equal, the needle shows horizontally indicating that the aircraft is precisely on the glide path. With 1.4° of overlap, the glide slope area is approximately 1500ft thick at 10nm reducing to less than 1ft at touchdown. A single instrument provides indication for both vertical and lateral guidance.

13.12. Triangulation

Though triangulation is not really a range or angle estimation technique, it is probably the most common method in use today for measuring the position of a target in space. The basic technique is thousands of years old and includes the following options

- Bearing only
- Range only
- Hybrid range and bearing

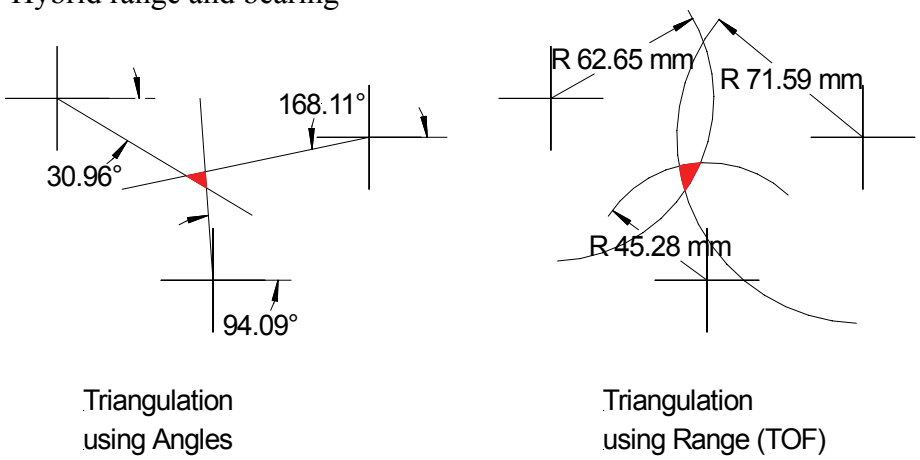


Figure 13.42: Position estimation by triangulation

Applications include

- Navigation: GPS, Omega, Loran-C
- Optical surveying
- Laser triangulation
- Radio emission detection
- Pinpointing earthquakes

13.12.1. Loran-C

LoRaN is an acronym for Long Range Navigation. It is a highly accurate (though not as accurate as GPS) highly available, all weather system for navigation in the coastal waters around the US and many other countries. An absolute accuracy of better than 0.25nm is specified within the region of coverage. And as with GPS, it is also used as a precise time reference,

A chain of three or more land based transmitting stations each separated by a couple of hundred km is used instead of a constellation of satellites. Within the chain, one station is designated as the master (M), and the other transmitters as secondary stations conventionally designated V,W,X,Y,Z.

The master station and the secondaries transmit radio pulses simultaneously at precise intervals and a Loran-C receiver on-board a ship or aircraft measures the slight difference in the time of arrival of the pulses. The difference in the time of arrival for a given master-secondary pair observed at a point in the coverage area is a measure of the difference in distance from the vessel to the two stations

The locus of points having the same TD from the pair, is a curved line of position (LOP).These curved lines are hyperbolas, or more correctly spheroidal hyperbolas on the curved surface of the earth and the intersection of two or more LOPs from different master-secondary pairs determines the position of the user as shown in the figure below.

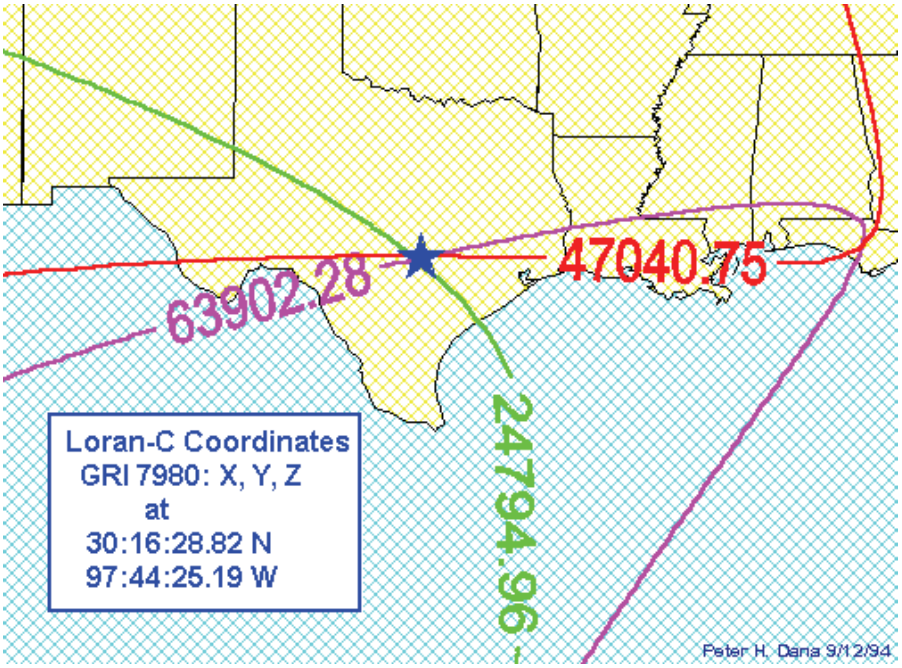


Figure 13.43: Intersection of two LOPs determines the receiver position

Why would anyone want Loran-C?

- it is inexpensive to operate, at $17 million per year and user equipment is low-cost and of proven capabilities. GPS requires in excess of $500 million per year.
- Loran's signal format has an integrity check built in, the GPS does not. Loran is easy to service ÷ just drive to the transmitting station, GPS replacement requires a new satellite launch schedule.
- On-air time for Loran-C is about 99.99%, GPS about 99.6% These features have convinced over 25,000 users to sign petitions to keep Loran-C online to its original date, 2015.
- Foreign nations have purchased new solid-state Loran-C transmitters to form new Loran-C chains in Europe, China, Japan and Russia. This land-based navigation system is viewed as a desirable, stable complement to the future GNSS of the world civilian community.
- And Loran is totally unclassified and is operated by host country authorities ÷ not the U.S. military.

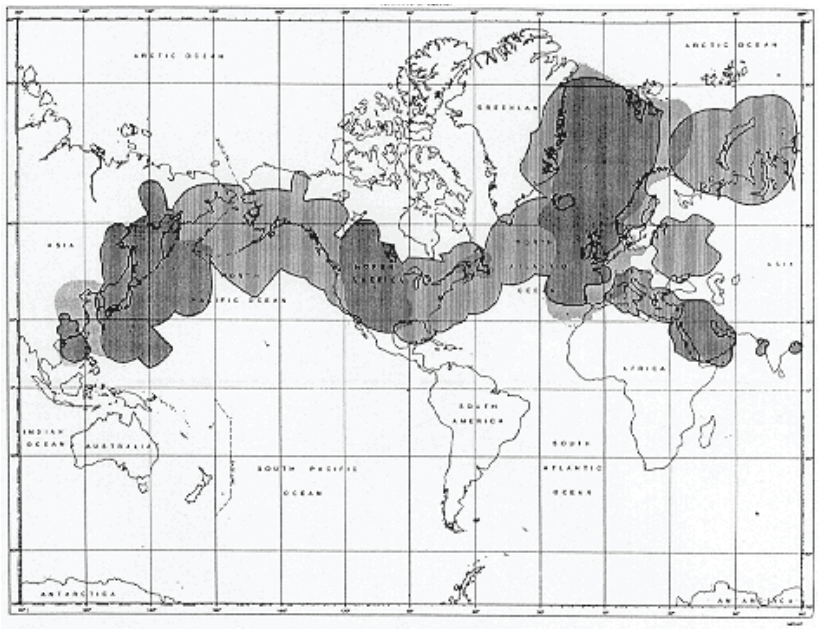


Figure 13.44: Loran-C coverage

Why does the military, who designed the GPS system, not rely on it totally?

- The ease of which GPS can be jammed either on purpose or by unexpected interferences is certainly a major reason. The deliberate jamming is well documented by the U.S. military, the International Association of Lighthouses and the Civil Aviation Authority (UK).
- A one watt jammer about 5x5x10cm with a 10cm antenna can block out a 60km diameter circle. Picture that near your local airport (such a unit costs about $100.USD).
- If you want a jammer for GPS and Glonass, (the Russian equivalent to the GPS), such units were offered for sale by the Aviaconversia Company, Russia, which displayed them at a recent Moscow Air Show. Their jamming range was said to be 200km. What was the reaction by the FAA? "Nothing new" because there are "hundreds of these devices on the market".

13.13. References

[1] D.Barton, *Radar Systems Analysis*, Artech 1976.
[2] M.Skolnik, *Introduction to Radar Systems 2nd ed,* McGraw Hill, 1980
[3] G.Morris, *Airborne Pulsed Doppler Radar,* Artech, 1988.
[4] FG21-P –Principle of Operation, http://www.riegl.co.at/fg21p/21p_prin.html, 04/04/2000
[5] SpeedTrap, http://www.audicoupe.demon.co.uk/speedtrap_types.html, 22/02/2000
[6] M.Skolnik, *Radar Handbook,* McGraw Hill, 1970
[7] N.Currie (ed), *Radar reflectivity Measurement: Techniques & Applications,* Artech House, 1989.
[8] M.Skolnik, *Introduction to Radar Systems,* McGraw Hill, 1980
[9] Navigation Systems- The Instrument Landing System, http://www.allstar.fiu.edu/aero/ils.html, 29/02/2000.
[10] Marine Radio Beacons, http://www.navcen.uscg.mil/policy/loran/h-book/book-1.txt, 28/02/2000.
[11] Loran C, http://www.landings.com/_landings/reviews-opinions/loran-c.html, 18/04/2001

Chapter 14.

Doppler Measurement

14.1.The Doppler Shift

The apparent difference between the frequency at which sound or light waves leave a source and that at which they reach an observer, caused by relative motion of the observer and the wave source. This phenomenon is used in astronomical measurements, in radar and modern navigation. It was first described (1842) by Austrian physicist Christian Doppler.

Examples of the Doppler effect include the following:

- As one approaches a blowing horn, the perceived pitch is higher until the horn is reached and then becomes lower as the horn is passed.
- The light from a star, observed from the Earth, shifts toward the red end of the spectrum (lower frequency or longer wavelength) if the Earth and star are receding from each other and toward the violet (higher frequency or shorter wavelength) if they are approaching each other. The Doppler effect is used in studying the motion of stars.

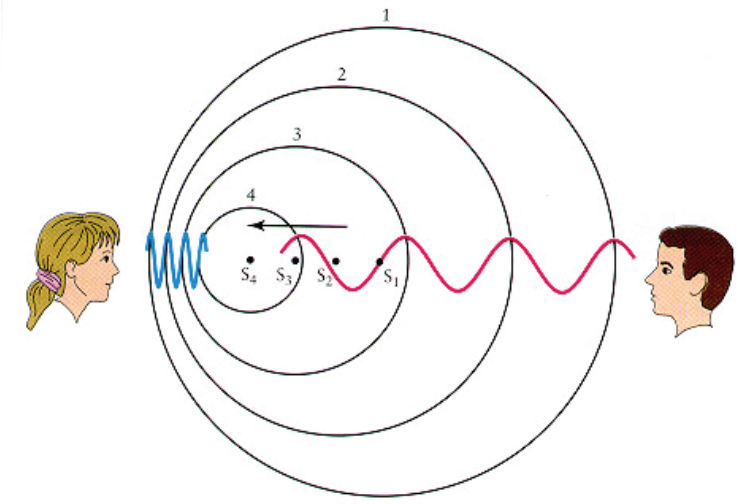


Figure 14.1: Illustration of Doppler shift for a sound source moving from S_1 to S_4

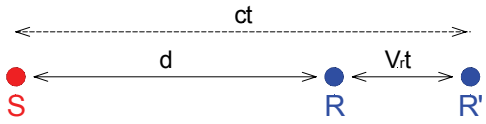
14.1.1. Doppler Shift Derivation

Consider the relationship between the frequency of sound produced by a source moving with velocity v_s and the frequency received by a receiver moving with velocity v_r. For simplicity, assume that the both the source and the receiver are moving in a straight line in the same direction.

At time t = 0, the source, S, and receiver R are separated by a distance *d*.



The source emits a wave that propagates at velocity *c* and reaches the receiver after time *t*.



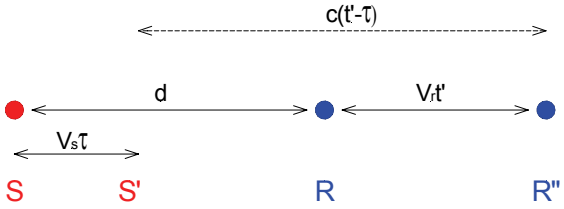
As the receiver has moved $v_r.t$ metres

$$ct = d + v_r t\,, \tag{14.1}$$

or

$$t = \frac{d}{c - v_r}\,. \tag{14.2}$$

At time τ the source, S, would have moved τv_s. Let the wave emitted at that instant be received at time *t'* by R. In this time R would have moved $v_r t'$



$$c(t' - \tau) = (d - v_s\tau) + v_r t'\,, \tag{14.3}$$

making

$$t' = \frac{d + (c - v_s)\tau}{c - v_r}\,, \tag{14.4}$$

Thus for the receiver, the interval between the waves has been

$$\tau' = t' - t = \frac{c - v_s}{c - v_r}\tau\,. \tag{14.5}$$

Whereas for the source, the interval between waves has been τ. Now the number of waves emitted in τ by the source must equal the number of waves received by the receiver in τ'

$$f_r \tau' = f_s \tau \,, \tag{14.6}$$

making

$$f_r = \frac{c - v_r}{c - v_s} f_s \,. \tag{14.7}$$

For v_s and $v_r << c$

$$f_r = \frac{1 - v_r / c}{1 - v_s / c} f_s = \left[1 - \frac{v_r}{c}\right]\left[1 - \frac{v_s}{c}\right]^{-1} . \tag{14.8}$$

Expanding the last term using the binomial expansion

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + ... \tag{14.9}$$

For $x<<1$, the higher order terms can be ignored

$$f_r \approx \left[1 - \frac{v_r}{c}\right]\left[1 + \frac{v_s}{c}\right] = \left(1 - \frac{v_{rs}}{c}\right) f_s \,. \tag{14.10}$$

where $v_{rs} = v_r - v_s$ is the velocity of the receiver relative to the source

The Doppler shift is thus

$$f_d = f_r - f_s = \frac{-v_{rs}}{c} f_s \,. \tag{14.11}$$

The frequency moving away from the source will be less than the frequency measured at the source, whereas the frequency measured at a receiver moving towards the source will be greater than the frequency measured at the source

14.2.Doppler Geometry

In most Doppler sensors both the transmitter and receiver are stationary, and they illuminate a moving target. In addition, for some applications (Doppler ultrasound imaging) they may not be co located as shown below

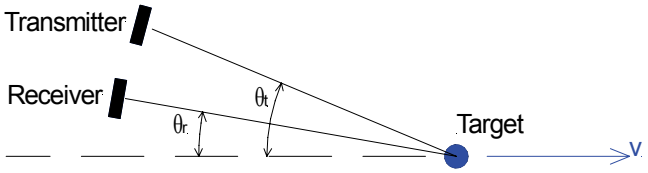


Figure 14.2: Doppler geometry: separated transducers

The velocity of the target relative to the transmitter will be $v\cos\theta_t$ and the velocity of the target relative to the receiver will be $v\cos\theta_r$

The Doppler shift arising under these circumstances can be calculated assuming that:

- The target is a receiver moving away from the source with a velocity $v\cos\theta_t$
- The receiver is moving away from the target (source) with a velocity $v\cos\theta_r$

This is equivalent to the receiver moving away from the source with velocity $v\cos\theta_t + v\cos\theta_r$ even though both are stationary.

14.2.1. Targets moving at low velocities (v<<c)

The Doppler frequency for separated transducers is

$$f_d = -\frac{f_s v}{c}\left(\cos\theta_t + \cos\theta_r\right), \tag{14.12}$$

$$f_d = -\frac{2f_s v}{c}\cos\left(\frac{\theta_r + \theta_t}{2}\right)\cos\left(\frac{\theta_r - \theta_t}{2}\right). \tag{14.13}$$

If the transmit and receiver transducers are co located then $\theta_r \approx \theta_t = \theta$ and the formula for the Doppler frequency reduces to

$$f_d = -\frac{2f_s v}{c}\cos\theta = -\frac{2v}{\lambda_s}\cos\theta\,. \tag{14.14}$$

14.2.2. Targets Moving at High Speed (v<c)

In this case it is not possible to use the EM radiation approximation, and the whole formula must be used for separated transducers. This is almost always the case with ultrasound measurements in air.

$$f_r = \frac{c - v\cos\theta_r}{c + v\cos\theta_t} f_s\,, \tag{14.15}$$

And the Doppler frequency is

$$f_d = f_r - f_s\,. \tag{14.16}$$

For co-located transducers $\theta_r \approx \theta_t = \theta$ just substitute for a common angle.

14.3.Doppler Frequency Extraction

This section outlines the instrumentation required to detect Doppler shifts in a received signal.

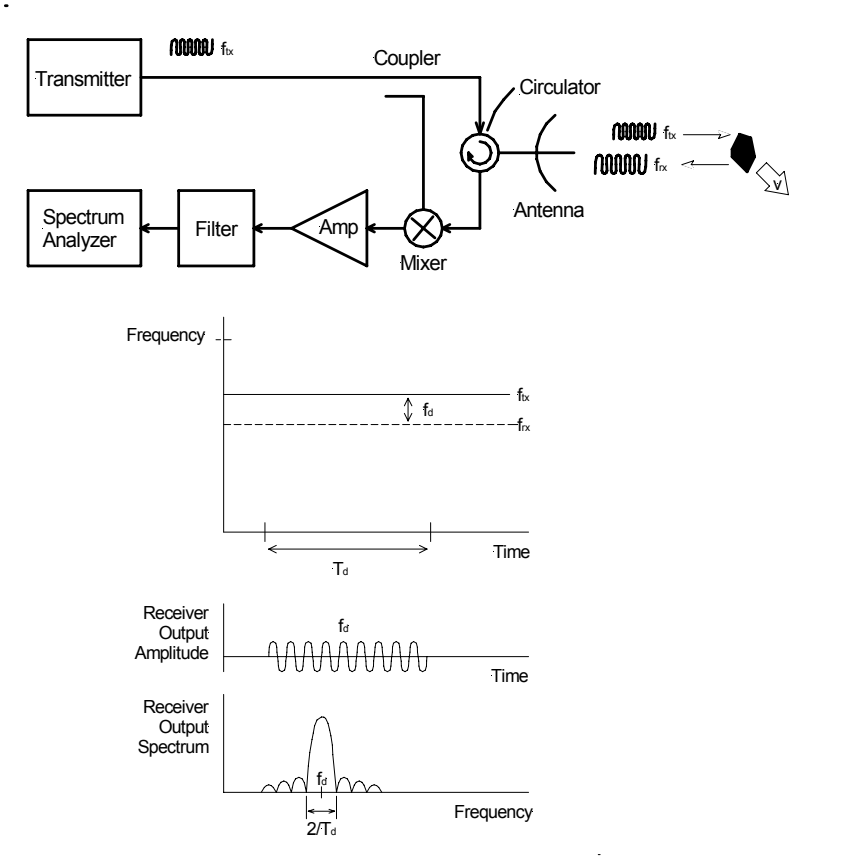


Figure 14.3: Configuration of a generic Doppler sensor

The transmitted signal is of the form

$$x_t(t) = \xi_t \cos(\varpi_s t) \tag{14.17}$$

The corresponding received signal from a single target will be

$$x_r(t) = \xi_r \cos([\varpi_s + \varpi_d]t + \phi), \tag{14.18}$$

where ϕ - phase term dependent on the distance to the target (rad),
ϖ_s - $2\pi f_s$ (rad/s),
ϖ_d - $2\pi f_d$ (rad/s).

The two signals are mixed (multiplied together) to produce

$$\begin{aligned} x_t(t)x_r(t) &= \xi_t\xi_s \cos(\varpi_s t)\cos([\varpi_s + \varpi_d]t + \phi) \\ &= \frac{\xi_t\xi_r}{2}\{\cos(\varpi_d t + \phi) + \cos([2\varpi_s + \varpi_d]t + \phi)\} \end{aligned} \tag{14.19}$$

This signal is low pass filtered as shown in the figure below to remove the component at $2f_s$ leaving only the Doppler signal

$$x_d(t) = \frac{\xi_t \xi_r}{2} \cos(\varpi_d t + \phi) . \tag{14.20}$$

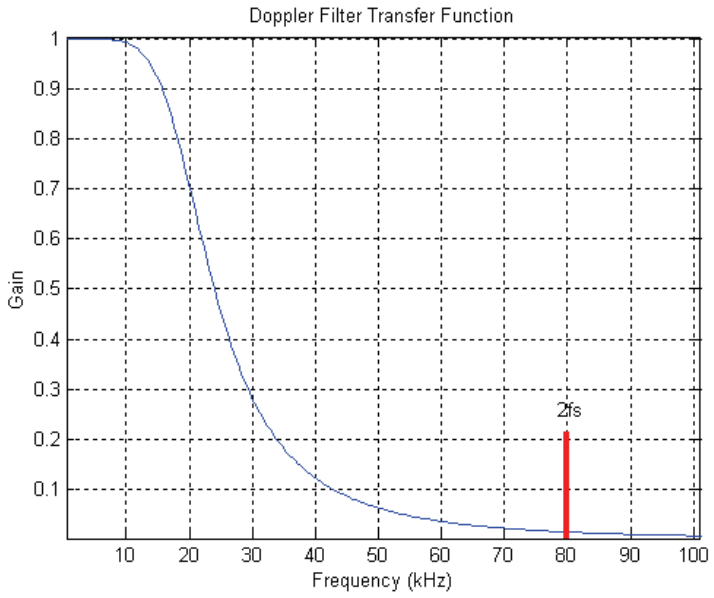


Figure 14.4: Lowpass filter to remove the component at $2f_s$

The reflected signal amplitude from non-moving objects in the beam will be 40 to 50dB larger than the Doppler signal, and so additional high pass filtering is often required to remove this.

14.3.1. Direction Discrimination

The Doppler process discussed above can provide only an absolute difference frequency, it contains no information regarding the direction of motion.

A number of techniques can be applied to preserve this directional information:

- Sideband filtering
- Offset carrier demodulation
- In-phase/ quadrature demodulation

In the descriptions that follow, it mist be remembered that

- $\varpi_d > 0$ Target velocity towards the sensor
- $\varpi_d < 0$ Target velocity away from the sensor

Sideband Filtering

The received signal is split and passed through two bandpass filters, one passing signals over the range $\varpi_s < \varpi < \varpi_s + \varpi_m$ and the other passing signals over the range $\varpi_s - \varpi_m < \varpi < \varpi_s$. The output of each filter passes through a mixer and filter as usual.

If the target is approaching the signal appears in the first bin, and if it is receding then it appears in the second

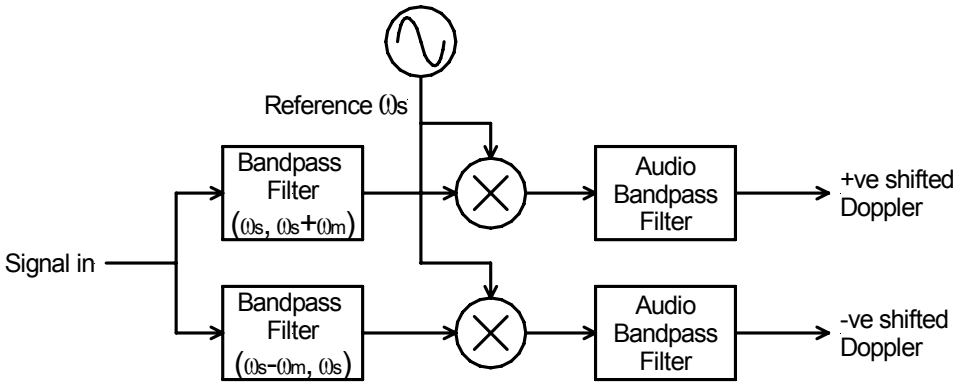


Figure 14.5: Sideband filtering

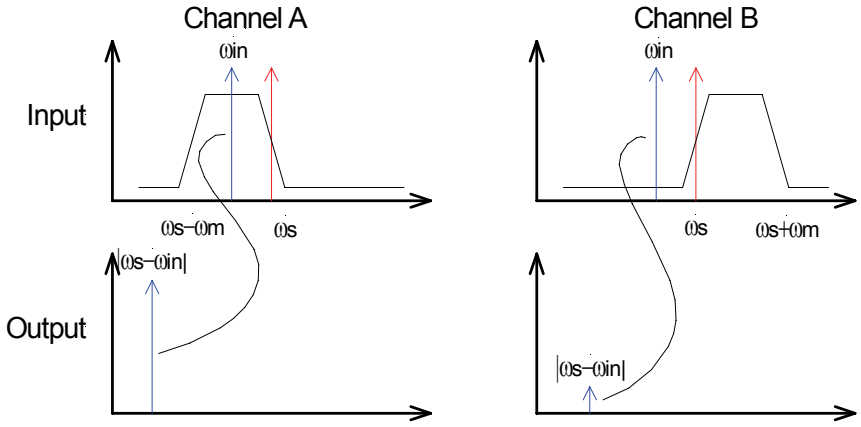


Figure 14.6: Sideband filtering spectra in the two channels

Offset Carrier Demodulation

This process involves heterodyning (mixing) the received signal by a reference signal $\varpi_1 + \varpi_s$.

The received signal is

$$x_r(t) = \xi_r \cos\left(\left[\varpi_s + \varpi_d\right]t + \phi\right), \tag{14.21}$$

and the reference signal is

$$x_1(t) = \xi_1 \cos\left(\left[\varpi_s + \varpi_1\right]t\right). \tag{14.22}$$

Mixing the two signals (14.21) and (14.22) gives

$$x_1(t)x_r(t) = \frac{\xi_1\xi_r}{2}\{\cos([\varpi_1 + \varpi_d]t + \phi)\cos([2\varpi_s + \varpi_1 + \varpi_d]t + \phi)\}, \qquad (14.23)$$

where ϖ_1 is chosen so that $\varpi_1 > |\varpi_{d\,max}|$,

and as usual the mixed signal is filtered to remove the component at $2\varpi_s$.

$\varpi_1 + \varpi_d > \varpi_1$ +ve shift Doppler

$\varpi_1 + \varpi_d < \varpi_1$ -ve shift Doppler

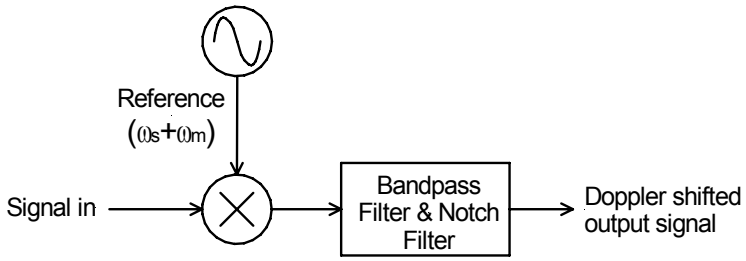


Figure 14.7: Offset carrier demodulation

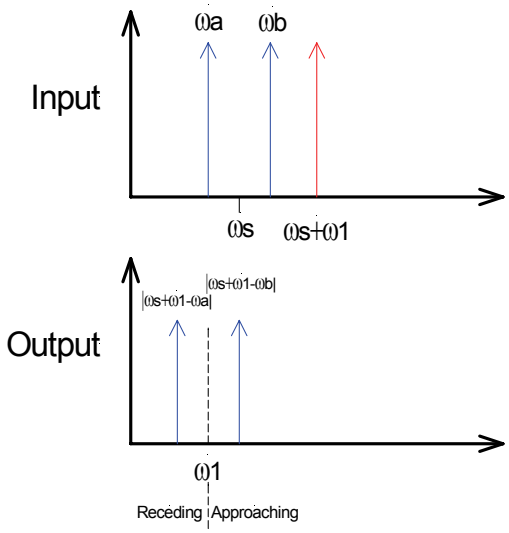


Figure 14.8: Offset carrier spectra

In Phase/ Quadrature Demodulation

The received signal is split into two channels. In the in-phase channel, it is mixed with the transmitted signal, and in the quadrature channel it is mixed with the transmitted signal but phase shifted by π/2.

Mixing and filtering as before results in the following signals

$$\begin{aligned} i(t) &= \cos(\varpi_d t + \phi) \\ q(t) &= \sin(\varpi_d t + \phi) \end{aligned} \qquad (14.24)$$

The direction of the Doppler shift, and hence the direction of flow is determined by noting the phase relationship between i(t) and q(t).

- $\varpi_d > 0$, $q(t)$ is $\pi/2$ retarded with respect to i(t)
- $\varpi_d < 0$, $q(t)$ is $\pi/2$ advanced with respect to i(t)

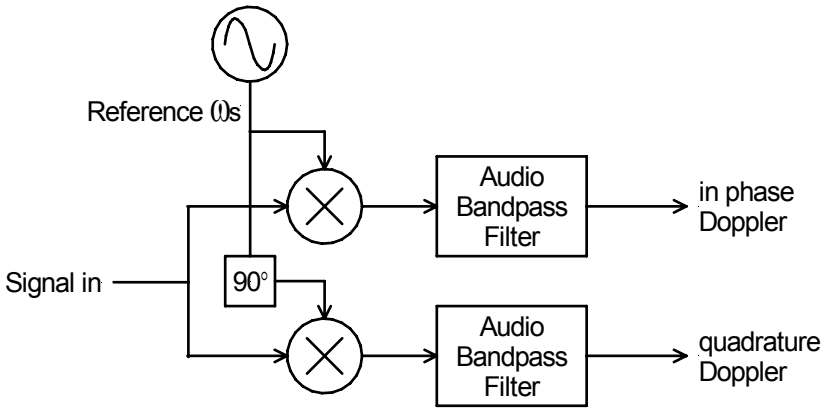


Figure 14.9: In-phase / quadrature demodulation

14.4. Pulsed Doppler

Pulsed Doppler is identical to the CW version except that the transmit signal is pulsed. This allows the technique to measure range as well as velocity. As shown in the figure, a reference version of the transmitted signal must be maintained so that the received signal can be synchronously detected

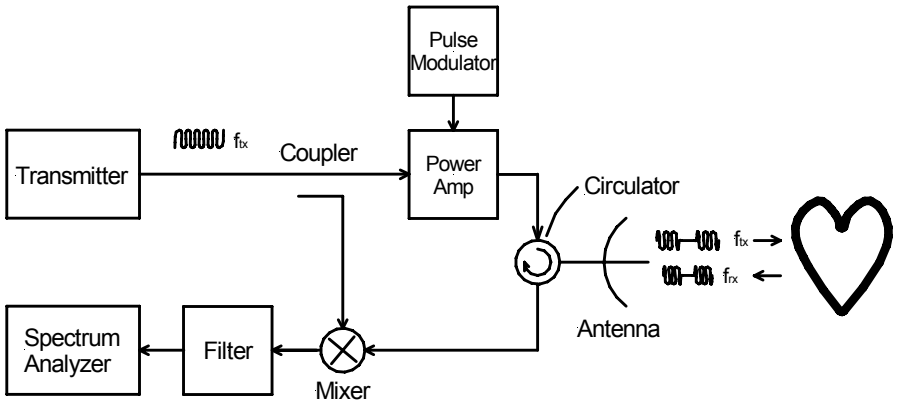


Figure 14.10: Pulsed Doppler ultrasound schematic diagram

The waveform that is received from a pulsed Doppler sensor in a specific range gate is shown in the figure below in which it should be noted that the number of cycles received during each return is a function of the pulsewidth and the Doppler frequency.

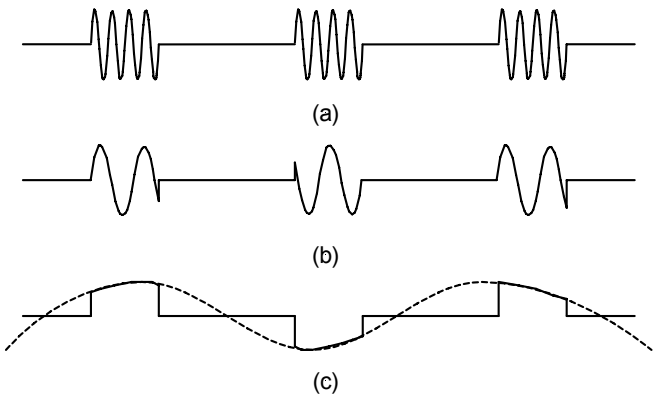
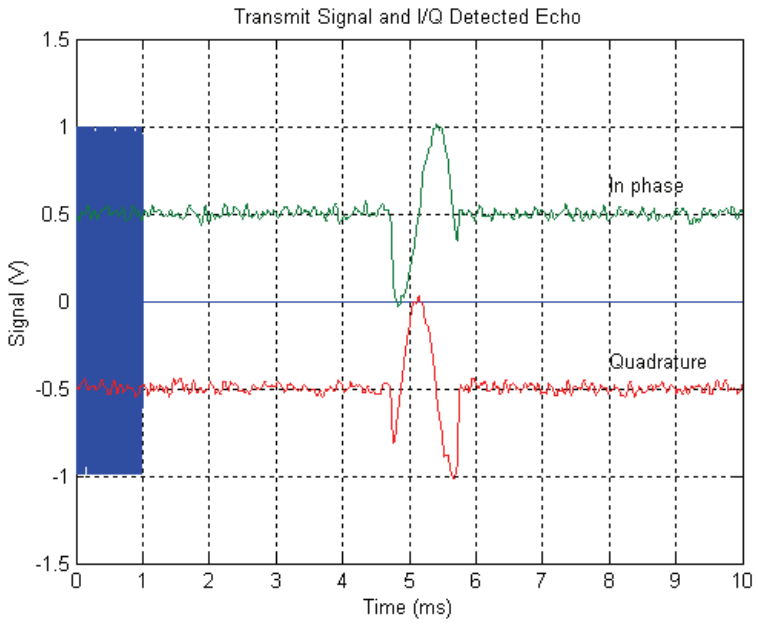


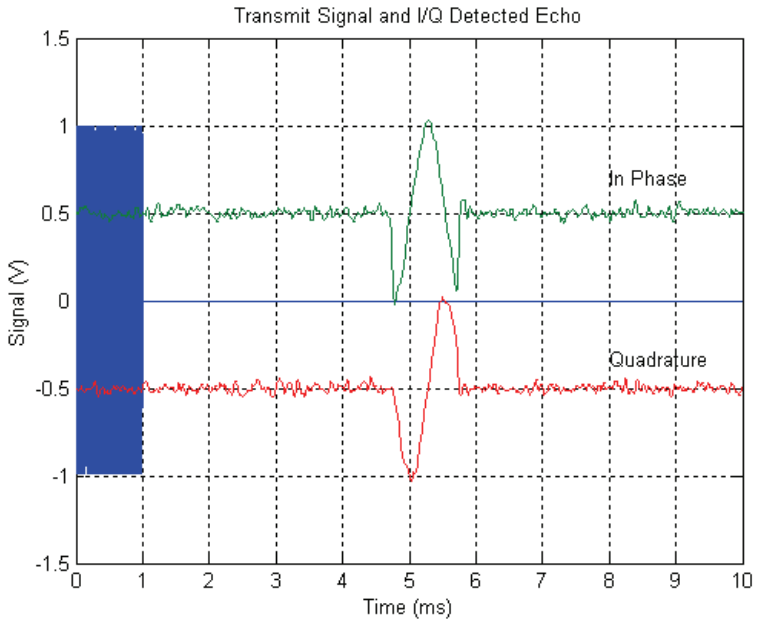
Figure 14.11: Pulsed Doppler waveforms (a) RF echo pulse train, (b) video pulse train for $f_d > 1/\tau$ and (c) video pulse train for $f_d < 1/\tau$

The following waveforms show the basic pulsed Doppler technique extended to incorporate In-phase and quadrature detection so that the direction of travel can be determined. In this figure it can be seen that the Quadrature signal leads the In-phase signal for a receding target (a) and that it lags the In-phase signal for an approaching target (b).

Note that the in-phase and quadrature traces have been shifted on the display so that they are easier to interpret.



(a)



(b)

Figure 14.12: I and Q Doppler frequencies for a receding (a) and an approaching (b) target

14.5.Doppler Sensors

Many sensors using both ultrasonic and radar technology make use of the Doppler principle to measure target motion effects. Such sensors can either operate using continuous wave (CW) or pulsed sources.

In general CW sources have no range discrimination, while pulsed sources can determine both range and velocity.

14.5.1. Continuous Wave Doppler Ultrasound

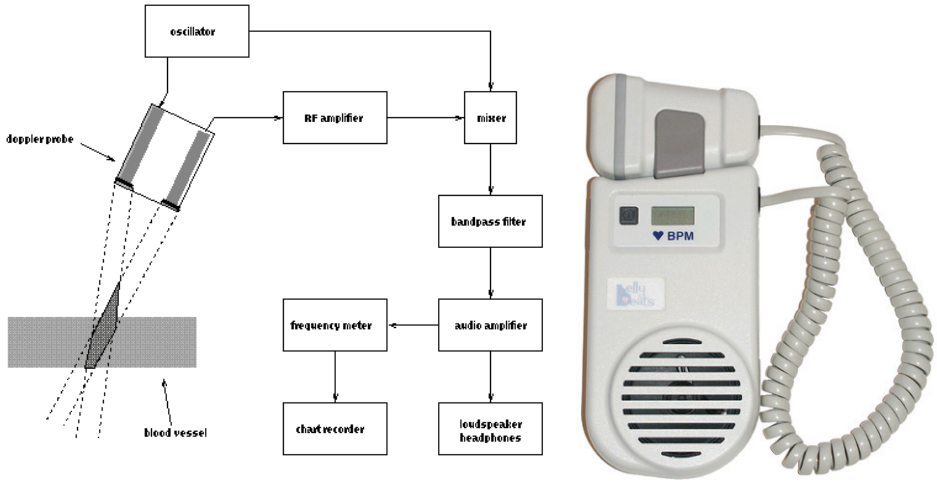
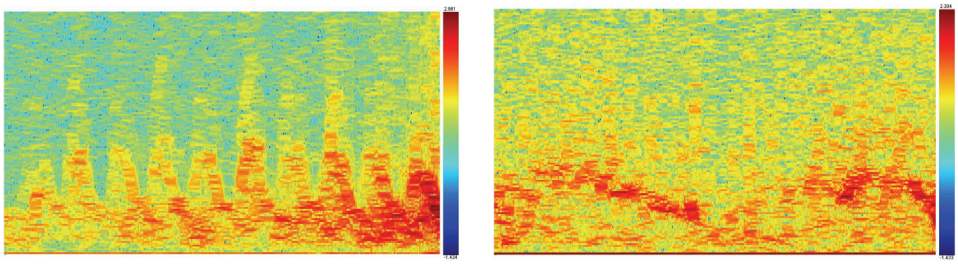


Figure 14.13: Generic CW Doppler instrumentation measuring blood flow in an artery

CW ultrasound systems are used to measure fluid flow and the movement of internal organs (particularly the heart) while industrial Doppler flowmeters measure flow in pipes and canals. In all cases they rely on the acoustic signal reflecting off the moving medium.

If a spectrogram is produced, then it is possible to identify specific target characteristics that vary with time. For example, the following figures show the difference between the return from a large dog and a human subject.



(a) (b)

Figure 14.14: Doppler spectrograms of (a) human subject and (b) large dog both walking towards the radar

14.5.2. Continuous Wave Doppler Radar

CW Doppler radar has a myriad of uses. These include short range intruder detection, moderate range sports and police radar for measuring the velocity of moving targets, and long range aircraft tracking applications.

Intruder Detection

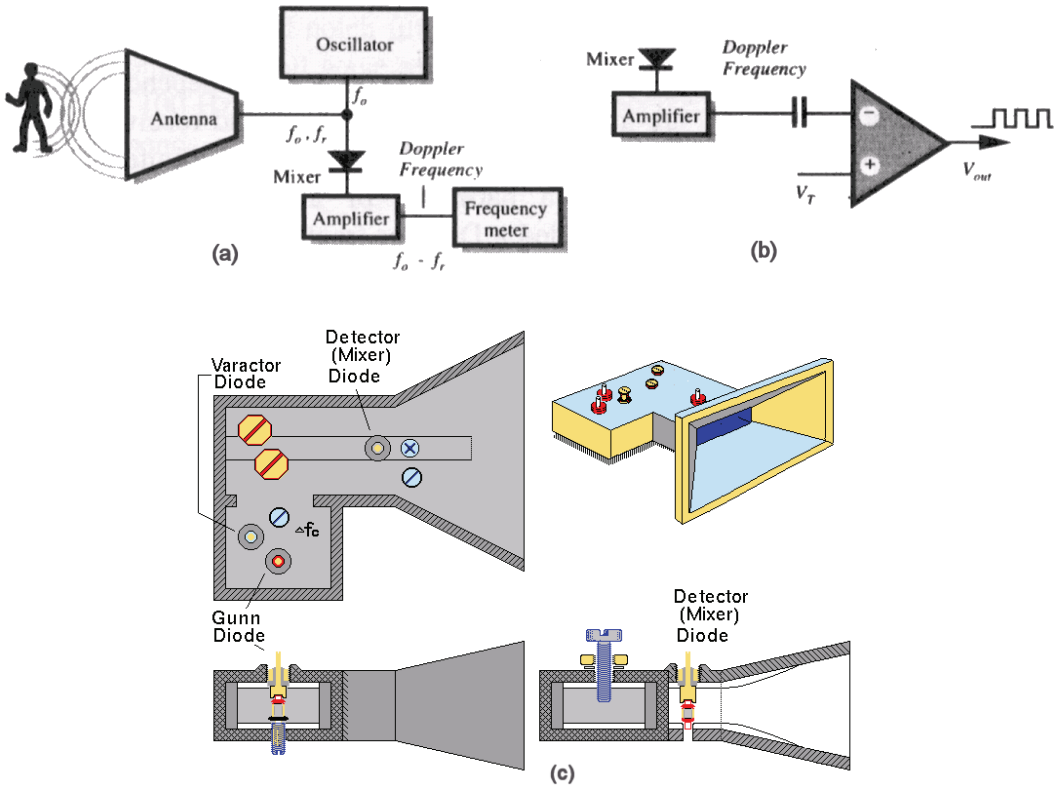


Figure 14.15: CW Doppler intruder detection. (a) a circuit for measuring Doppler frequency, (b) circuit with a threshold detector and (c) Gunn microwave Doppler module

Specifications of a typical intruder system:

- Operational frequency typically X-band (8-12GHz)
- Output power 1-10mW
- Operational principle, Gunn oscillator iris-coupled to the common antenna port that includes a mixer diode.
- Sometimes a varactor diode is mounted in the Gunn diode cavity to allow for electronic control of the oscillation frequency
- Antenna port is generally flared into a horn to constrain the beam angle

Sports Radar



(a) **(b)**

Figure 14.16: Doppler sports radar units, (a) tripod mounted and (b) hand held

Specifications of a typical sports radar

- Operational frequency typically K-band (24.15GHz)
- Output power 40-100mW
- Antenna generally a large horn or a horn-lens
- Range depends on antenna size, output power and target RCS
- Accuracy +/-1km/h (typical)

Police Speed Trap



Figure 14.17: Police radar

Specifications of a typical police radar

- Operational frequency, X-band, K-band or Ka-band
- Accuracy +/- 0.2% Microwave +/- .15 MPH internal signal processing
- +1/4, -3/4 mph readout truncation system accuracy at 100 MPH +0.6/-1.1 mph
- Patrol speed range +16 to +75 mph
- Target Speed Range Stationary/Moving 15 mph thru 199 mph
- Target detection distance: Approx. 1000m.

Aircraft Tracking Radar

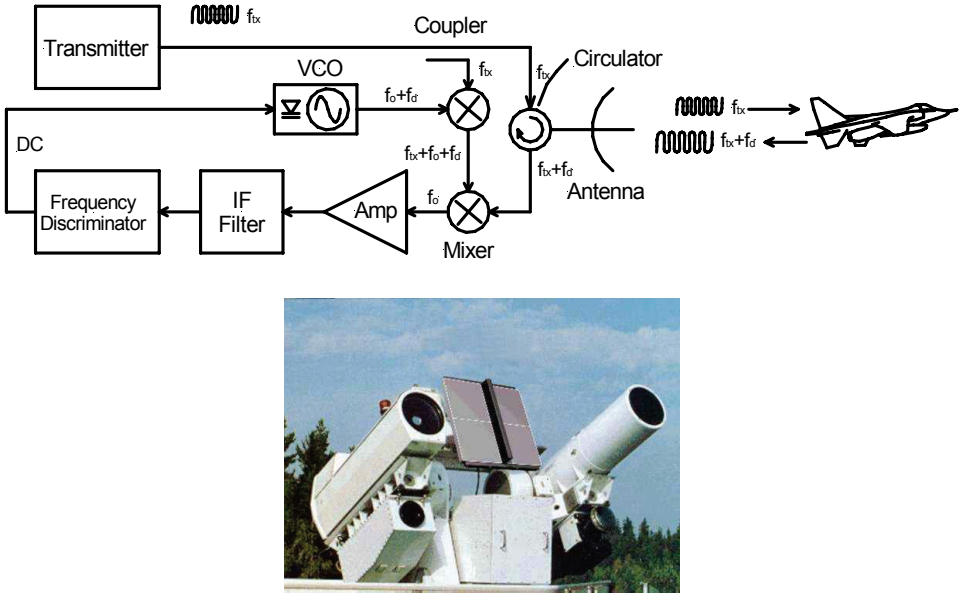


Figure 14.18: CW Doppler aircraft tracking

The block diagram shows a simple aircraft tracking Doppler radar similar to the radar shown in the photograph. The antenna is tracked manually in angles using one of the telescopes. Automatic tracking is carried out only on the velocity. A narrow-band tracking filter in the IF is used to reject noise and crosstalk from the transmitter. This filter is followed by a discriminator that provides an error signal that drives an automatic frequency control (AFC) loop which ensures that the received frequency (f_{tx}+/-f_d) is downconverted to a constant IF, f_o, irrespective of the actual Doppler frequency.

There will be a minimum Doppler frequency below which the transmitter spillover will swamp the received signal.

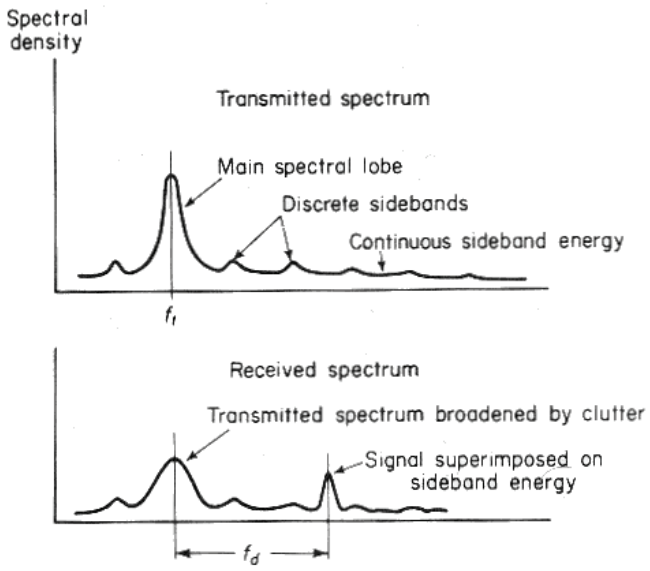


Figure 14.19: CW Doppler spectra for the tracking radar example

14.5.3. Pulsed Doppler Ultrasound



Figure 14.20: Doppler ultrasound of blood flow through the heart

Conventional CW sensors are unable to measure range, and so they are unable to separate Doppler signals returned from different targets within the beam except when beam pairs are used as shown earlier in this chapter.

The primary differences between CW Doppler and pulsed Doppler are as follows:

- A single transducer can be used for transmission and reception because the two are separated in time
- Pulsed Doppler can be incorporated into conventional pulsed echo ultrasound (often known as duplex scanning)
- Periodic bursts of ultrasound typically only a few cycles long are used.
- Pulsed Doppler is only sensitive to flow within a region termed the sampling volume.

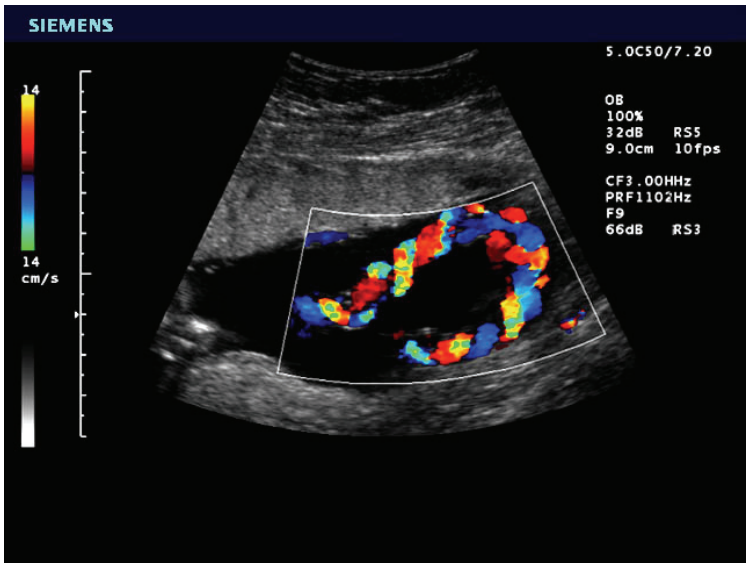


Figure 14.21: Doppler ultrasound image of umbilical cord

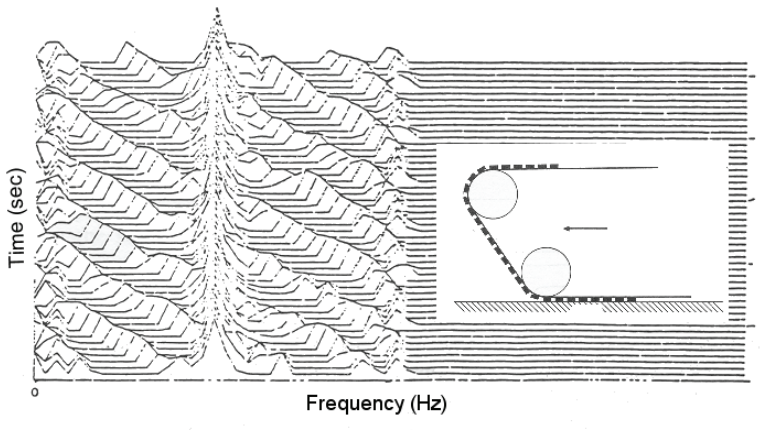
Range resolution is achieved by transmitting a short burst of ultrasound.

Following the burst, the received signal is mixed with a delayed version of the transmitted pulse as reference. The time of flight of the transmitted pulse to the target of interest and back again is equal to this delay. This allows the sampling volume to be moved to different positions along the beam by altering the delay, and in this way, flow at different depths can be selectively monitored. The width of the sampling volume is equal to the transmitted beamwidth, and the length is equal to the pulse width.

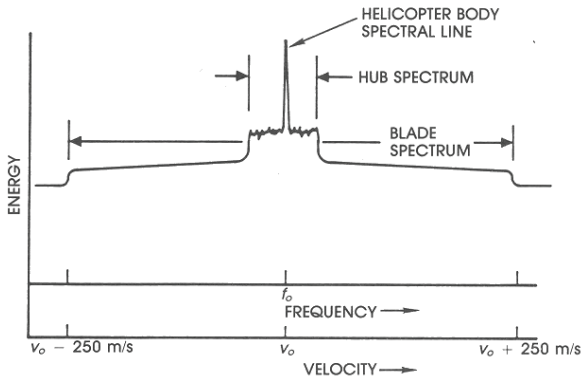
Imaging systems such as the one shown in the figure process the Doppler in a number of consecutive range bins by running a reference CW oscillator continuously as shown in the diagram.

14.5.4. Doppler Target Identification

Doppler signatures of moving targets can be used to differentiate between target types at long range. From a military perspective, this discriminatory capability is extremely useful, for example a high cost missile seeker needs to know whether the target it has detected is a high-value tank or helicopter and a low-value truck.



(a)



(b)

Figure 14.22: Idealised Doppler spectra of (a) moving tank and (b) helicopter that can be used to discriminate between different classes of moving targets

14.5.5. Pulsed Doppler Radar



- Pulsed Doppler radar systems such as the NEXRAD weather surveillance radar shown here operate using similar principles to the imaging Doppler ultrasound discussed above.
- The RF return signal in each range gate is down-converted to baseband using the I/Q principle described earlier in this section.

Figure 14.23: Doppler weather radar

The amplitude and phase of this echo pulse-train will vary with time as shown in (c) for $f_d < 1/\tau$) in Figure.14.11. The Doppler spectrum is extracted from the time waveform of each gate by processing blocks of data through a complex FFT algorithm. The frequency and phase information that is output from each range gate is converted to a radial velocity component from which actual wind velocity can be inferred from the variation in radial velocity of adjacent gates and an understanding of airflow dynamics.

Range and velocity, or range and reflectivity maps such as the one displayed below can then be produced by displaying the data for all range gates over the 360° scan angle of the radar.

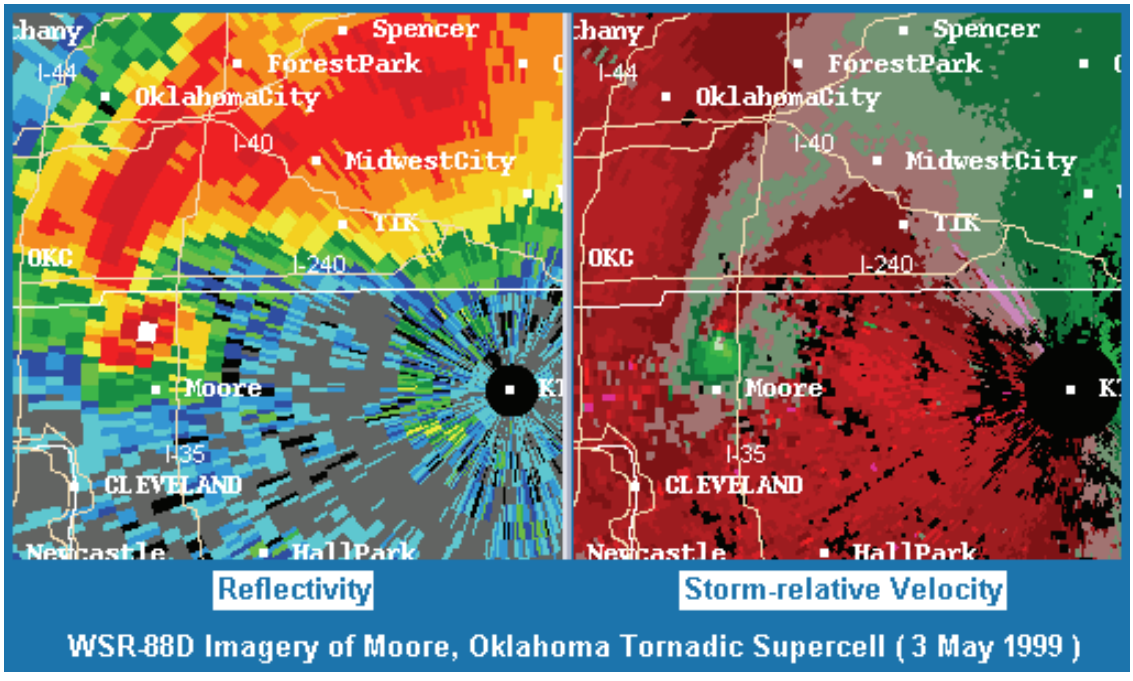


Figure 14.24: Pulsed Doppler radar images of a storm over Oklahoma

An integral and important part of RNDSUP (Radar Network and Doppler Services Upgrade) of the Australian Bureau of Meteorology is the implementation of a new

national radar training program for Australian forecasters. As the Doppler radars are implemented in each state, forecasters will undergo extensive radar training.

The image below is taken from the Doppler radar at Buckland Park. The air is moving in the directions shown by the arrows.

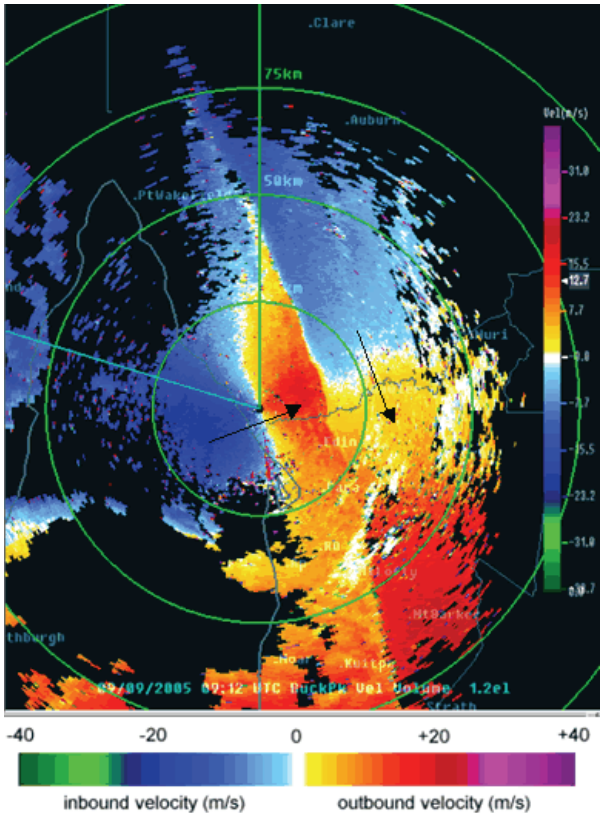


Figure 14.25: Doppler radar display from the Australian bureau of meteorology used to improve weather forecasting

14.6.References

[1] D.Liley, The Doppler Principle, http://liley.physics.swin.oz.au, 03/03/2001.
[2] Sports radar Gun, http://www.radarguns.com/specs/radar2000.html, 22/02/2000.
[3] Weather Surveillance Radar, http://www.mws.noaa.gov/modernize/88dtech.html, 23/02/2000.
[4] D.Barton, Radar Systems Analysis, Artech, 1976
[5] How Stuff Works: Doppler Ultrasound, http://howstuffworks.com, 4/4/2001.

Chapter 15.

Tracking Moving Targets

15.1. Track While Scan (TWS)

The track of a target in 2D can be determined from a surveillance radar Plan Position Indicator (PPI) display by plotting the target coordinates as they move when measured from scan to scan.

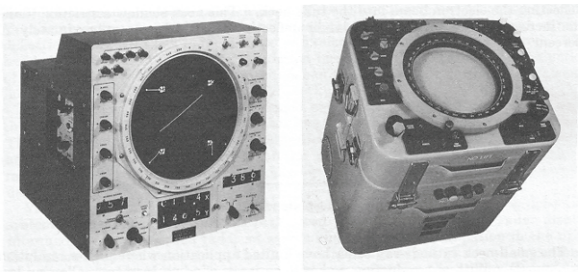


Figure 15.1: Examples of plan position indicator display hardware

At its most simple, this tracking function can be performed by a radar operator marking the face of the cathode ray tube with a pen. This is an inaccurate process and limits the number of targets that can be handled at one time.

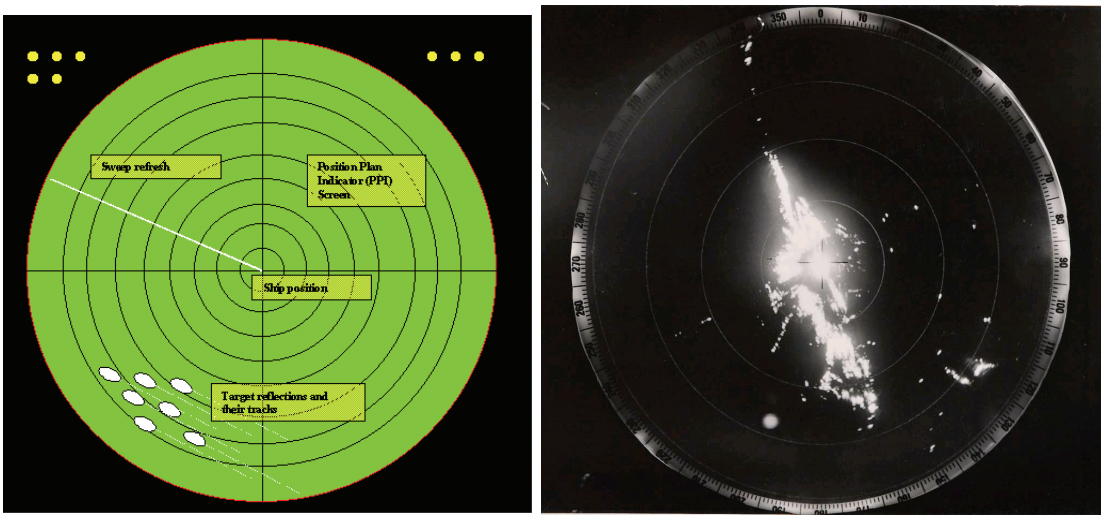


Figure 15.2: Plan position indicator showing a simulated and a real display

Automatic trackers operate as follows (for a single target)

1. Target is detected as the received echo exceeds a threshold. There is no information about its velocity. The software constrains the uncertainty to a reasonable value for an aircraft target (large circle)
2. Target is detected again displaced in range and angles, but within the uncertainty boundary. A crude velocity estimate is made and the position where the target will appear next is predicted. The uncertainty is still large as the position and velocity estimates are not good.
3. The target appears within this uncertainty boundary. Tracking filters estimates of position and velocity improve, and the next sample prediction is made with a smaller position uncertainty
4. As with (3)
5. The actual target position falls outside the position uncertainty boundary because it has accelerated, and the prediction algorithm only used position and velocity. Track is lost
6. A new target is detected with unknown velocity.

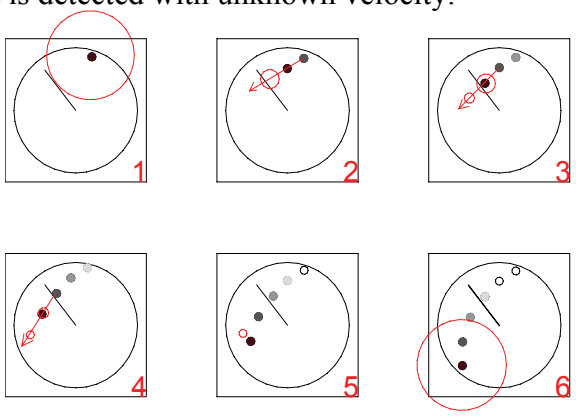


Figure 15.3: PPI Display sequence to illustrate the target tracking process

One method of performing the filtering and prediction function is to use αβ filters (see Chapter 13) for the polar co-ordinates (R,θ). An alternative would be to convert the measured positions from polar to Cartesian (x,y) before filtering as shown in the figure.

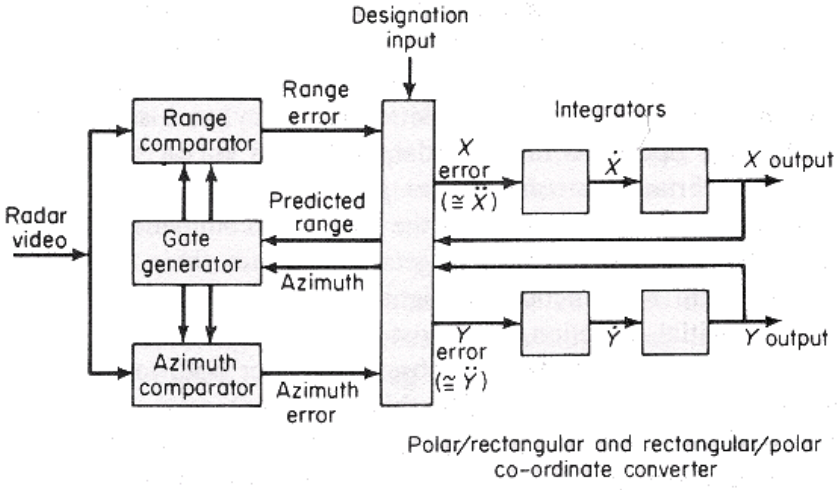


Figure 15.4: Track while scan processing

15.2. The Coherent Pulsed Tracking Radar

Pulsed tracking radar systems, such as the BAE Systems unit shown in the figure below, are mostly used for military applications such as fire control. These radars track fast moving aircraft or missiles with high accuracy, and then use the estimates of the target position and velocity to direct missiles or anti-aircraft guns.



Figure 15.5: Cluster of fire control tracking radars

Such coherent radar systems extract both amplitude and phase information from the signal reflected by a target. This is required because the length of a single pulse is too short to resolve typical target Doppler frequencies as shown in (c) of the figure below.

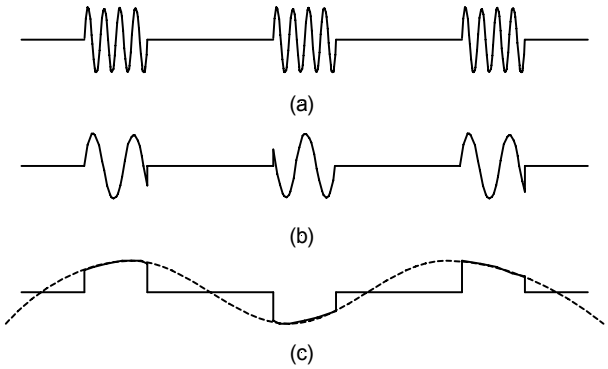


Figure 15.6: Doppler modulation of a pulse train

To extract Doppler shift, the returns from many pulses over an observation time T must be analysed so that the spectrum can be resolved down to a bandwidth $\beta \approx 1/T$. For this process to work, a deterministic phase relationship must be maintained over the observation time T.

15.2.1. Single Channel Detection

The block diagram of a single channel coherent radar is shown in the figure below.

The transmitter configuration is Master Oscillator Power Amplifier (MOPA)

- The transmitted signal is obtained by mixing an RF signal from a Stable Local Oscillator (STALO) with a frequency f_{rf} with a coherent oscillator (also very stable) with a frequency f_{if}.
- The resulting frequency (after filtering) id f_o is pulse modulated to form a pulse train, amplified and transmitted.

The received signal is down converted to IF (typically 30-60MHz) by mixing with the STALO signal. After amplification and filtering it is down converted further to baseband (video) by mixing with the COHO signal

A consequence of this single channel down conversion is that there is no direction information in any Doppler modulation since a target receding would produce exactly the same Doppler signature as one approaching and the response is

$$V_1 = k\sin\left(2\pi f_d t + \phi_o\right), \tag{18.1}$$

where V_1 – Video output voltage,
k – Amplitude of the video signal,
f_d – Doppler frequency (Hz),
ϕ_o –Phase shift (rad).

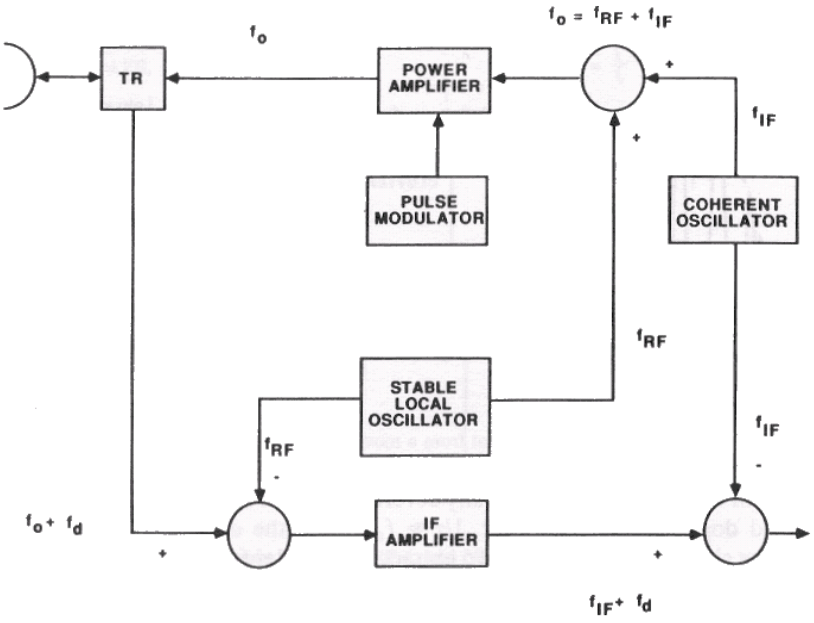


Figure 15.7: Single channel coherent pulsed Doppler radar

15.2.2. I/Q Detection

In an I/Q detector, the IF signal is split into two channels with the quadrature (Q) channel being phase shifted by 90° with respect to the in-phase (I) channel

Though the two Doppler signals output by the I and Q channels will have identical frequency whether the target is approaching or receding, their phase relationship with each other will reverse, and so direction information can be obtained as discussed in Chapter 13.

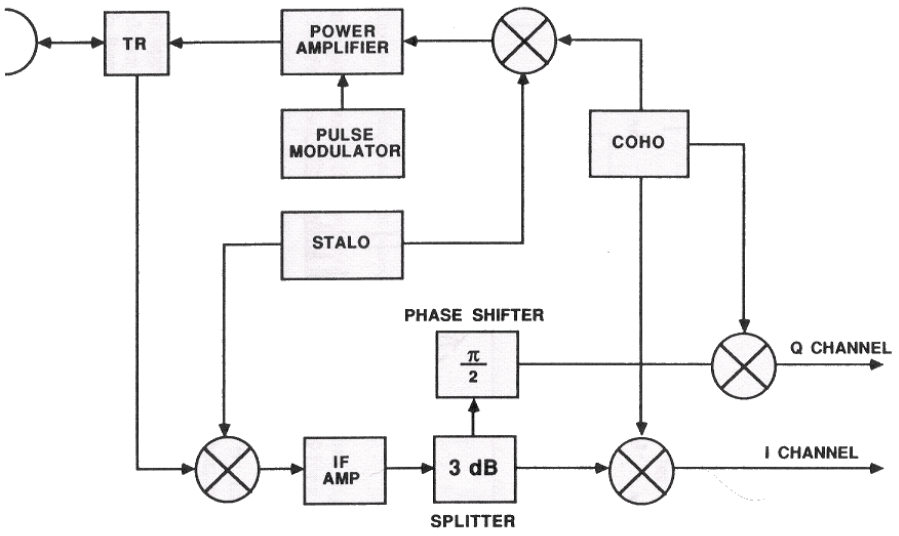


Figure 15.8: Coherent pulsed radar with a synchronous (i/q) detector

By sampling the I and Q outputs and using a complex fast Fourier Transform (FFT), the magnitude and phase of the combined Doppler spectrum can be obtained.

Another benefit of I/Q detection over single channel detection is that a 3dB gain in SNR is obtained at the output of the Doppler spectral analysis function.

If the actual Doppler frequency is not important, but only the fact that the target is moving at important, then a process called Moving Target Indication (MTI) can be applied.

15.2.3. Moving Target Indicator (MTI)

For a moving target. if the video output of the I or Q channel is examined, the amplitude will vary on a pulse to pulse basis, as shown earlier, due to the changing phase between the transmitted and received signals. For a static target, the phase will remain unchanged, and the amplitude will remain constant.

An MTI based on a delay-line canceller operates by taking the difference of the amplitudes of successive pulses as shown below.

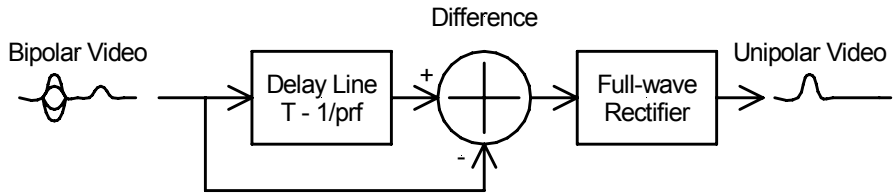


Figure 15.9: Delay line canceller used to suppress the returns from static targets

This is in effect a Finite Impulse Response (FIR) filter with a high-pass characteristic that rejects signals that are unchanging or that are changing very slowly. However, because it becomes in effect a sampled data system, the frequency response is repeated with a period of $1/T$ as shown in the figure below.

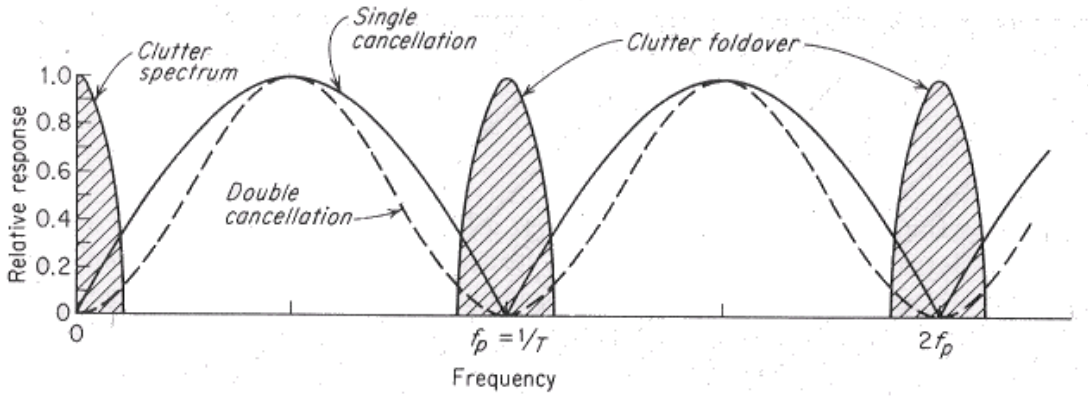


Figure 15.10: Transfer function of single and double delay line cancellers for MTI

Standard FIR filter techniques can be used to adjust the weightings of a canonical filter implementation, as shown below, to alter the filter response and narrow the stop band and flatten the pass band.

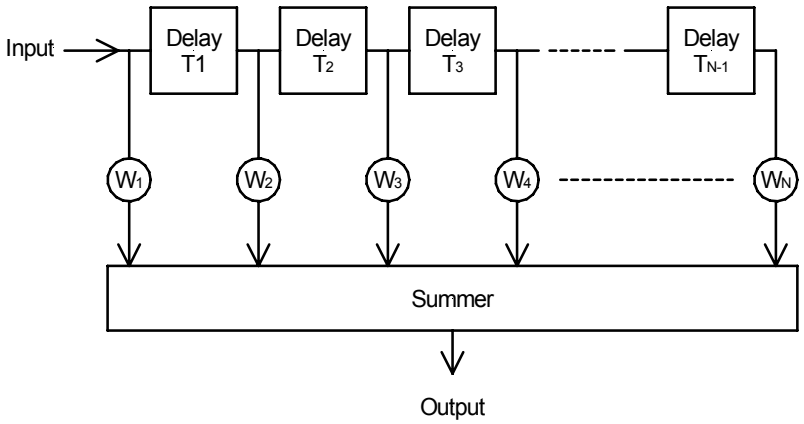


Figure 15.11: Canonical MTI filter structure

One of the disadvantages of classical MTI techniques is that they cannot discriminate between a moving target and moving clutter, nor do they cope very well if the radar is also moving.

Blind Speeds

If the target Doppler frequency lies in the region where $f_d \approx 1/T$, it can be seen, from the transfer function above, that it is attenuated. Zeros also occur at $2/T$, $3/T\ldots$ n/T

$$f_d = \frac{n}{T} = nf_p, \tag{18.2}$$

where f_d – Doppler frequency (Hz),
f_p – Pulse repetition frequency (PRF) (Hz).

The actual blind speeds are given by the following

$$v_n = \frac{n\lambda}{2T} = \frac{n\lambda f_p}{2} \text{ m/s}, \tag{18.3}$$

where v_n – Blind speed (m/s),
λ - Transmitter wavelength (m),
n – 1,2,3,…

To obtain a high first blind speed, it is necessary to either operate with a long wavelength (often not practical), or to operate with a high PRF. However, a high PRF becomes ambiguous at a short range which is also not ideal for surveillance or long range tracking radar applications.

Staggered PRF and Blind Speed

The effect of blind speeds can be reduced by operating at more than one PRF as shown below.

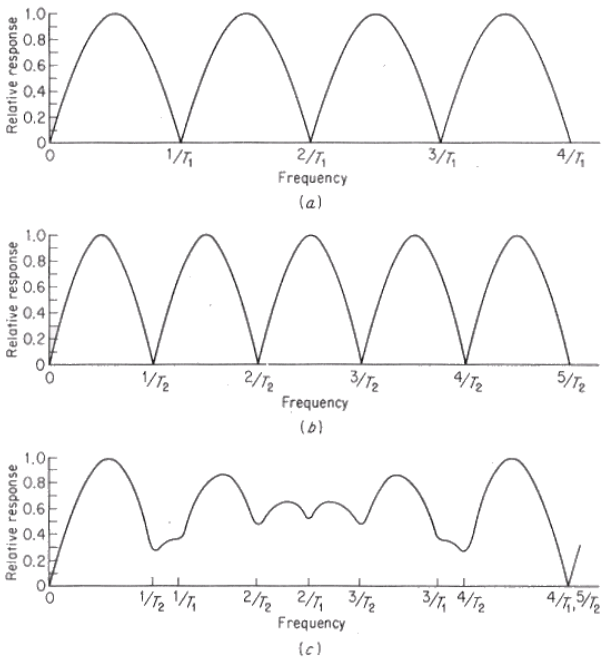


Figure 15.12: Effect of staggered PRF on MTI transfer function

As the ratio of the pulse repetition interval, PRI, T_1/T_2 approaches unity, the greater will be the value of the first blind speed. However, the first null also gets deeper and so the rejection of slowly moving clutter will be compromised.

To cater for the moving clutter problem, a bank of Doppler filters can be implemented instead of a delay line canceller. In modern radars, this process is generally implemented digitally using the Complex FFT. The outputs (excluding sidelobes) of such a filter implementation are shown below.

To obtain sufficient rejection of unwanted signals in adjacent bins, the filter sidelobes must be made as low as possible.

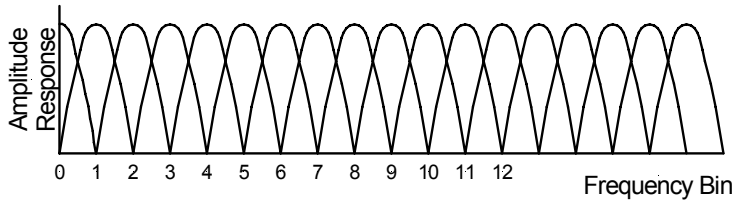


Figure 15.13: Filter bank implemented using the FFT

A tradeoff exists between the width of the mainlobe and the sidelobe level with different windowing (weighting) functions as discussed in Chapter 11.

Table 15.1: Properties of weighting functions

Weighting Function	Peak Sidelobe Level (dB)	MainLobe Width
Rectangular	-13.26	0.886
Hanning	-31.5	1.42
Hamming	-42.5	1.32
Taylor n=5	-34	1.19
Taylor n=6	-40	1.25
Dolph Chebyshev	-40	1.2

15.3.Limitations to MTI Performance

Sub-clutter visibility is the ratio by which the moving-target power may be lower than the clutter in the same range bin, and still be detected with a specified P_d and P_{fa}.

It is usually limited by internal instabilities in the amplitude and phase of the various waveforms generated by the radar, by the motion of a scanning antenna and the finite time on target which has the effect of widening the clutter spectrum and also by the bandwidth characteristics of real clutter. Rain is blown by the wind, the sea moves, as do leaves and grass.

15.4.Range Gated Pulsed Doppler Tracking

The Doppler tracking loop and the range tracking loop are interrelated. The Doppler tracking loop operates on video sampled at the range indicated by the range tracker while range tracking is usually accomplished by means of the split (early/late) gate technique.

Before the signals in the early and late gates are compared to derive the range tracking error, each is passed through a Doppler tracking filter to reject returns from stationary objects at the same range.

The range tracking loop can be updated at the PRF unless an FFT that processes N_p pulses is used to perform the Doppler analysis in which case the update frequency is reduced to PRF/N_p.

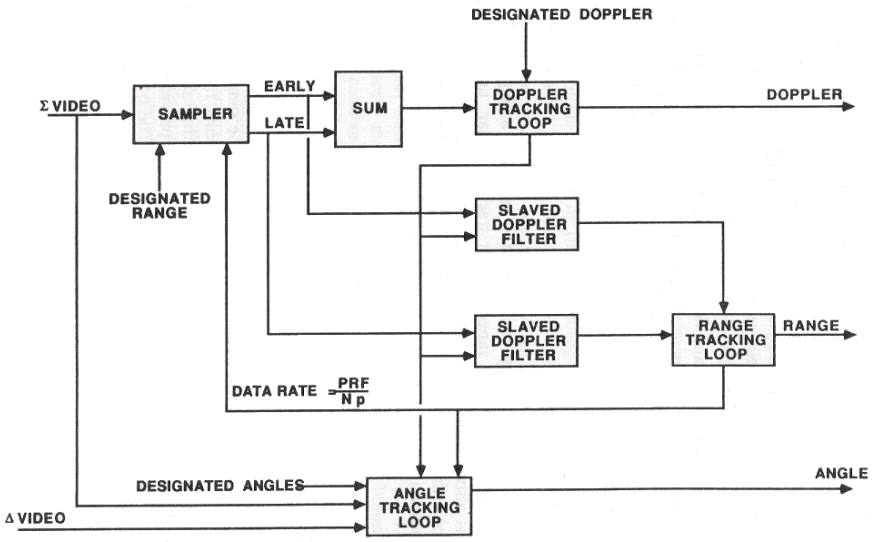


Figure 15.14: Schematic diagram of a range gated doppler tracker

Initiation of tracking requires the near simultaneous initialisation of all four tracking coordinates: range, two angles and Doppler frequency, this is achieved as follows.

- The operator who identifies a target during the search phase designates the selected target on the display using a cursor.
- On the following scan, if the target is still present, the range, azimuth and elevation angles and Doppler frequency are recorded.
- The scan sequence is interrupted and the radar antenna returns automatically to the designated area.
- The range gate pair are moved to the designated range where they begin a small search to compensate for uncertainties in the designation.
- The sum channel video received by the combined early and late gates, or by a third target gate that straddles the two is processed by a Doppler analyser (eg FFT)
- If a moving target is found at the designated range and sufficiently close to the Doppler of the detection that triggered the acquisition sequence, then the Doppler tracking gate is placed over the target.
- The Doppler loop is closed by using a frequency discriminator to maintain the target in the centre of the tracking filter.
- Video signals in both the early and late gates are analysed by Doppler filters slaved to the frequency of the Doppler tracker.
- After allowing the Doppler loop time to settle, the range tracking loop is closed by using the difference in the relative amplitudes of the early and late gates to control the range position

- The video signals in bot the azimuth and elevation monopulse angle error channels are sampled by a range gate slaved to the target gate of the range tracker.
- The sampled video for each of the two error signals is filtered by a Doppler filter slaved to the frequency of the Doppler tracker. Thus angle errors are derived only from the same range-Doppler cell that contains the target.
- After a brief period to allow the range tracking loop to settle, the angle tracking loop is closed.

The transition from search to track is not a trivial problem, and it is made even more difficult in the military scenario, if the target either starts to manoeuvre or tries to disrupt the process by deploying chaff or by some electronic means

15.5.Co-ordinate Frames

15.5.1. Measurement Frame

Radar measurements are made in polar space (R,θ,ϕ) as the radar can only measure range, elevation and bearing (azimuth).

15.5.2. Tracking and Estimation Frame

The equations of motion that govern the profile of a target operate in Cartesian space, (x,y,z), so it is advantageous to transform the co-ordinate system from polar to Cartesian space.

A point P can be located by spherical coordinates (r, θ, ϕ) as well as rectangular coordinates (x, y, z).

The transformation between those coordinates is

$$\begin{cases} x = r \sin\theta \cos\phi \\ y = r \sin\theta \sin\phi \\ z = r \cos\theta \end{cases}$$

or

$$\begin{cases} r = \sqrt{x^2 + y^2 + z^2} \\ \phi = \tan^{-1}(y/x) \\ \theta = \cos^{-1}(z/\sqrt{x^2 + y^2 + z^2}) \end{cases}$$

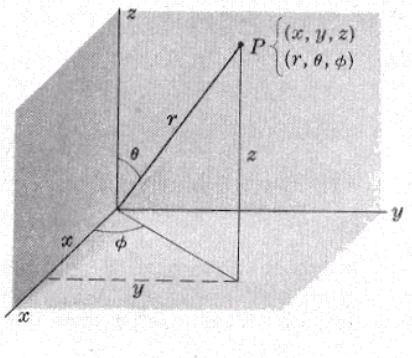


Figure 15.15: Rectangular <–> Polar (spherical) transfer function

Generally this frame of reference will remain centred at the radar, however, some fire control systems translate and rotate the frame to make it target centred as the aircraft dynamics can be better modelled in this frame.

If more than one sensor may be involved in the tracking function, and these sensors are not co-located (they may be on different platforms that move relative to each other), then an earth centred Cartesian frame is generally used.

15.6.Antenna Mounts and Servo Systems

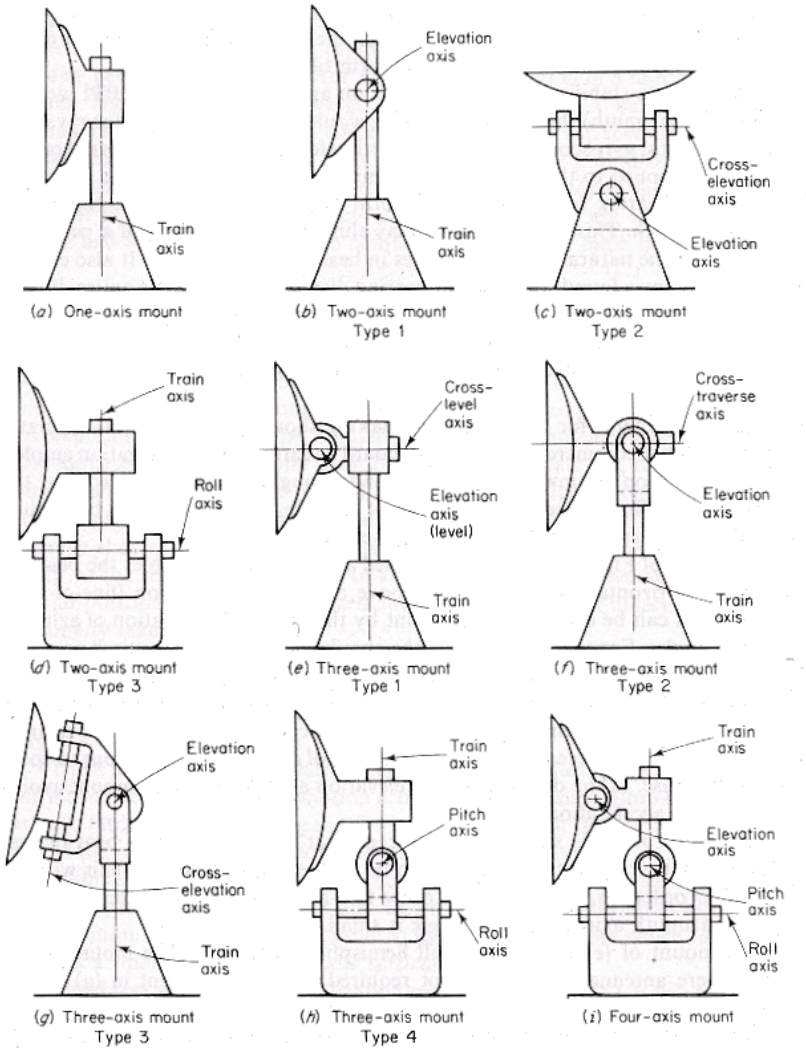


Figure 15.16: Antenna mounts

The pencil beam of a tracking radar must be pointed at the target for tracking to occur. This is quite a challenge as a typical tracking radar has a 3dB beamwidth between 1° and 2°, or smaller in the case of the radar for a close-in weapon system onboard ship.

A servo system is used to drive the antenna in the direction that minimises the tracking errors. Most servo systems are Type II, or zero velocity error systems since, in theory, no steady-state error exists for a constant velocity (angular rate) input. With Type II systems, dynamic lags proportional to the magnitude of the target acceleration do occur. To accommodate this, the tracking bandwidth is adjusted to minimise the tracking error which is due to a combination of measurement noise and dynamic lag.

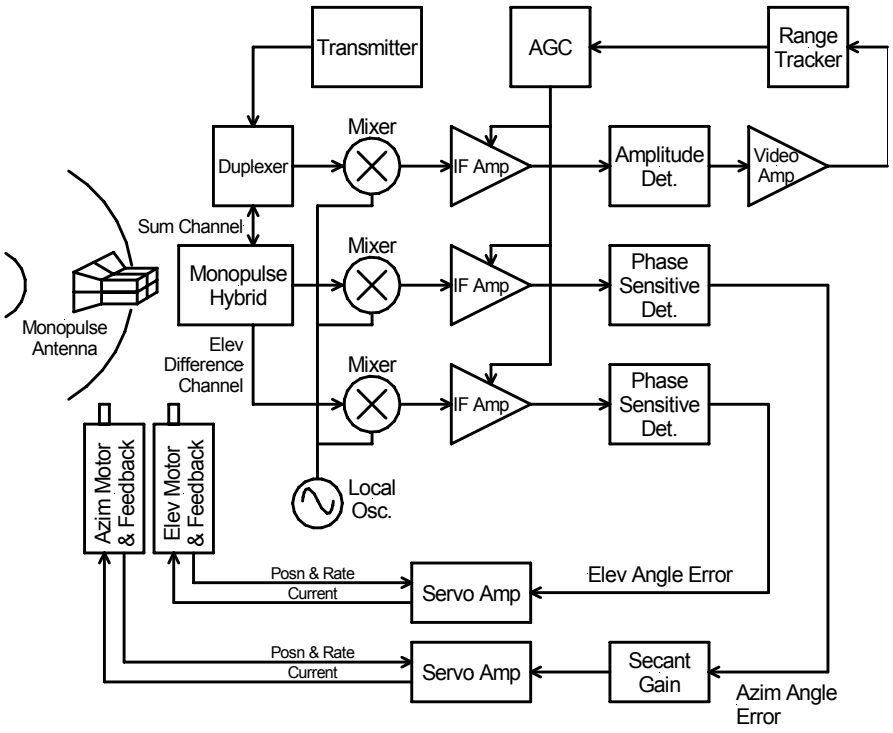


Figure 15.17: Monopulse tracking radar

At long range where the angular motion of the target is small, a very small tracking bandwidth can be tolerated. However, at short range where target angular rates and accelerations are large, a wider bandwidth becomes acceptable. Secant correction increases the azimuth error signal gain as a function of the elevation angle.

Another restriction on tracking bandwidth is that it must be small (10%) compared to the lowest natural resonant frequency of the antenna and mounting structure to reduce the risk of instabilities occurring.

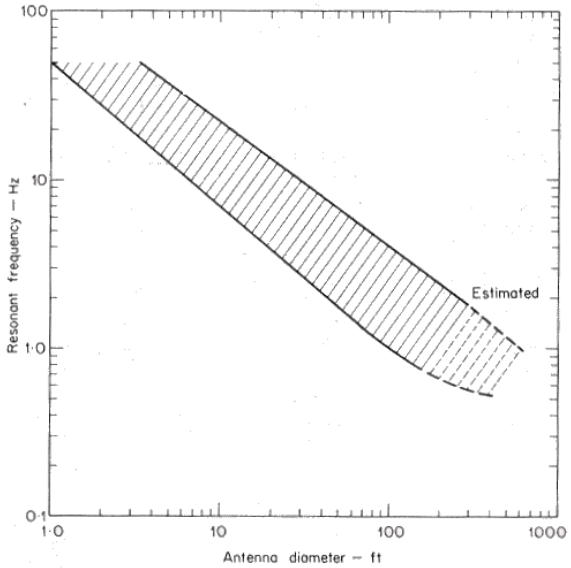


Figure 15.18: Lowest resonant frequency as a function of antenna diameter

The pointing system, generally known as a pedestal must be able to accommodate both target motion as discussed and its own base motion (if mounted on a moving vehicle)

There are many different types of antenna mounts that can be used depending on the tracking and stabilisation requirements.

15.6.1. On-Axis Tracking

The best tracking occurs using null-steering when the antenna is pointed towards the target with an accuracy of only a few milliradians. This is known as on-axis tracking.

It reduces cross coupling between the axes by minimising cross polar levels and reducing the effects of system nonlinearities.

It requires the following:

- the removal by prior calibration of biases
- a filter than can perform one sample ahead prediction
- the selection of the appropriate co-ordinate system for tracking.

Target dynamics that dictate the real and apparent acceleration and tracking loop bandwidth determines the tracking accuracy.

15.6.2. Crossing Targets and Apparent Acceleration

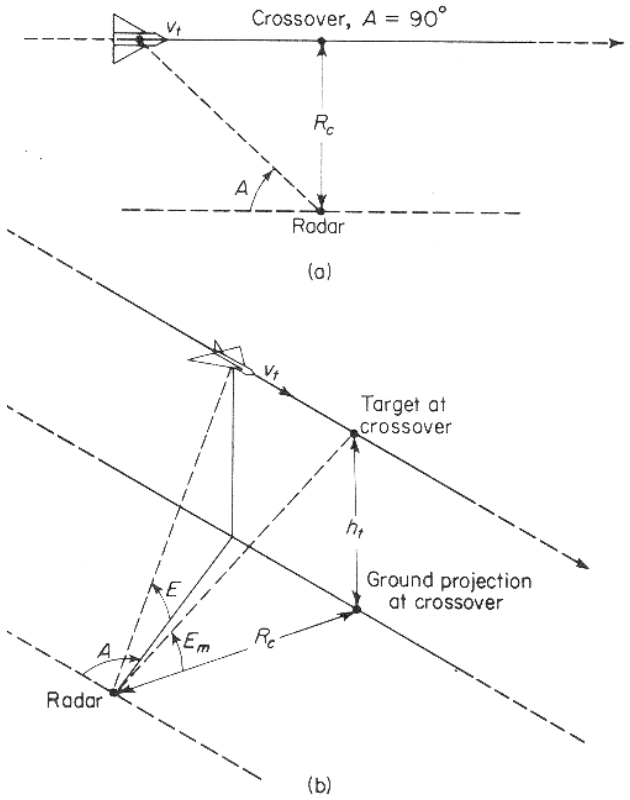


Figure 15.19: Geometry of a crossing target (a) in plan and (b) in perspective

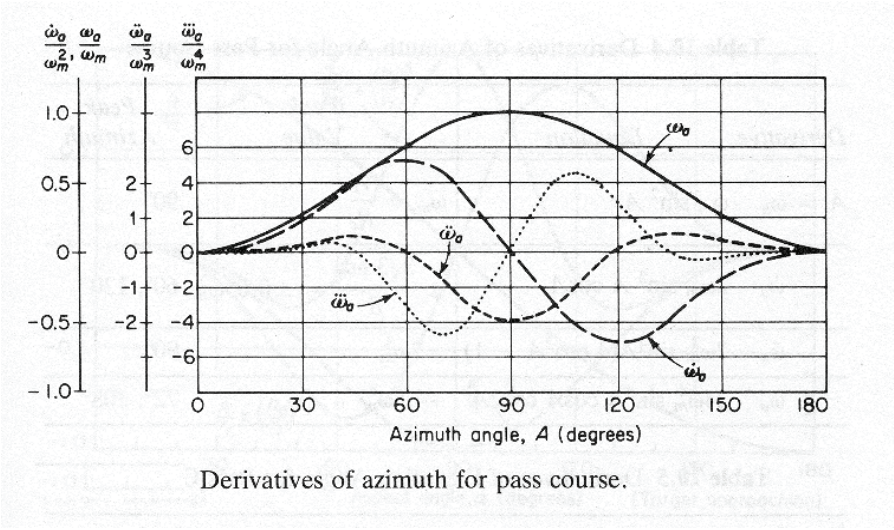
When viewed in radar polar co-ordinates, the target velocities will be

$$\dot{R}_{\max} = v_t, \dot{A}_{\max} = \frac{v_t}{R}, \dot{E}_{\max} = \frac{v_t}{R}. \tag{18.4}$$

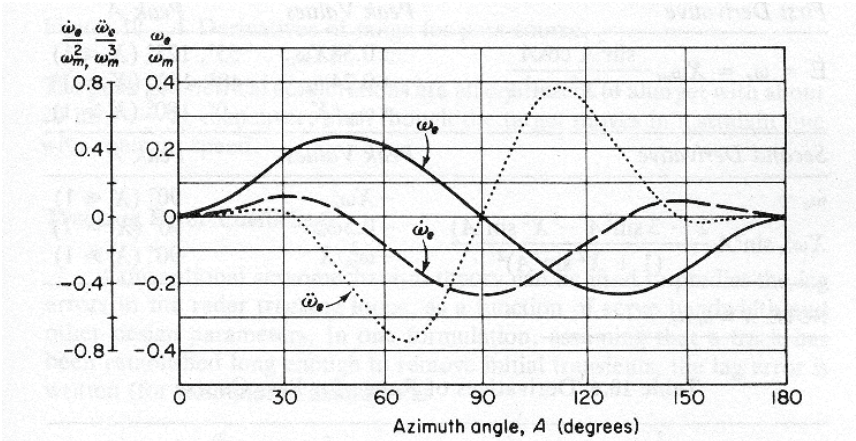
The real accelerations are

$$\ddot{R}_{\max} = a_t, \ddot{A}_{\max} = \frac{a_t}{R}, \ddot{E}_{\max} = \frac{a_t}{R}. \tag{18.5}$$

The geometric accelerations and other derivatives are shown below. They are normalised in the angular co-ordinates to $\dot{A}_{\max} = \varpi_m = v_t / R_c$. Where R_c is the ground range at crossover.



Derivatives of azimuth for pass course.



Derivatives of elevation angle for pass course ($X = 1$).

Figure 15.20: Angular derivatives for crossing targets

In range, the derivatives are normalised to $R_a = R_{min}$, at the point of closest approach.

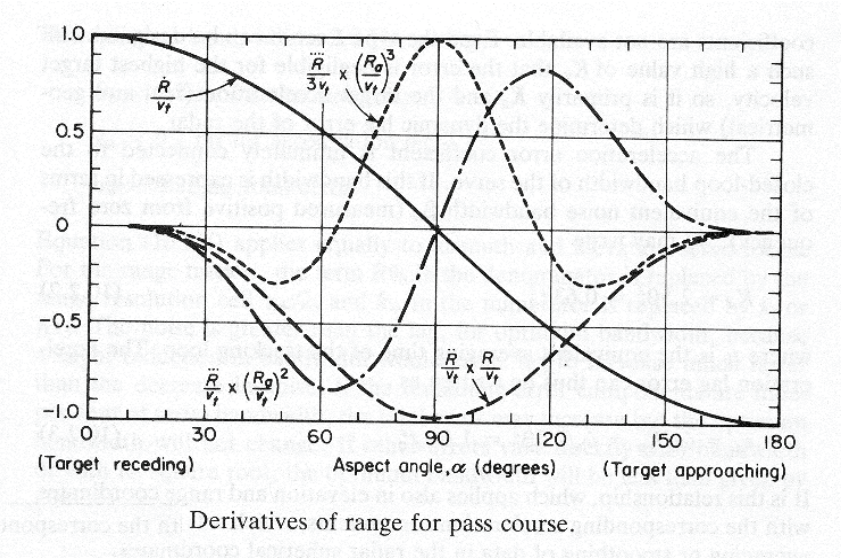


Figure 15.21: Range derivatives for crossing targets

Conventional servomechanism theory can be used to determine the lag errors in radar tracking loops if conventional loops are implemented.

The lag error can be written as follows

$$\delta_a = \frac{\varpi_a}{K_v} + \frac{\dot{\varpi}_a}{K_a} + \frac{\ddot{\varpi}_a}{K_3} + \ldots \tag{18.6}$$

Where the coefficients K_v, K_a and K_3 are the servo error coefficients, the values of which increase with increasing loop gain and bandwidth.

Servos are classified according to the first coefficient that is finite in the loop design. A type 1 servo has a finite K_v (but infinite position error constant K_o), a type 2 has finite K_a but infinite K_v, etc.

The acceleration error coefficient is intimately connected to the closed loop bandwidth of the servo. If the bandwidth is expressed in terms of the equivalent noise bandwidth B_n then

$$K_a = 2.5B_n^2 = 0.63/t_o^2. \tag{18.7}$$

Where t_o is the equivalent averaging time for the tracking loop. The lag will be.

$$\varepsilon_a = \frac{\dot{\varpi}_a}{K_a} = \frac{\dot{\varpi}_a}{2.5B_n^2} = 1.6\dot{\varpi}_a t_o^2. \tag{18.8}$$

This formula is applicable to azimuth, elevation and tracking lags with their corresponding acceleration components.

If thermal noise and dynamic lag are the primary sources of error, then an optimum bandwidth can be obtained to minimise the total error variance. According to Barton, this is

$$\sigma_\theta^2 = \frac{\theta_3^2 B_n}{k_m^2 f_r B\tau (S/N)} + \frac{a_t^2}{6.3 R^2 B_n^2}, \tag{18.9}$$

$$B_n = \left[\frac{a_t^2 k_m^2 f_r B\tau (S/N)}{1.6\theta_3^2 R^2} \right]^{1/5}, \tag{18.10}$$

where: a_t – acceleration (real or geometric) (m/s^2),
k_m – Monopulse Gain Constant (typ 1.6),
f_r – Pulse repetition frequency (Hz),
$B\tau$ - IF bandwidth and pulse width (see matched filter),
S/N – Single pulse signal to noise ratio,
θ_3 – Antenna beamwidth (rad),
R – Crossing Range (m).

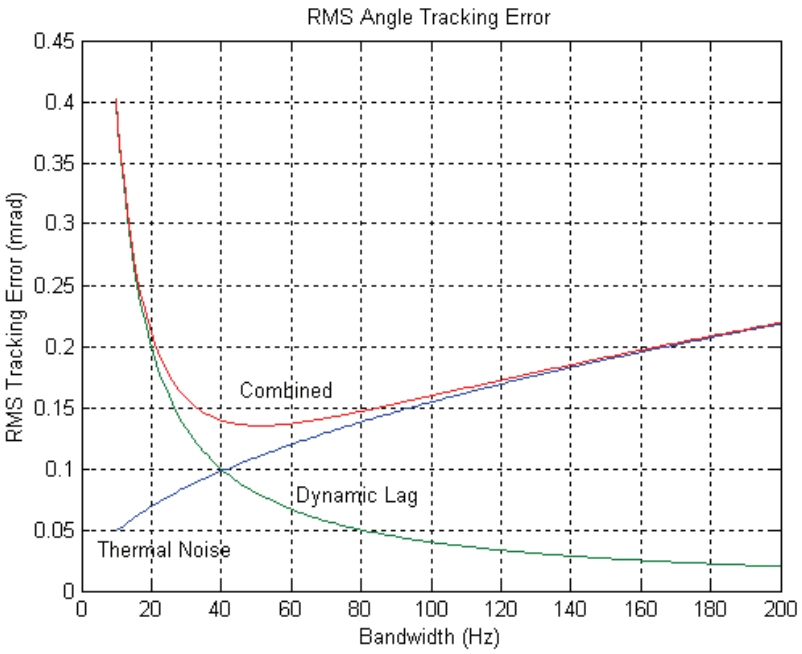


Figure 15.22: Typical noise and lag optimisation

15.6.3. Tracking in Cartesian Space

One method of maintaining the noise performance of the system while minimising the dynamic lag is to operate with a wide angle-servo bandwidth, and to perform the tracking and smoothing in Cartesian space.

As there are no geometric accelerations in Cartesian space, it is possible to reduce the filter bandwidth to less than 1Hz. This determines the noise performance of the tracker.

Because the angle servo bandwidth is wide (typically 10 to 100Hz for a real system), the dynamic lag for geometric accelerations is limited.

A simplified block diagram showing the loop configuration is shown below.

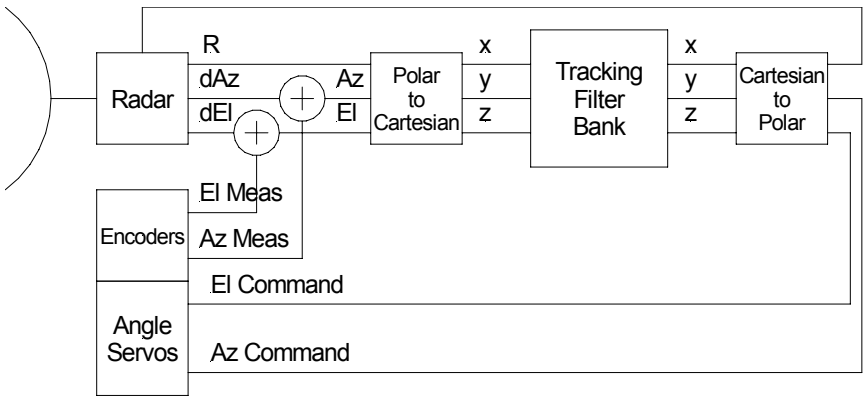


Figure 15.23: Tracking in Cartesian space

15.7.Fire Control Radar Design

15.7.1. Requirements

Ships Motion

- Radar on the deck of a ship 8m above the water
- Roll +/-25° Pitch +/- 10° Yaw +/-5°
- Period about 5s

Designation

- From a surveillance radar (one time per second)
- Elevation accuracy +/-5°, Azimuth accuracy +/-2°
- Range accuracy +/-25m
- No velocity information

Environment

- Up to sea state 5 (very rough, wave height 2.4 to 3.6m)
- Rainfall up to 25mm/hr

Target Types

- Fixed Wing Aircraft (frontal RCS $1m^2$ independent of frequency)
- Aircraft Height > 60m
- Sea Skimming Missiles (frontal RCS $0.1m^2$ independent of frequency)
- Sea Skimmer height 3m

Detection Performance

- Probability of detection $P_d = 0.95$
- Probability of false alarm $P_{fa} = 10^{-6}$
- Detection time from receipt of designation (excluding slew) 0.5s
- Detection Range

Table 15.2: Tracking performance

Weather	Aircraft	Sea Skimmer
Clear Air	15km	6km
Rain 12.5mm/hr	10km	4km
Rain 25mm/hr	5km	2km

Tracking Performance

- Minimum tracking range $R_{min} = 50m$
- Range tracking accuracy (<1m RMS)
- Angle tracking accuracy (<1mrad RMS)
- Tracking aircraft directly overhead at h>100m
- Aircraft and missile velocity <280m/s

Safety Constraints

- Average Transmit Power <100W

15.7.2. Selection of Polarisation

For low flying aircraft over the sea we want to minimise the sea clutter. In calm seas at low grazing angles the difference between the reflectivity for horizontal polarisation and vertical polarisation exceeds 12dB. As the roughness increases the difference decreases.

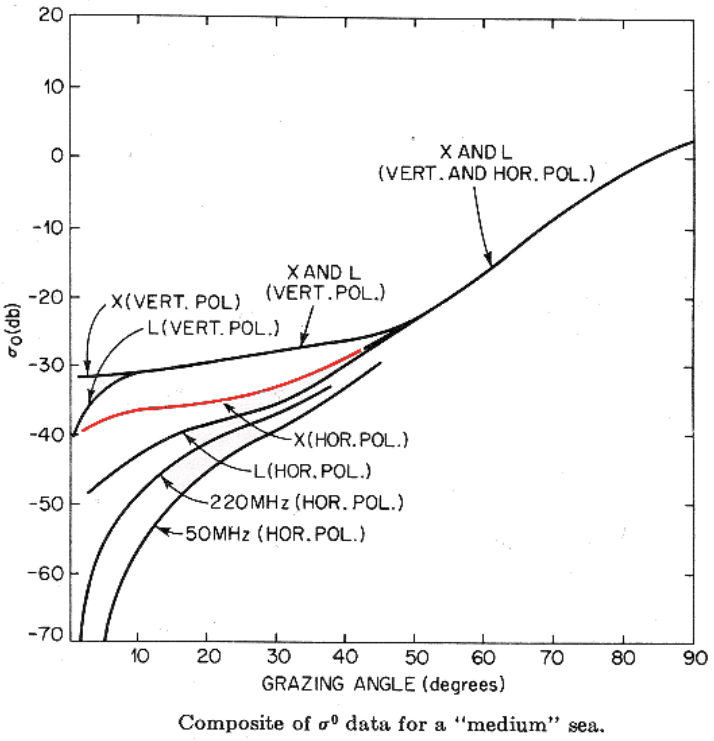


Figure 15.24: Sea Clutter reflectivity as a function of grazing angle

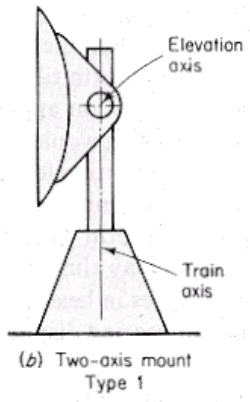
The radar will operate using horizontal polarisation, and it will be assumed that the surface reflectivity σ° is –35dB.

15.7.3. Pedestal Specifications

The maximum combined roll and pitch angle is

$$\varphi_{\max} < \sqrt{25^2 + 5^2} \ .$$

This does not exceed 30°



(b) Two-axis mount Type 1

The antenna can be stabilised with respect to this tilt adequately without resorting to a 3rd axis, however, it can result in a significant rotation of the polarisation

The pedestal will be of the type elevation, θ, over azimuth, ϕ.

θ defined as +ve up from the horizontal

ϕ defined as +ve anticlockwise from the x-axis

For a mounting height ≈ 8m above the sea with the minimum tracking angle for a sea skimmer at R = 50m (θ = -6°) and a combined roll and pitch angle of 30° requires a minimum angle of -36° (θ_{min} = -40°).

To allow the antenna to track over the vertical, and to have time to slew around in azimuth without losing lock at the maximum combined roll and pitch angle requires ϕ_{max} = 90+30 = 120° (use 125°)

15.7.4. Radar Horizon

For h_r = 8m and h_t = 60m, the radar horizon is given by

$$d = 130(\sqrt{h_r(km)} + \sqrt{h_t(km)} = 43km$$

The radar horizon is not a consideration for this design

15.7.5. Selection of Frequency

A reasonable maximum diameter for the antenna on a shipboard radar is 1.5m.

The beamwidth at various frequencies will be as follows:

Table 15.3: Beamwidth for different frequencies

Frequency (GHz)	Band	Beamwidth (deg)
10	X	1.4
35	Ka	0.4
94	W	0.13

A narrow beam decreases the effect of multipath and limits the area of clutter within the tracking gate, while attenuation increases with frequency (particularly in the rain)

The minimum elevation angle when tracking at target at h = 60m and R = 15km θ = 0.19° and the minimum tracking angle when tracking a sea skimmer at h = 3m and R = 50m θ = -6°

Though it may be possible to minimise the effects of multipath, it is not possible to eliminate them when the radar is looking down at the target.

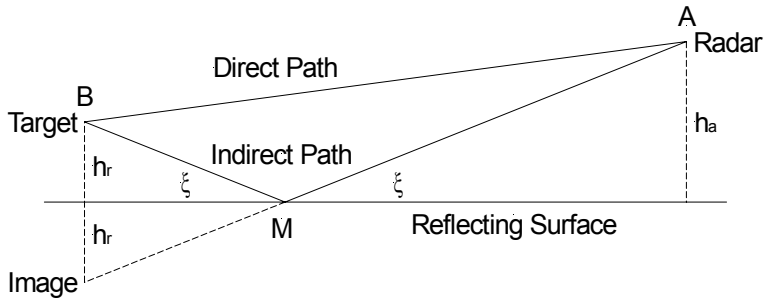
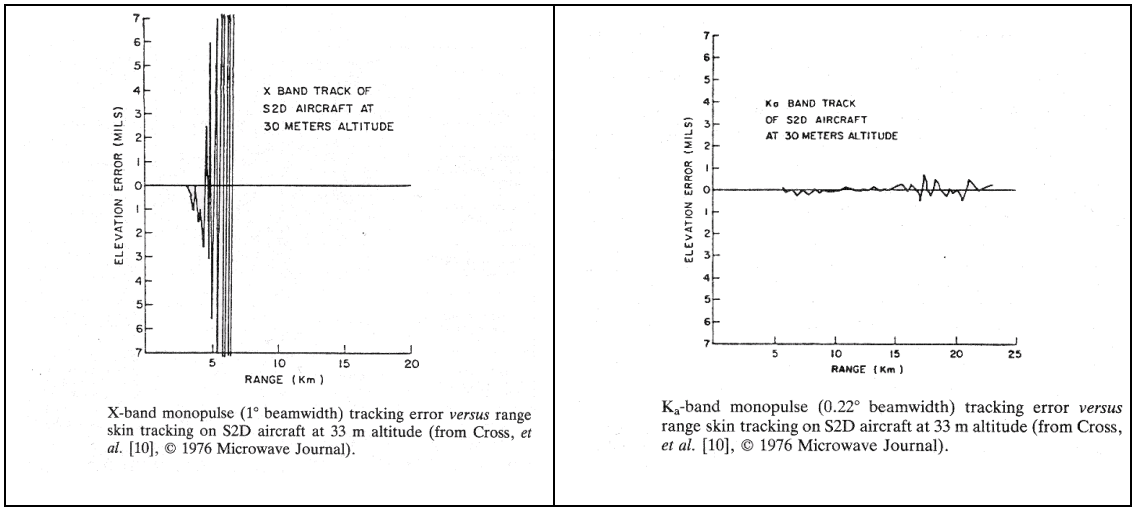


Figure 15.25: Multipath propagation

Multipath has a major effect on tracking accuracy as shown by the following measured data for two systems tracking the same target.



X-band monopulse (1° beamwidth) tracking error *versus* range skin tracking on S2D aircraft at 33 m altitude (from Cross, *et al.* [10], © 1976 Microwave Journal).

K_a-band monopulse (0.22° beamwidth) tracking error *versus* range skin tracking on S2D aircraft at 33 m altitude (from Cross, *et al.* [10], © 1976 Microwave Journal).

Figure 15.26: Effect of beamwidth on low angle tracking accuracy

The best compromise is to use the narrowest beamwidth possible and to use multipath reduction techniques. According to the literature application of these techniques can maintain an RMS elevation tracking accuracy of between 0.05 and 0.1 beamwidths

Assuming that we can only manage 0.1 beamwidths, then the tracking accuracy will be:

Table 15.4: Tracking accuracy with multipath

Band	Beamwidth (deg)	RMS Tracking (mrad)
X	1.4	2.4
Ka	0.4	0.69
W	0.13	0.22

This excludes the X-band option as the tracking accuracy does not meet the accuracy criteria defined earlier.

15.7.6. Adverse Weather Effects

Typical attenuation as a function of frequency under different weather conditions is shown below:

Table 15.5: Atmospheric attenuation

Band	Clear (dB/km)	Rain 12.5mm/h (dB/km)	Rain 25mm/h (dB/km)
X	0.02	0.25	1
Ka	0.15	3	7
W	0.3	7	12

15.7.7. Required Single Pulse Signal to Noise Ratio

The required S/N to achieve the specified P_d and P_{fa} is a function of the target distribution and its fluctuation characteristics. For a non-fluctuating target this is 13.6dB.

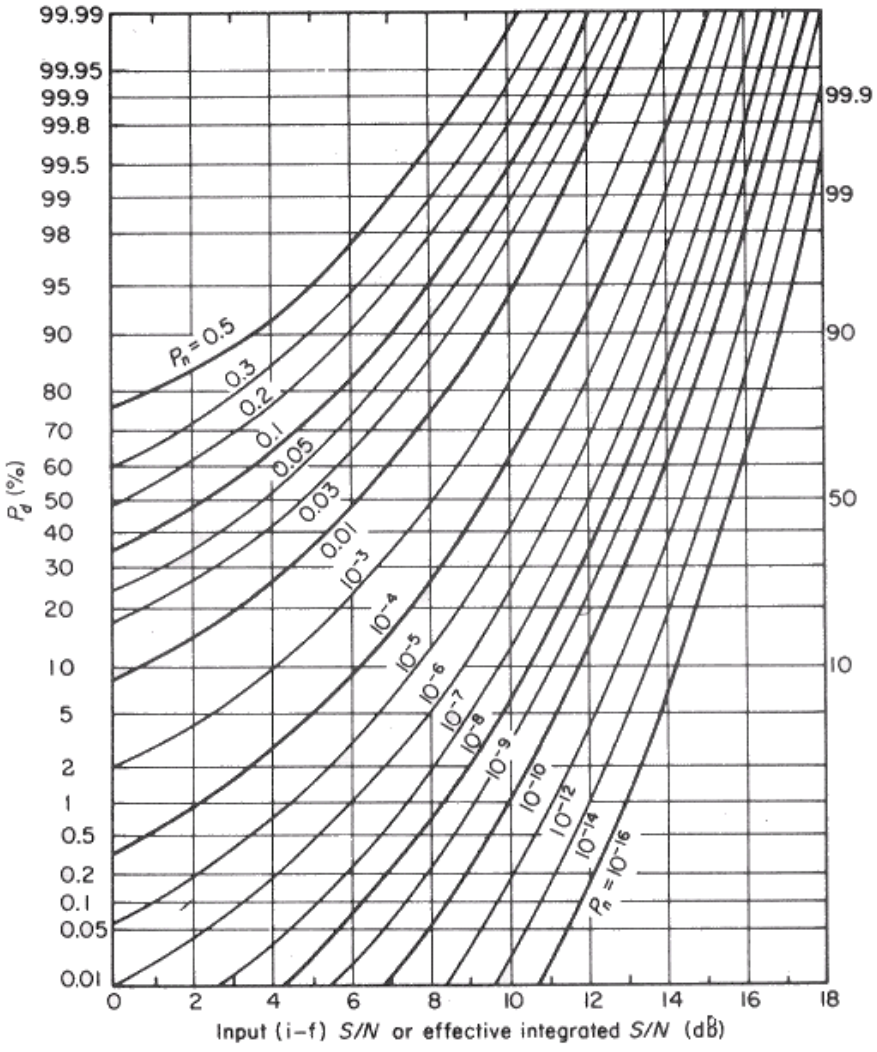
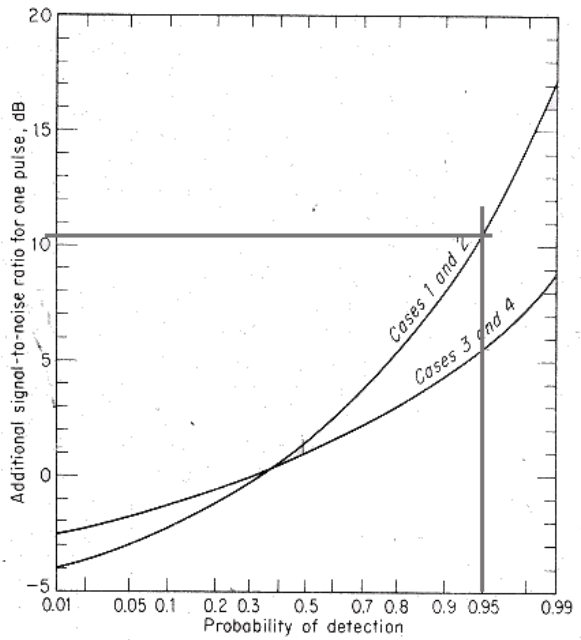


Figure 15.27: Detection probability curves

Swerling 1: Many independent scatters of similar RCS. This results in slow fluctuations with time.

Swerling 2: One major scatterer and many smaller scatterers. This is typical of an aircraft nose on. This results in fast fluctuations with time.



Additional signal-to-noise ratio required to achieve a particular probability of detection, when the target cross section fluctuates, as compared with a nonfluctuating target; single hit, $n = 1$.

Figure 15.28: Fluctuating target effects

The additional SNR required to achieve these probabilities of the target is fluctuating is determined from this graph. It will be 10.4dB. The single pulse SNR required is thus 13.6+10.4 = 24.0dB.

15.7.8. Tracking Gate Size

Ideally, the tracking gate size is matched to the target length

A typical Jet fighter would be 15m long, so a gate size between 15 and 20m would be ideal. We will use 20m.

This minimises the amount of clutter received, it does however complicate the target acquisition process as the designation accuracy is only +/-25m, so at least 3 gates would be required to span the uncertainty

15.7.9. Signal to Clutter

We assume that the reflectivity of the sea is the same at X, Ka and W bands

We assume that a similar SCR is required as SNR

For a gate size of 20m, the illuminated area will be a function of the antenna beamwidth

Table 15.6: Illuminated sea area as a function of frequency

Band	Beamwidth (deg)	Area (6km) (m^2)	Area (15km) (m^2)
X	1.4	2928	7320
Ka	0.4	840	2100
W	0.13	264	660

For an average reflectivity σ°= -35dB in all cases, the clutter RCS will be

Table 15.7: Clutter RCS as a function of frequency

Band	RCS (6km) (dBm^2)	RCS (15km) (dBm^2)
X	-0.33	3.6
Ka	-5.8	-1.8
W	-10.8	-6.8

For an aircraft RCS of $1m^2$ ($0dBm^2$) or a sea skimmer with an RCS of $0.1m^2$ ($-10dBm^2$), we are looking for an extra 20 to 30dB of signal to achieve the required SCR of 24dB (for the required single pulse detection probability).

Integration improvement cannot be used for improving the SCR because, unlike the white thermal noise used for target detection, clutter is correlated, so integration will not be as effective.

The primary difference between the targets and the clutter is that generally the target has a significant radial velocity and the clutter is static (or slow moving in high seas).

15.7.10. Moving Target Indicator

Some form of MTI will have to be implemented that can achieve a sub-clutter visibility as follows

Table 15.8: Subclutter visibility requirements sea skimmer

Band	Clutter RCS At 6km (dBm^2)	Sea skimmer RCS (dBm^2)	Subclutter Visibility For SCR=24dB
X	-0.33	-10	33.67
Ka	-5.8	-10	28.2
W	-10.8	-10	23.2

Table 15.9: Subclutter visibility requirements aircraft

Band	Clutter RCS At 15km (dBm^2)	Aircraft RCS (dBm^2)	Subclutter Visibility For SCR=24dB
X	3.6	0	27.6
Ka	-1.8	0	22.2
W	-6.8	0	17.2

This can be achieved using a delay-line canceller, however because the ship will be moving, the clutter will also have an effective velocity, so this is not a good technique to use.

A better alternative is to take more samples, to window them and to use an FFT to isolate moving from "static" targets.

With a Hamming window, a static return rejection of 42dB can be obtained. However, it is unlikely that if the measurement is unambiguous in range, that it will be unambiguous in velocity.

15.7.11. The Pulse Repetition Frequency

To be unambiguous in range out to 15km, the maximum allowed PRF is calculated as

PRF = $c/2R_{max}$ = 10kHz

Using the Nyquist criterion, this is unambiguous in velocity up to a Doppler frequency of 5kHz

The maximum unambiguous radial velocity is given by the following formula:

$$v_r = \frac{f_d \lambda}{2} = \frac{PRF.\lambda}{4}$$

Table 15.10: Unambiguous velocity as a function of frequency

Band	Unambiguous Velocity v_r (m/s)
X	75
Ka	21.5
W	8

All velocities will be folded down into this range, so it is possible that the folded target velocity will fall directly into the bin (or bins) containing the clutter.

This can be ameliorated by changing the PRF on a block by block basis, which will shift the relative positions of the clutter and the folded target Doppler frequencies until an one is found in which the target is not in the same bin as the clutter.

It is quite likely that the clutter velocity spread will exceed 8m/s, so there may not be an acceptable PRF at W-Band.

Because both the clutter and the ship may be moving, the actual gate in which the former will be found must be tracked.

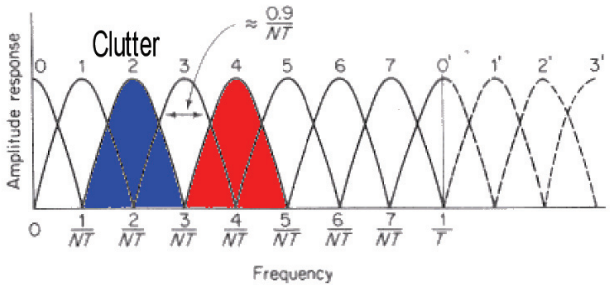


Figure 15.29: Separating target and clutter returns in a Doppler filter bank

15.7.12. Search Requirement

Need to search a volume 10° × 4° × 50m in less than 0.5s to meet the specification.

Any of the following patterns can be adopted

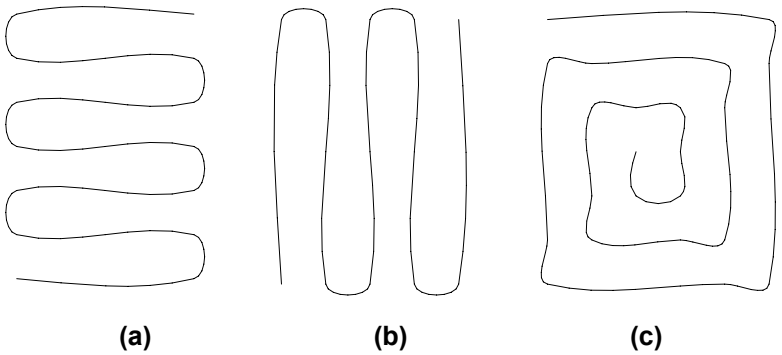


Figure 15.30: Search pattern options

Option (a) requires that the pedestal changes direction more often than option (b), and there is more of an overlap between scans with option (c), so adopt option (b)

For a 50% overlap between scans, the total distance that must be travelled to search each area is determined from the number of vertical scans required to cover the 4° in azimuth.

Table 15.11: Total search scan distance as a function of frequency

Band	Vertical Scans (n)	Total Distance (deg)	Scan Speed (deg/s)
X	4/1.4x1.5=4.28 [5]	50	100
Ka	4/0.4x1.5=15 [15]	150	300
W	4/0.13x1.5=46 [46]	460	920

Hits per scan (the number of pulses that illuminate the target as the beam passes over it) is determined from the required scan rate and the antenna beamwidth.

Table 15.12: Hits per scan

Band	Hits Per Scan (N)
X	140
Ka	13.3
W	1.4

There are not sufficient hits at either Ka band or W band to generate an FFT with sufficient bins

Decreasing the overlap to 0%

Table 15.13: Scan distance and hits per scan for 0.5s detection time

Band	Vertical Scans (n)	Total Distance (deg)	Scan Speed (deg/s)	Hits per Scan (N)
X	3	30	60	233
Ka	10	100	200	20
W	31	310	620	2.1

There is the added burden that the target may have moved up to 140m during the search, so a large bank of gates spanning the original 50m uncertainty plus 140m on each side (330m) must be examined.

At least 17 gates will be required to span this range.

It is now feasible to use a 16 point FFT at Ka band or a 128 point FFT at X-band.

15.7.13. Integration Gain

The FFT process is the equivalent of a coherent integrator, which will produce gains of $10\log_{10}(N)$. However, because the target is fluctuating, this effect will vary with the observation time compared to the fluctuation rate and the probability of detection.

We use the following graph (which is not quite correct as it assumes post detection integration).

For a $P_d \approx 0.95$ (we use 0.9) and a Swerling 2 target, the integration gain is about 17dB for the 16 point FFT at Ka band.

For the 128 point FFT at X band, the integration gain is about 24dB.

Though these are higher than the theoretical maxima of 12 and 21dB respectively, the extra is taken care of in the 10.4dB required for detection described earlier.

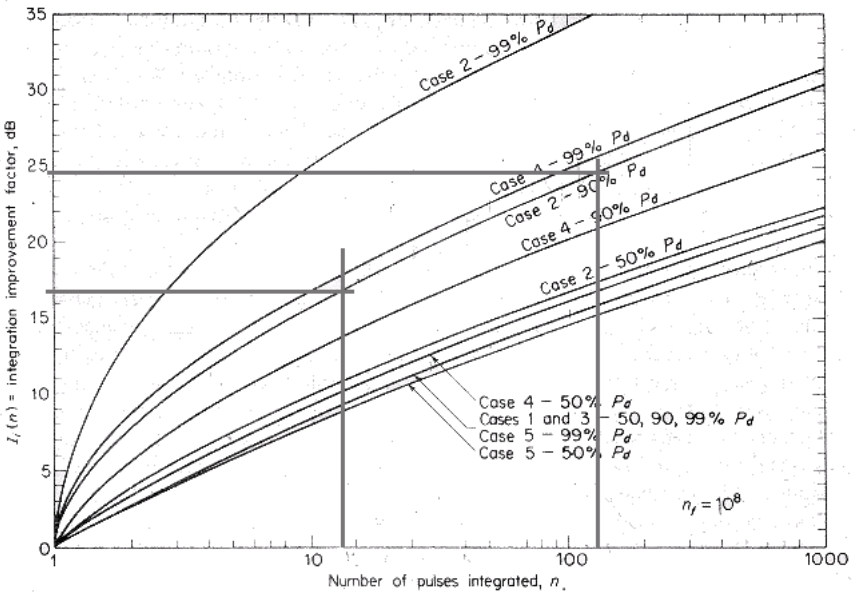


Figure 15.31: Integration improvement factor

15.7.14. Matched Filter

We assume that the pulse is rectangular and that the filter is made up of 5 cascaded tuned bandpass sections

For a pulse width of 20m (133ns), the optimum bandwidth $\beta = 5$MHz

Table 15.14: Matched filter characteristics

Input Signal	Filter	Optimum $\beta\tau$	Loss in SNR compared to Matched Filter (dB)
Rectangular Pulse	Single tuned circuit	0.4	0.88
Rectangular Pulse	Two cascaded tuned circuits	0.613	0.56
Rectangular Pulse	Five cascaded tuned circuits	0.672	0.5

15.7.15. Transmitter Power

The average power allowed is 100W, the pulse width is 133ns and the PRF is 10kHz making the duty cycle 0.133% and so the peak transmitted power $P_t = 75$kW.

This can be achieved using a Magnetron or Travelling Wave Tube at Ka or X band

15.7.16. System Configuration

A Magnetron can be used in a (pseudo) coherent radar if the LO is primed by the random start phase of the magnetron on every pulse

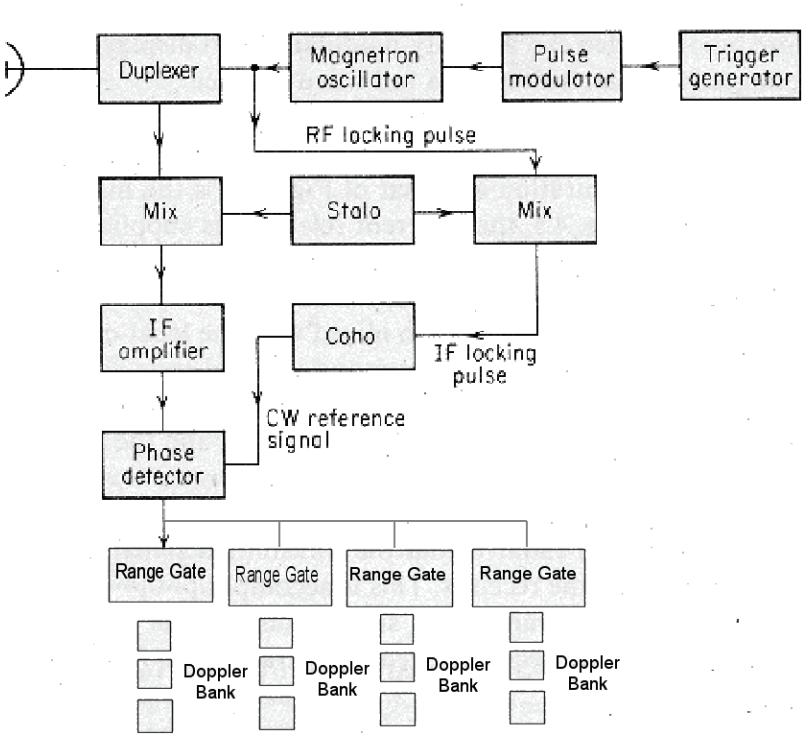


Figure 15.32: Schematic diagram of a pseudo coherent pulsed radar

15.7.17. Free Space Detection Range

Applying the radar range equation when there is significant attenuation is best achieved using MATLAB

Various losses are listed in the M-file

```
% Fire Control Radar

r=(10:10:75000);    % Range (m)
sigmat = 1;         % Aircraft (sqm)
sigmas = 0.1; % Sea skimmer (sqm)
dant = 1.5;         % Antenna Diameter (m)
pt = 75e03;         % Trasmit power (W)
lamx = 30e-03;      % Wavelength (m)
lamk = 8.6e-03;
nfxdb = 3;          % Noise fig (dB)
nfkdb = 4.5;
tau = 133e-09;      % Pulse width (s)
k = 1.38e-23; % Boltzmann (Js)
t = 290;
nintxdb = 24; % Integ i mprove(dB)
nintkdb = 17;
ltxx = 3;           % Tx plumbing loss (dB)
ltxk = 4;
lscanx = 3.2; % 2D Scan loss (dB)
lscank = 3.2;
lmiscx = 3;         % Misc loss (dB)
lmisck = 3;
alphax = 1;         % Atmos attn (dB/km) 0.02, 0.25, 1
alphak = 7;         % cREAR, 12.5, 25mm/hr 0.15, 3, 7

SNx = 13.6+10.4;    % S/N for Pd=0.95, Pfa=1.0e-06
SNk = 13.6+10.4;

% Calculate the antenna gain
gxdb = 10*log10(4*pi*0.7*pi*dant*dant/(4*lamx*lamx));
gkdb = 10*log10(4*pi*0.7*pi*dant*dant/(4*lamk*lamk));

% Propagation factor
propxdb = 10*log10(lamx*lamx/(4*pi).^3);
propkdb = 10*log10(lamk*lamk/(4*pi).^3);

% Transmit power
ptdb = 10*log10(pt);

% Target RCS
sigmatdb = 10*log10(sigmat);     % Aircraft
sigmasdb = 10*log10(sigmas);     % Sea skimmer

prxdb     =     ptdb+2*gxdb+propxdb+sigmasdb-ltxx-lscanx-lmiscx-40*log10(r)-
2*alphax*r/1000;
prkdb     =     ptdb+2*gkdb+propkdb+sigmasdb-ltxk-lscank-lmisck-40*log10(r)-
2*alphak*r/1000;
% Threshold for detection
thermxdb = 10*log10(k*t*0.672/tau);
thermkdb = 10*log10(k*t*0.672/tau);

threshxdb = (SNx + thermxdb + nfxdb --nintxdb)*ones(size(r);
threshkdb = (SNk + thermkdb + nfkdb --nintkdb)*ones(size(r);

semilogx(r,prxdb-threshxdb,r,prkdb-threshkdb);
grid
title('FIRE CONTROL RADAR: MAGNETRON: SKIMMER: Clear Air')
xlabel('RANGE (m)')
ylabel('SIGNAL LEVEL ABOVE Smin (dB)')
```

```
axis([0,100000,-50,200]);

lx=find((prxdb-threshxdb)<0);
lk=find((prkdb-threshkdb)<0);
min(r(lx))
min(r(lk))
```

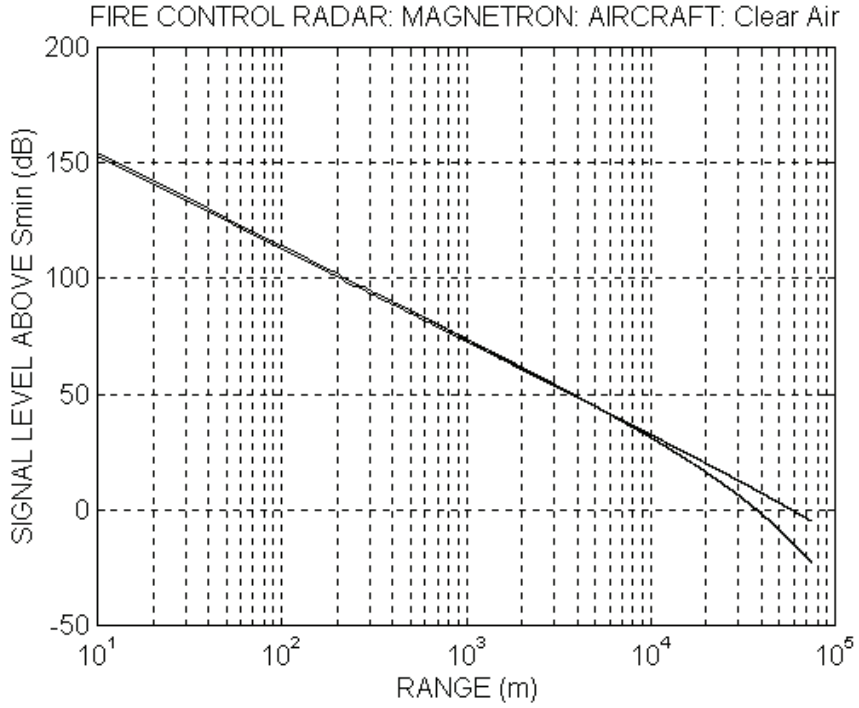


Figure 15.33: Radar performance, clear air

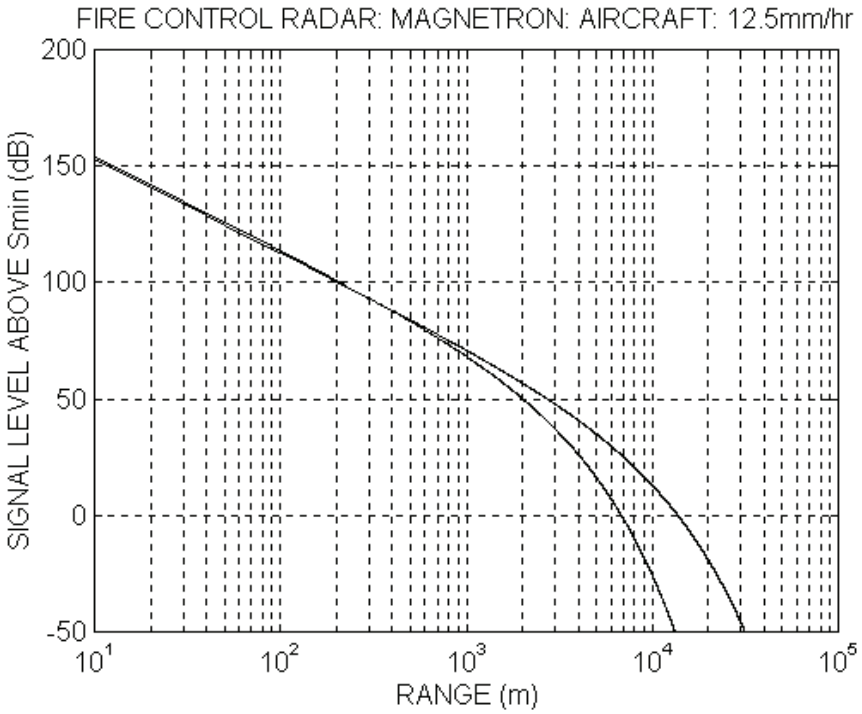


Figure 15.34: Radar performance, 12.5mm/h rainfall

Table 15.15: System performance

Target	Condition	X-Band Range (km)	Ka-Band Range (km)
Aircraft RCS=1sqm	Clear	58	37.5
	12.5mm/h	28.9	6.8
	25mm/h	13.7	3.6
Sea Skimmer RCS=0.1sqm	Clear	34.4	25.8
	12.5mm/h	20.6	5.7
	25mm/h	10.8	3.1

For aircraft detection, only the X-Band radar meets the specified range criteria of 15, 10 and 5km

Both the X-Band and the Ka-Band radars meet the sea skimmer detection requirements of 6, 4 and 2km

15.7.18. Effects of Multipath on Aircraft Detection

The following graphs show that the constructive and destructive interference caused by multipath can reduce the signal level S_{min} to below 0dB under some circumstances.

The fades are more pronounced at X-Band where detection may not occur between 13 and 14km, and again at just below 10km.

None of the fades at Ka-Band dip so low.

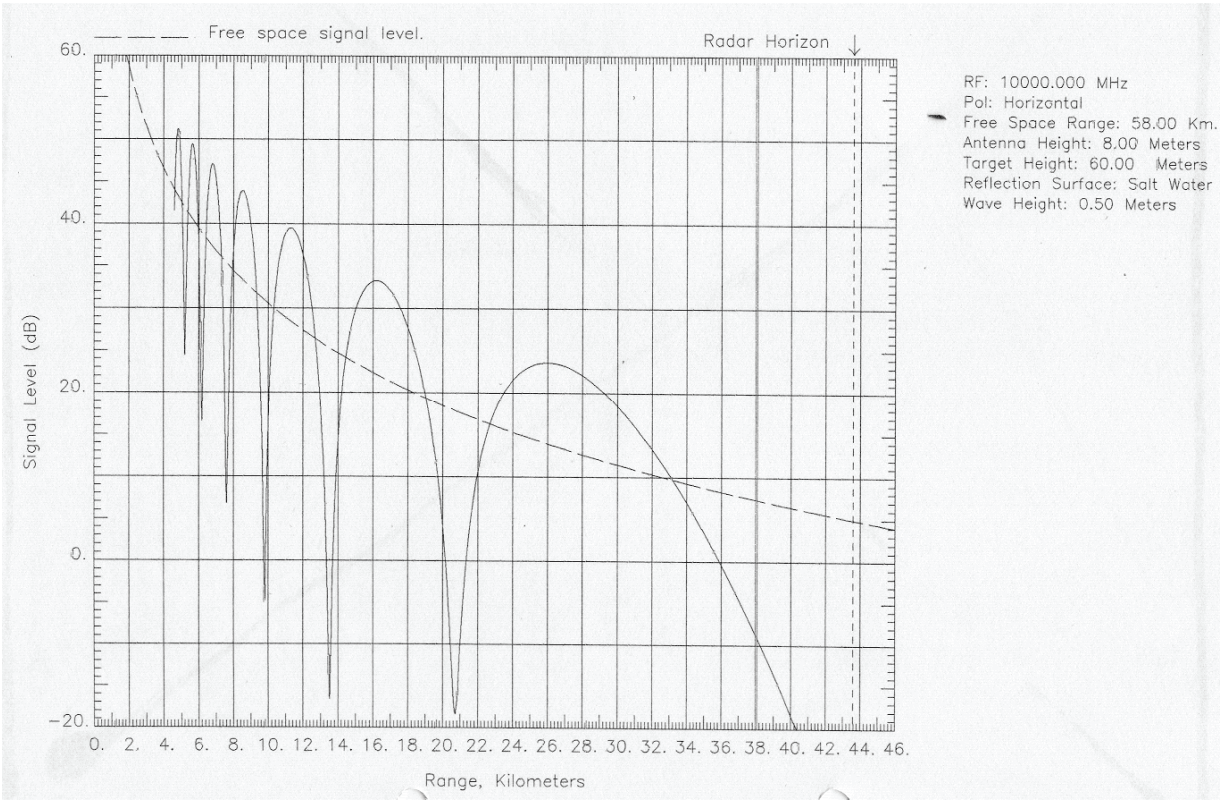


Figure 15.35: Multipath effects on aircraft tracking at X-band

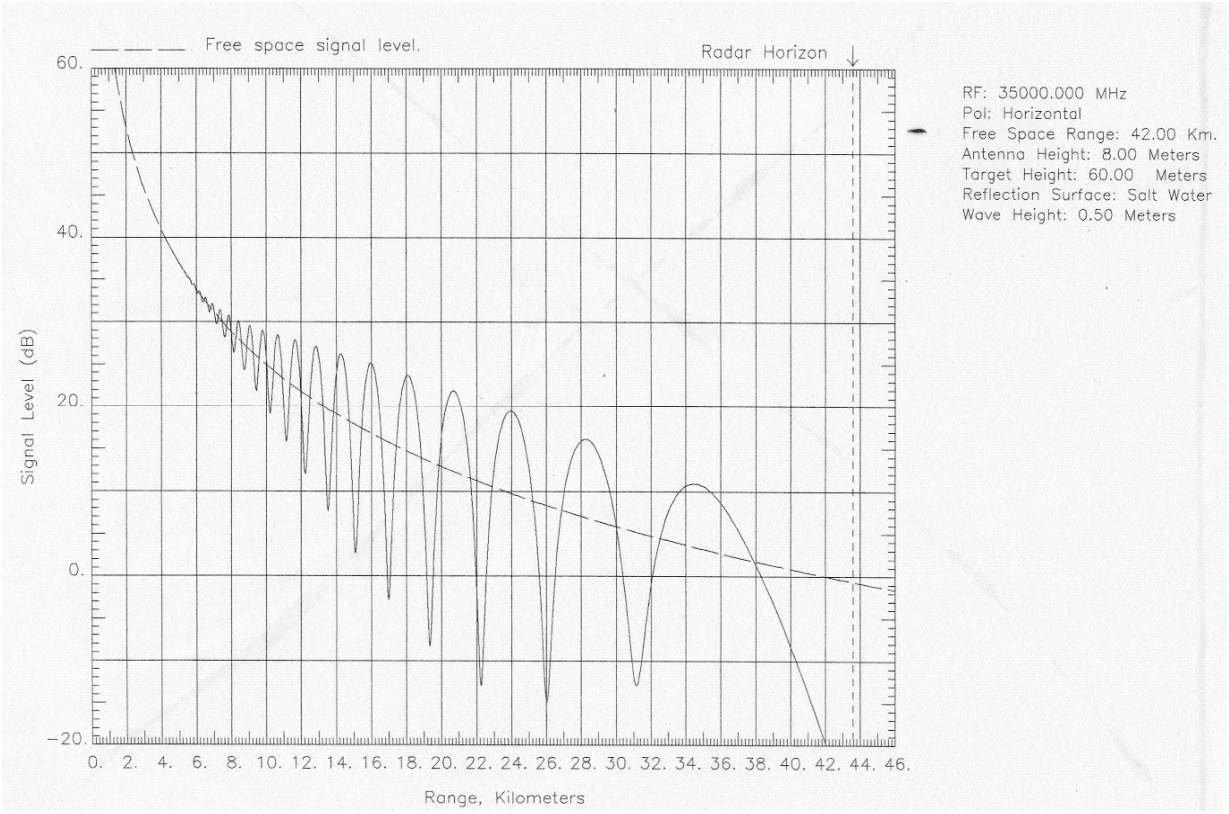


Figure 15.36: Multipath effects on aircraft tracking at Ka-band

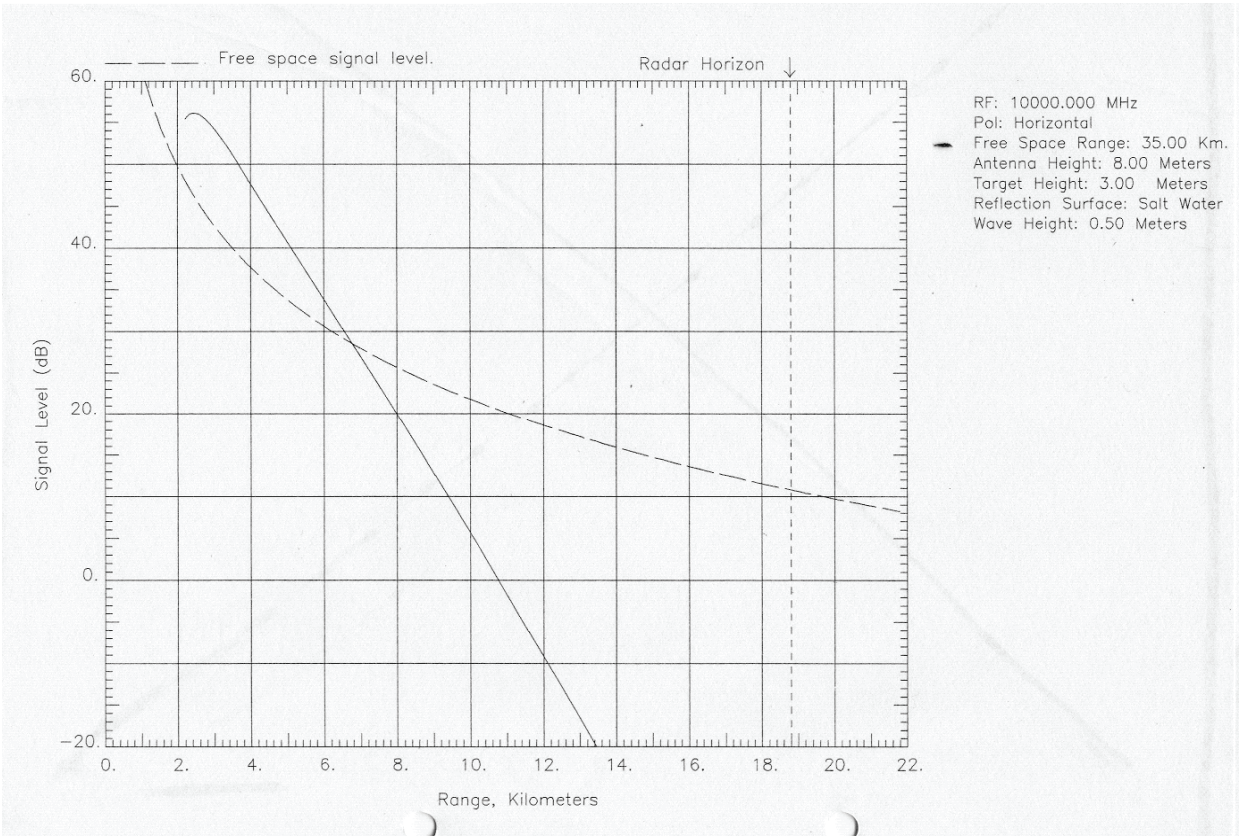


Figure 15.37: Multipath effects on sea skimmer tracking at X-band

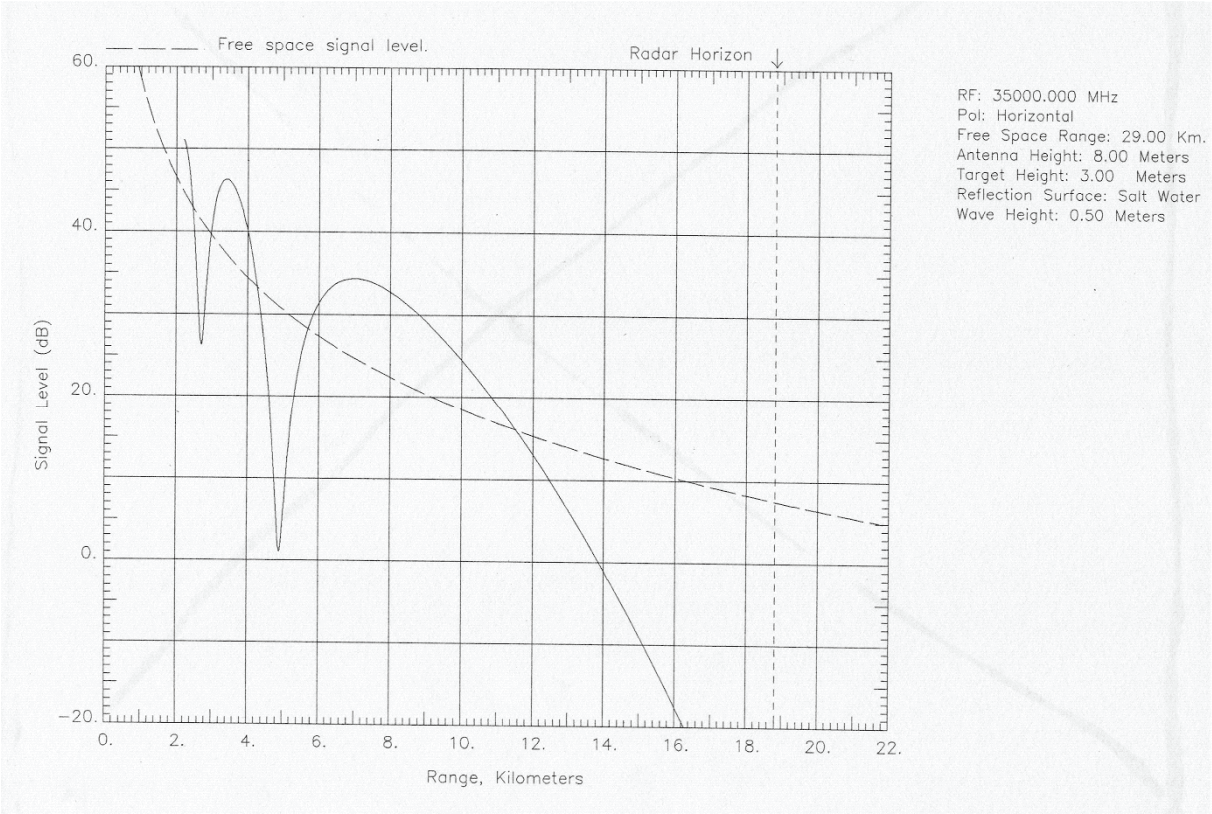


Figure 15.38: Multipath effects on sea skimmer tracking at Ka-band

15.7.19. Detection Threshold and CFAR

The false alarm rate is very sensitive to the setting of the detection threshold voltage.

Because no velocity information is available, the radar must transmit a block of pulses and look in all the Doppler gates in all of the range gates over the designated range of the target.

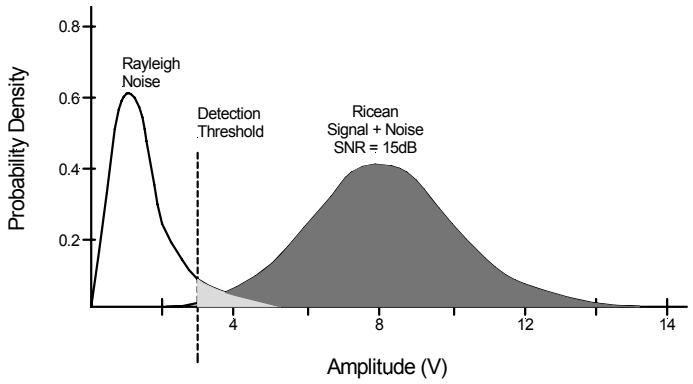


Figure 15.39: Selection of detection threshold

Changes in radar characteristics with time (ageing) and changes in the target background characteristics mean that a fixed detection threshold is not practical. A Constant False Alarm Rate Processor (CFAR) is required

We propose a cell-averaging CFAR processor that averages the returns from a particular Doppler bin across all of the range gates that span the designated range as shown in the highlighted row.

This averaging process excludes the cell under test which proceeds sequentially from gate 1 to gate N (it is shown in gate 3 in the Figure)

Range gates

Doppler Bins		1	2	3	4	5					N
	1										
	2			X							
	3										
	125										
	126										
	127										
	128										

Figure 15.40: Range/Doppler gate matrix showing cell containing target

15.7.20. Transition to Track

The mechanically operated FCR antenna moving at 60°/s (X-Band) or 200°/s (Ka-Band) during the search phase is unable to stop while the antenna is still pointing at the target.

The angles, range and Doppler bin at which detection takes place is recorded as the antenna sweeps past and the antenna returns to that designation more slowly.

If the target is not detected, then a slow search is conducted in angles while the range gates broaden their search until a detection occurs or the system times out. The angular rates of this new search are such that the antenna can stop while still illuminating the target, and a transition to track mode is made.

The process of detection involves closing the Doppler loop, followed by the range loop and finally the angle loop.

15.7.21. Target Tracking

A split tracking gate straddles the sum-channel range-gate in which the detection took place and early and late gate measurements are made only in the Doppler bin in which detection took place.

As the Doppler is generally ambiguous, it cannot be used to prime the range filter, so a long (typically 50ms or so) period is used to obtain a good range rate estimate to prime the range tracking filter.

The range tracking loop closes using early and late gate signals from the appropriate Doppler bin, and automatic tracking in range occurs using the difference between the early and late gate signals to keep the gates centred on the target

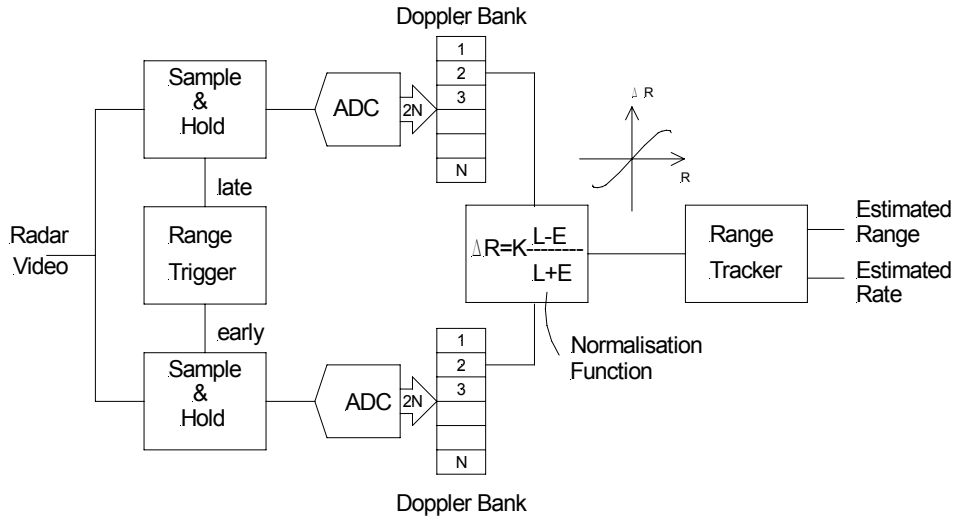


Figure 15.41: Schematic diagram of range tracking configuration

The angle tracking loops are closed.

The azimuth and elevation error signals sampled simultaneously with the sum channel signal drive the angle servos to keep the antenna pointed at the target being tracked.

Target dynamics that dictate the real and apparent acceleration and tracking loop bandwidth determines the tracking accuracy.

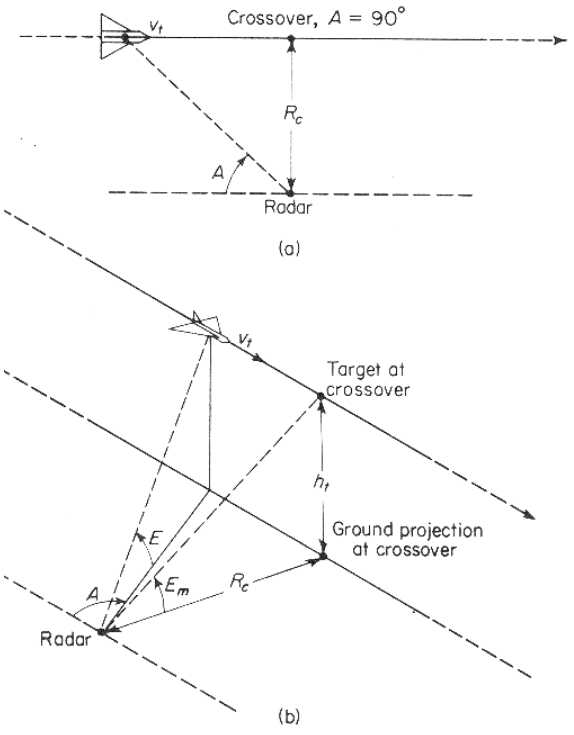


Figure 15.42: Crossing target geometry

When viewed in radar polar co-ordinates, the target velocities will be

$$\dot{R}_{\max} = v_t, \dot{A}_{\max} = \frac{v_t}{R}, \dot{E}_{\max} = \frac{v_t}{R}.$$

The real accelerations are

$$\ddot{R}_{\max} = a_t, \ddot{A}_{\max} = \frac{a_t}{R}, \ddot{E}_{\max} = \frac{a_t}{R}.$$

The geometric accelerations and other derivatives are shown below. They are normalised in the angular co-ordinates to $\dot{A}_{\max} = \varpi_m = v_t / R_c$. Where R_c is the ground range at crossover.

In range, the derivatives are normalised to $R_a = R_{min}$, at the point of closest approach.

Assuming that the target is travelling at 280m/s and flies past the radar at a height of 60m and a ground range of R_c = 60m, then the peak velocities and accelerations experienced (in polar co-ordinates)

ϖ_m = 4.66rad/s $\dot{\varpi}_a$ = 13.75 rad/s^2 (from the graph)

ϖ_e = 1.6 rad/s $\dot{\varpi}_e$ = 10.42 rad/s^2

$\dot{R}_{\max}$ = 280m/s $\ddot{R}_{\max}$ = 923 m/s^2

Assuming that thermal noise and dynamic lag are the primary sources of error, then an optimum bandwidth is calculated as follows:

$$B_n = \left[\frac{a_t^2 k_m^2 f_r B\tau (S/N)}{1.6\theta_3^2 R^2} \right]^{1/5}$$

where: a_t – acceleration (real or geometric) (m/s^2) [923]
k_m – Monopulse Gain Constant (typ 1.6) [1.6]
f_r – Pulse repetition frequency (Hz) [10kHz]
$B\tau$ - IF bandwidth and pulse width [0.672]
S/N –pulse signal to noise ratio [10]
θ_3 – Antenna beamwidth (rad) [0.024 or 0.007]
R – Crossing Range (m) [85]

The optimum angle servo bandwidth is 116Hz for the X-Band tracker and will be significantly higher for the Ka-Band unit. This is very high as the specified crossing range is extremely short.

If the bandwidth is expressed in terms of the equivalent noise bandwidth B_n then the acceleration coefficient

$K_a = 2.5B_n^2 = 0.63/t_o^2$

Where t_o is the equivalent averaging time for the tracking loop. The lag will be.

$$\varepsilon_a = \frac{\dot{\varpi}_a}{K_a} = \frac{\dot{\varpi}_a}{2.5B_n^2} = \frac{13.75}{2.5\times 116^2} = 0.409\times 10^{-3}\,\text{rad}$$

$$\sigma_\theta = \frac{\theta_3}{k_m\sqrt{2(S/N)f_r/B_n}} = \frac{0.024}{1.6\sqrt{2\times 10\times 10^4/116}} = 0.36\times 10^{-3}\,\text{rad}$$

The RMS tracking error caused by the thermal noise will be $\sigma_\theta = 0.36$mrad

The RMS tracking error caused by dynamic lag will be $\varepsilon_a = 0.41$mrad.

The alternative is to use a wide angle-servo bandwidth (≈200Hz) that will cope with the geometric accelerations, and to perform the filtering in Cartesian space with a bandwidth of between 1 and 4Hz that will cater for real target accelerations that should not exceed 6g for the aircraft.

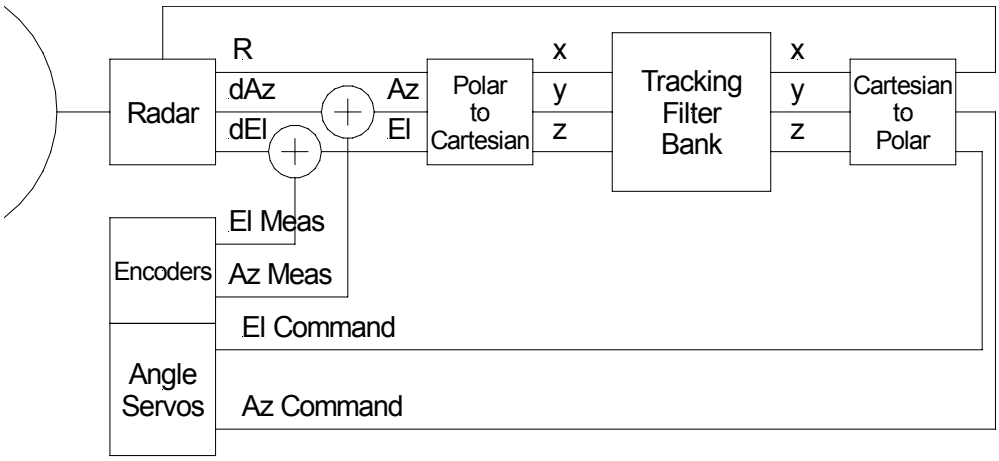


Figure 15.43: Tracking in Cartesian space

15.8. References

[1] M.Skolnik, *Radar handbook,* McGraw-Hill, 1970.
[2] M.Skolnik, *Introduction to Radar Systems,* McGraw-Hill, 1980.
[3] P.Bhartia, I.Bahl, *Millimeter Wave Engineering and Applications,* John Wiley & Sons, 1984
[4] D.Barton, *Radar Systems Analysis*, Artech 1976.
[5] C.Currie (ed), *Principles and Applications of Millimeter Wave Radar,* Artech, 1987.
[6] G. Morris, *Airborne Pulsed Doppler Radar,* Artech, 1988.
[7] N. Currie (ed), *Radar Reflectivity Measurement: Techniques & Applications,* Artech, 1989.
[8] S. Sherman, *Monopulse Principles and Techniques,* Artech
[9] M. Spiegel, *Mathematical Handbook of Formulas and Tables,* McGraw Hill, 1968.

Chapter 16.

Radio Tags and Transponders

16.1.Principle of Operation

Transponders were originally electronic circuits that were attached to some item whose position or presence was to be determined.

The transponder operated by responding to an interrogation request received from an interrogator, either by returning some data from the transponder such as an identity code, or returning the original properties of the signal received from the interrogator with a minimum time delay.

Because the interrogator signal was generally much stronger than the returned signal, the former would swamp the latter unless some characteristic of the response was changed. This was generally the frequency, but it could be polarisation.

16.2.History and Applications

Transponders were developed for aircraft in WWII for identification friend or foe (IFF) application. Both commercial and military aircraft still use this technology. Most air traffic control centres rely on the range and altitude information returned by commercial aircraft rather than raw radar information.

Since the Gulf War during which unacceptably large numbers of personnel were killed by their own side, transponder technology called Battlefield Identification System (BIS) has been developed for ground vehicles. It relies on interrogator – transponder technology operating at 38GHz.

Another important application of the transponder has been in the measurement of distance. Here, the interrogator sends a signal to the transponder which immediately responds on another frequency. By using the time of flight, the range can be determined with great accuracy. This technique is more accurate than using the skin echo as the ultimate signal to noise ratio remains very high.

A system called the Tellurometer (see Chapter 5) used this technique with improvements to measure ranges of hundreds of kilometres to an accuracy of a few cm.

The GSM network is being used increasingly as a form of transponder system, where the position of each phone can be determined with great accuracy from field strength readings at various base stations. An application of this technology includes monitoring traffic flow conditions on major roads by tracking phone positions with time.

Transponder systems have become major players in the field of electronic identification. Within this application, it is necessary to make the transponder as cheap as possible and to build the sophistication into the interrogator (reader).

Because it is not feasible to perform low cost frequency translation, these transponders (or tags) have given up the ability to supply range, and time slice the interrogation and response cycle to avoid swamping the response.

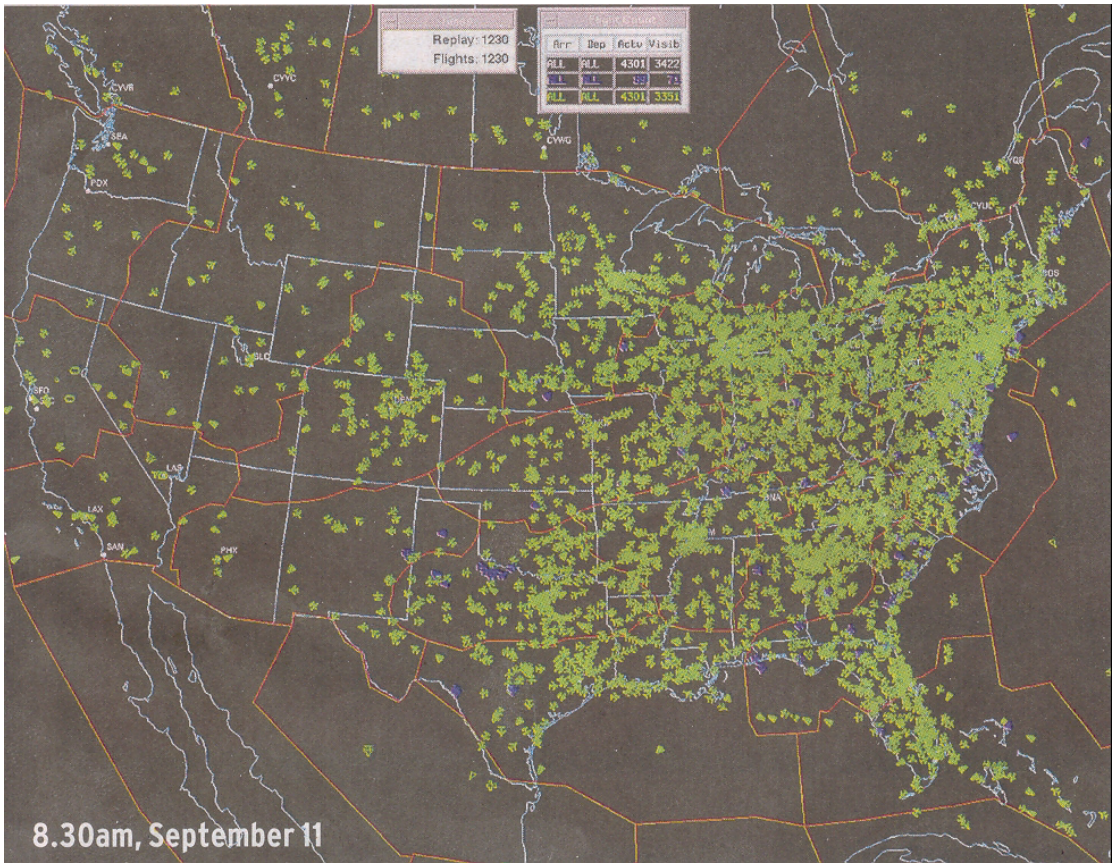


Figure 16.1: Transponder tracked air traffic over North America

16.3.RFID Systems

Radio Frequency ID systems covers a wide range of technologies from reading railway containers to applications in retail. The principle involved relies on the ability of EM radiation to penetrate certain goods and to read a tag that is not visible where the response can be in the form of an identity code or just a presence called Electronic Article Surveillance (EAS).

Commonly available tags operate at frequencies in the range from 60kHz to 5.8GHz.

There are three main implementations of this technology:

- EAS based
- Magnetic based
- Electric field based

16.3.1. Electronic Article Surveillance (EAS)

EAS systems are used to electronically detect goods that have not been authorised when they are removed from the retailer. They comprise a tag attached to the goods and a sensor mechanism. Tags can only be neutralised by the retailer when he wishes to authorise the removal of the goods.

In effect these are single-bit RFID systems. They convey presence but not identity.

There are four major technologies used for EAS systems

- Microwave
- Magnetic
- Acousto-Magnetic
- Radio Frequency

Market penetration is estimated at 6billion tags per annum at $0.12 each.

Magnetic and radio frequency tags are very cheap and are generally attached permanently to the goods or packaging, while microwave tags are expensive and are removed from the merchandise using a special tool.

Radio Frequency Tags

Comprise a tuned LC circuit that resonates at radio frequencies that absorbs energy from a field at the resonant frequency. These can be deactivated by applying a high power RF pulse that burns out a fuse in the circuitry.

Acousto-Magnetic Tags

This marker consists of a strip of magnetostrictive material and a strip of magnetic material of high coercivity. It resonates mechanically in the presence of a magnetic field of a particular frequency and this resonance can be detected by a receiver sensitive to the changing field created by the flexing magnetostrictive material. The marker is deactivated by modifying the magnetic bias on the strip of magnetic material. This technology is widespread for book protection in libraries.

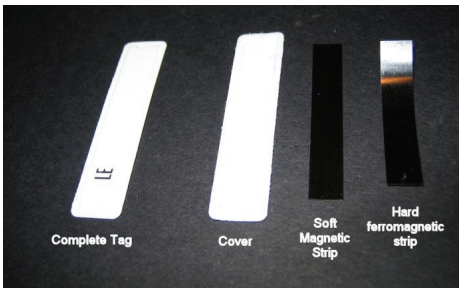


Figure 16.2: Photograph of a dismantled acousto-magnetic tag

Microwave Tags (E-tags)

Modern microwave tags operate at 2.45GHz using the interrogator transponder technique with an active low power transponder offering a battery life of 10 years. Operational range is typically 4m through dirt, glass, wood etc. These are the units that operate on Sydney's toll roads.

The Electronic Toll Collection (ETC) system works as follows:

- As a car approaches a toll plaza, the radio-frequency (RF) field emitted from the antenna activates the transponder.
- The transponder broadcasts a signal back to the lane antenna with some basic information.
- That information is transferred from the lane antenna to the central database.
- If the account is in good standing, a toll is deducted from the driver's prepaid account.
- If the toll lane has a gate, the gate opens.
- A green light indicates that the driver can proceed. Some lanes have text messages that inform drivers of the toll just paid and their account balance.

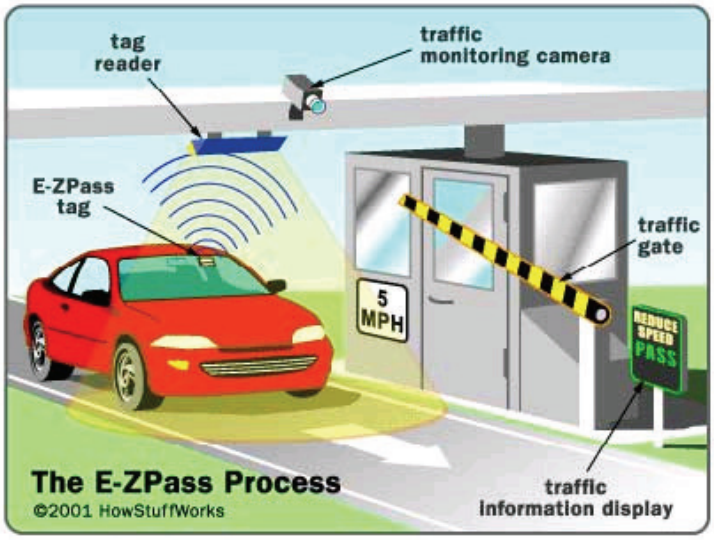


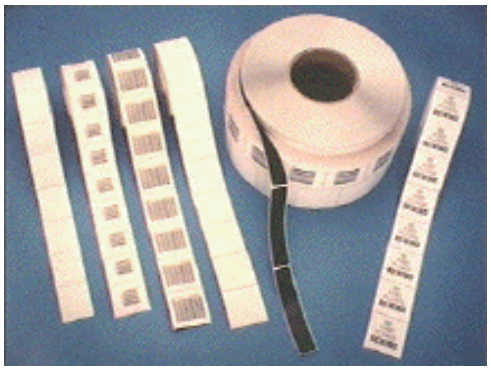
Figure 16.3: Microwave tag used in an electronic tolling system

A study has shown that a dedicated cash lane had an average transaction time of 10.5 seconds for a passenger car and 29.5 seconds for a commercial vehicle. A dedicated ETC lane processed 1000 vehicles per hour, with an average transaction time of 3.6 seconds per vehicle. The typical savings is 6.9 seconds per passenger car. The greater time savings, however, is realized with the elimination of queues due to the reduction of transaction time. In the study, queues of over 20 vehicles, which took up to three minutes to process had been observed at some toll plazas prior to ETC. With the introduction of ETC, there are virtually no queues at interchanges and barriers where there once was often heavy congestion.

16.3.2. Multibit EAS Tags

These are generally of the LC tuned circuit design, and have a series of capacitors that can be selectively shorted (mechanically or electrically) to alter the resonant frequency.

16.3.3. Magnetic Coupled RFID Transponder Systems



These are the most common transponders available today

Operate at frequencies in the order of 125kHz

Antenna systems comprising a multi turn coil designed to collect energy from the readers magnetic field.

Because of the magnetic coupling, the range is limited to a couple of cm.

Figure 16.4: Printed magnetic coupled tags

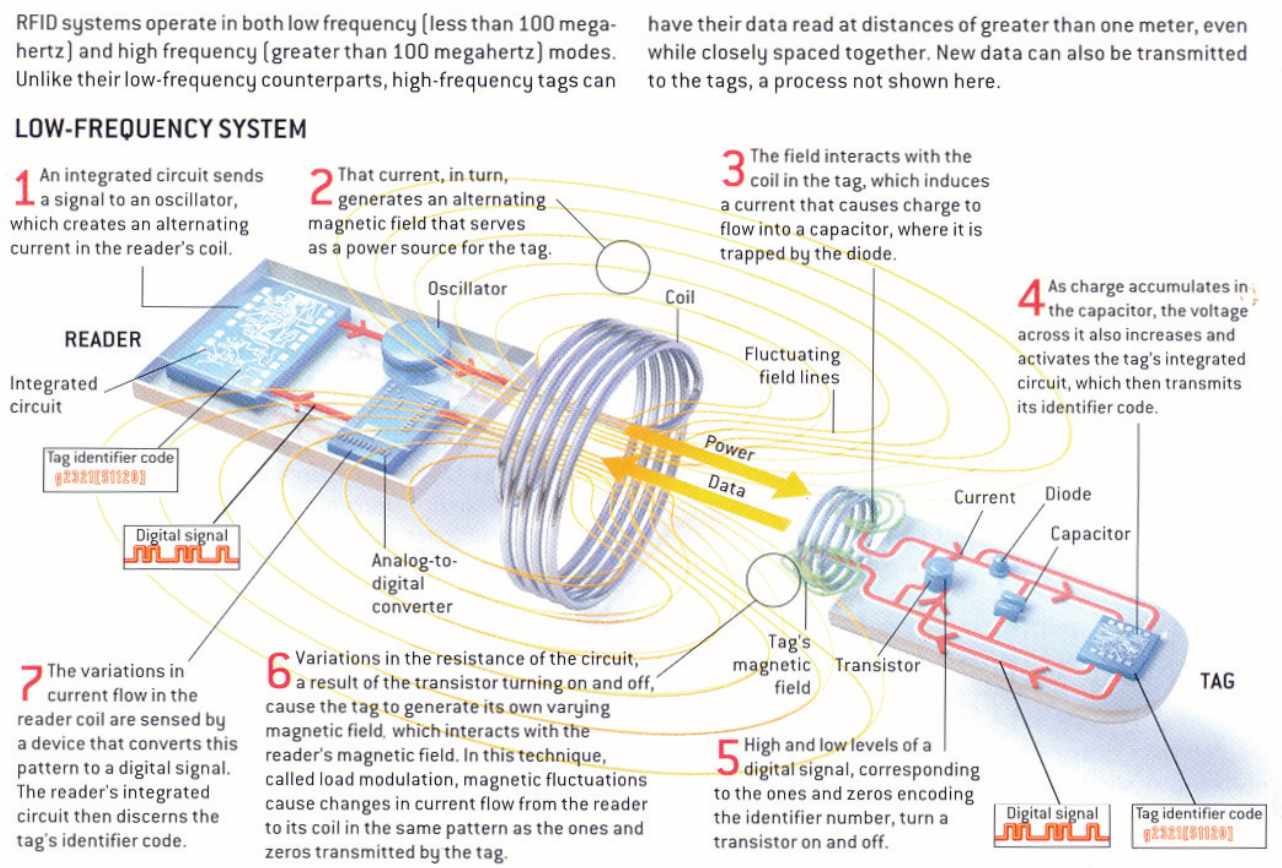
Operational Principles

Using the 125kHz energising field, the transponder sends data back to the reader using one of the following techniques:

- Transmits a signal back at half the frequency while the transmitter operates in CW mode.
- Transmit back at the transmitter frequency after the transmitter has switched off, during the decay period of the LC circuit (called flyback)
- Let the tag load the transmitter field with a fluctuating load being sensed by the transmitter circuitry.

A new generation of tags that operate at high frequency (up to 29MHz) with read/write capabilities have been available since 1998. These tags allegedly have some anti-collision properties which allow many tags to be in the reader field simultaneously.

Reading and writing distance is generally limited to 20cm, but some manufacturers claim operational distances up to 1m



16.3.4. Electric Coupled RFID Transponder Systems

Rather than using magnetic lines of force to couple the interrogator and transponder, this technology relies on EM propagation which offers much longer range that magnetic coupling technology.

Antennas are typically half-wave dipoles, and so to make the transponders small, the operational frequency must be high. However higher frequency components cost more and the Energy transfer is proportional to $1/f^2$ so high frequency devices cannot transfer much energy.

The field coupling is proportional to $1/R^2$ so the energy transfer becomes even more difficult.

The power density at a range, R, from the transmitter is determined as follows

$$E = \frac{P_t G_t}{4\pi R^2} \quad \text{W/m}^2, \tag{16.1}$$

where E – Power density (W/m^2),
P_t – Transmitted power (W),
G_t – Transmitter gain (wrt isotropic radiator),
R – Operating range (m).

The received power is the product of the power density, E, and the receiver antenna effective aperture, A_e,

$$P_r = EA_e \,. \tag{16.2}$$

The effective aperture of a half wave dipole is a function of the receiver gain, G_r and the square of the frequency

$$A_e = \frac{G_r \lambda^2}{4\pi} = \frac{G_r c^2}{4\pi f^2} \,. \tag{16.3}$$

Assuming that antenna gains and matching remain constant for all frequencies, then the minimum received power, P_{min}, determines the maximum operational range, R_{max}.

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi R)^2} = P_{\min} \,, \tag{16.4}$$

$$R_{\max} = \frac{\left(P_t G_t G_r\right)^{1/2} \lambda}{4\pi P_{\min}^{1/2}} \,. \tag{16.5}$$

The maximum allowable power of the interrogator is determined by health and safety regulations. In general a safety factor of 10 is applied to this data.

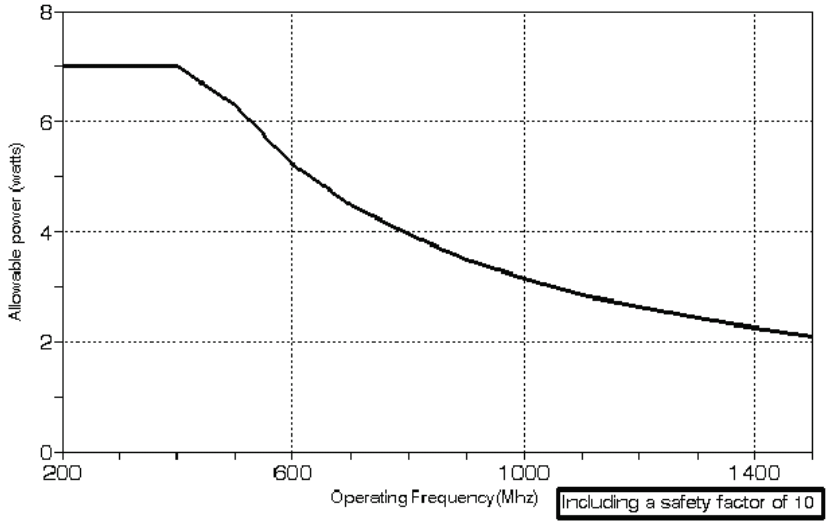


Figure 16.5: Allowable radiated power

Passive transponders operating at 400MHz can operate to a maximum range of 10m, down to less than 1m at 2.5GHz.

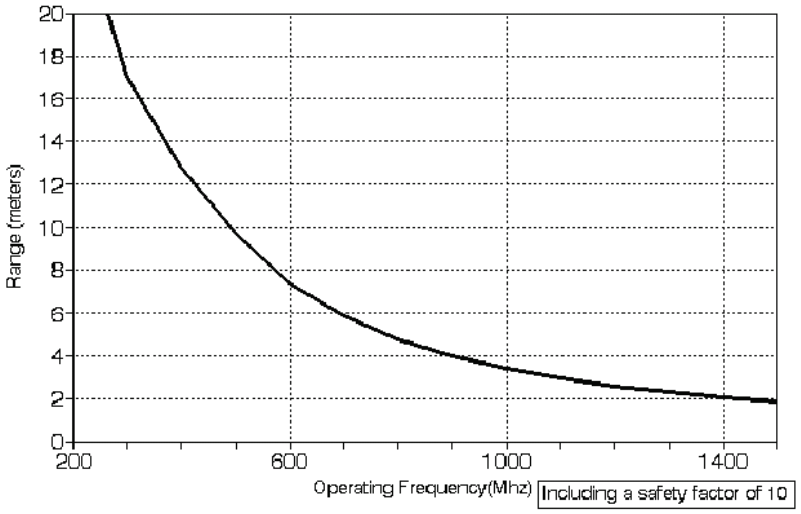
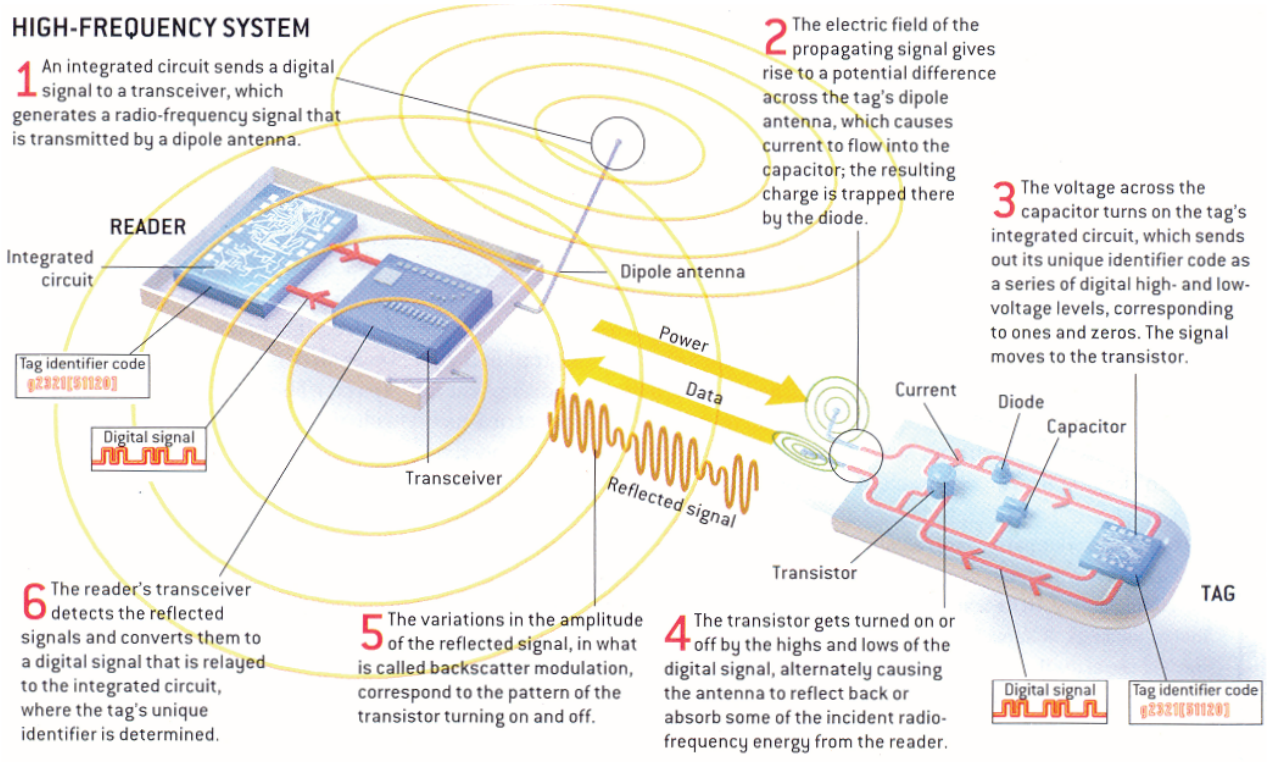


Figure 16.6: Detection range as a function of frequency for passive transponders

Using the backscatter modulation principle developed at Lawrence Livermore Laboratories has meant that electric field tags can be manufactured for 10c (US).



16.3.5. Law Enforcement

Offenders charged with serious crimes are electronically tagged while on bail awaiting trial, in an effort to stop them re-offending. The system is designed "to stop thugs cocking a snoot" at police and the legal system which is often forced to free them to re-offend. The move is in response to soaring levels of street crime and robberies in the UK and other western countries.

The offender wears an unobtrusive transmitter (tagging device) on his or her ankle or wrist which emits a signal that is picked up by a monitoring unit when the offender is within range. The monitoring unit is linked by a telephone line to a central computer system where the information about the offender's presence or absence is permanently stored. If the offender leaves the range the central computer system will be alerted.

It is not possible to remove the tag and leave home unmonitored and any tampering with the electronic monitoring unit will alert the central computer and be acted upon immediately.



Figure 16.7: Electronic tag for conditions of bail monitoring

16.3.6. Social Issues

Privacy advocates suggest that the identification of specific items linked by their RFID, to a credit card would give marketing companies unprecedented access to an individuals purchasing profiles which would allow consumers to be targeted with specifically tailored sales pitches (with a resulting increase in credit card dept that society can ill afford)

Audit trails of commercial transactions would allow the police to trace the movement of individuals through logs maintained by the RFID tracking infrastructure. This would be in breach of an individuals right to privacy, and laws are being formulated as regards access to these logs.

Job losses due to the further automation of many processes will lead to labour unrest in certain circumstances

A number of large scale RFID trials, particularly those by Benetton, Wal-Mart and Gillette have been curtailed after pressure from privacy groups who were against the insertion of tags into individual items.

Even placing the tags into packaging has been criticised as it would allow a criminal element to determine (from your rubbish) that you have just purchased an expensive item.

It has been proposed that children be tagged at birth which would allow parents, police, and even school administration an unprecedented and invasive knowledge of an individuals whereabouts – from the cradle to the grave.

Issues with regard to the ethics of electronic tagging of offenders is also under consideration. However, it is generally believed that keeping the offenders under what is effectively "house arrest" is less damaging than a stint in an overcrowded jail would be.

16.4. Technical Challenges

RFID tags and readers are orientation dependent. Tags must be positioned properly relative to the readers so that antennas can exchange signals. This can be overcome by using arrays of readers orientated appropriately.

RFID signals are easily blocked. Over short range they can be attenuated by packaging, particularly that made from metal foils or metallised plastic. Over longer ranges, because the signal levels are so low, they can be blocked by many common objects, even the human body. Again this can be solved by having arrays of readers

At an average cost of 20 to 30 cents (American), tags are still to costly for inclusion into individual items. For this reason, mass market consumer retail business that operates on low margin, high turnover has yet to embrace the technology

Competing technical standards and protocols prevent their universal adoption, though the International Organisation for Standardisation (ISO) is working to overcome this problem

The proposed 96 bit standard ID would allow each of the six billion human beings alive to acquire 1.3×10^{19} individually tagged items during their lifetime. This should be sufficient!

16.5. Battlefield Combat ID System (BCIS)

Millimetre wave (38GHz) interrogator transponder system for ground vehicles is capable of identifying friendly combat ground vehicles at ranges between 150 and 5500m ground to ground and 150 to 8000m air to ground in under 1sec.

The BCIS interrogation is activated automatically on activation of the shooter platforms laser range finder or interrogation button. It sends an encrypted directional query to the targeted vehicle, which, if fitted with BCIS will reply with an encrypted omnidirectional friend message.

A light and voice confirmation informs the gunner that the target is "friendly" or if no response is received, that the target is "unknown".

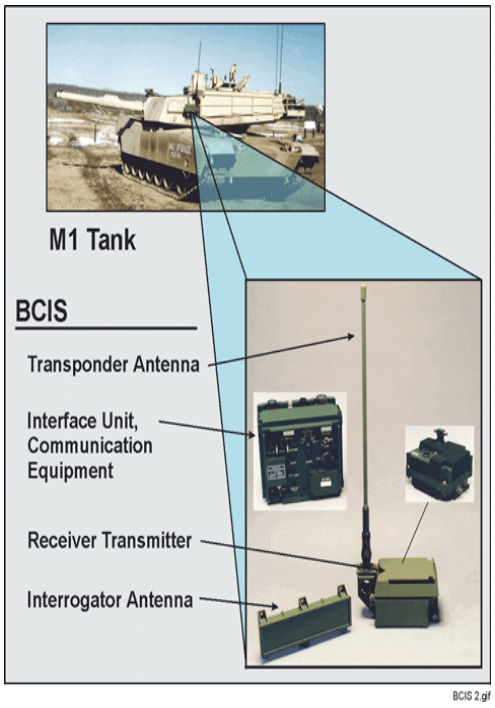


Figure 16.8: 38GHz battlefield IFF .

16.5.1. Combat Identification: The Future

The long range vision for combat identification is to enable the military forces to identify all targets in the battle space for all combat mission areas.

The fighting forces must be able to: rapidly and positively identify enemies, friends and neutrals with the primary purpose of lowering combat attrition, increase enemy losses and minimise the risk of fratricide.

The aspects of this vision that are purely transponder based are the following:

- BCIS Battlefield combat identification system
- BCIS(-) Lightweight handheld BCIS interrogator for use by soldiers
- (A-G) BCIS Air to ground BCIS
- CIDDS Combat Id Dismounted Soldier.(laser/RF based system for soldiers)
- SINCGARS GPS coordinate based "don't shoot me" net
- (A-G) SBCI The air to ground portion of the SINCGARS net

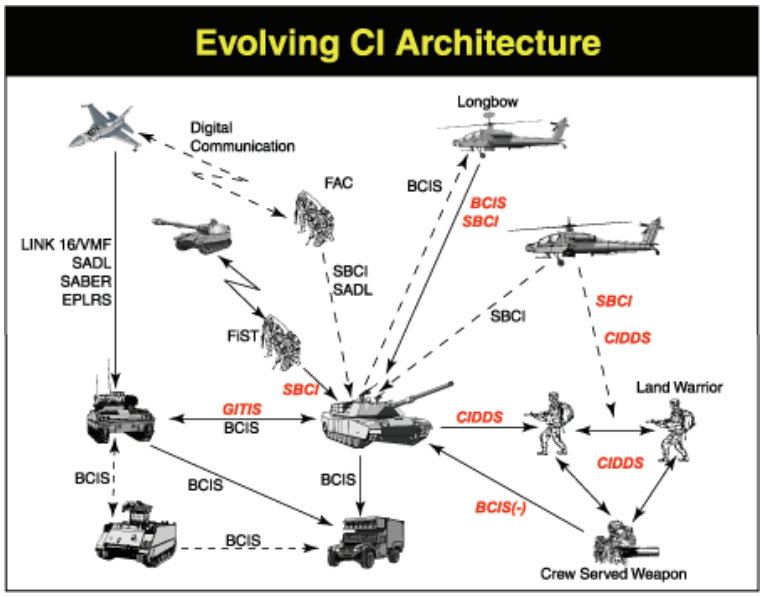


Figure 16.9: Battlefield IFF in the future

16.6.References

[1] Current trends in transponder systems, http://www.rapidttp.com/transponder/trends.html, 23/02/00.

[2] Range vs Power and frequency for passive electric coupled tags, http://www.rapidttp.com/transponder/freqpwr.html, 23/02/2000.

[3] Sentech tags and Labels, http://sentecheas.com/taglabel.html, 22/02/2000.

[4] Microwave ID pays off in the distribution chain, http://www.rapidttp.com/_transponder/confiden.html, 28/02/2001

[5] Battlefield combat ID system, http://www.sarda.army.mil/sard-zs/hti/hti_programs/bcis.html, 28/02/02.

[6] Science and Technology, http://peoiews.monmouth.army.nil/ci/webpg/cid8.html, 28/02/2001.

[7] RFID, A key to Automating Everything, Scientific American, Jan 2000

[8] Freeway Operation and Management Handbook, ops.fhwa.dot.gov/.../chapter8_02.htm

Chapter 17.

Tomography and 3D Imaging

17.1.Principle of Operation

The word tomography derives from the Greek word tomos meaning section, so the process of tomography involves the generation of narrow sections through an object. It is assumed that this is a non-invasive process which is performed using sensors outside the object of interest. In many applications sequences of measurements made in 2D are combined to produce a pseudo 3D image. The process is when applied to X-rays is referred to as Computerised Axial Tomography (CAT or CT).

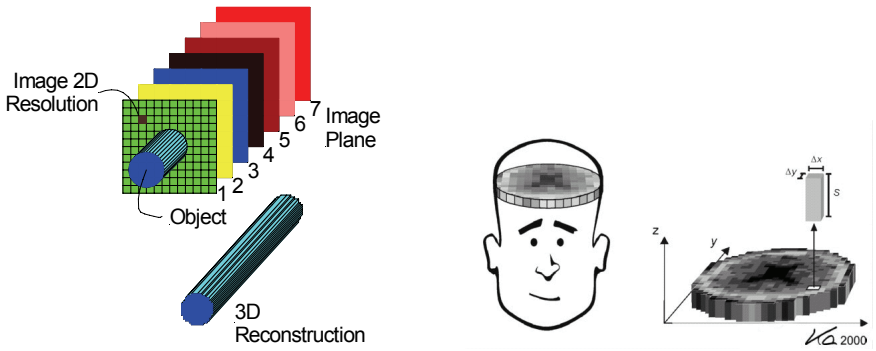


Figure 17.1: Computerised axial tomography (a) general principle and (b) as applied to a head

Computed tomography only became feasible with the development of computer signal processing capabilities in the sixties, but many of the basic principles were developed many years before. In 1917, a mathematician, J.Radon showed that the distribution of material, or the material properties can be determined if the integral values along any number of lines passing through that layer, are known.

Such processing is common in medical applications where doctors require a 3D representation when planning invasive surgery.

Examples of this include:

- 3D Ultrasound
- 3D Magnetic Resonance Imaging (MRI)
- Positron Emission Tomography (PET)
- X-ray CT Scans

Ultrasound technology has been discussed in earlier lectures. It uses the different elasticity characteristics of the target to produce an image while MRI imaging relies on the varying amounts of Hydrogen (in general contained in water) and PET Imaging relies on the release of positrons from radioactive decay.

The best understood, and possibly the most common is X-ray CT imaging which relies on target density variations that affect the attenuation of high frequency EM waves.

17.2.CT Imaging

Between 1957 and 1963, a South African physicist, A. Cormick, independently developed a method of calculating radiation absorption distributions in the human body, based on transmission measurements. He postulated that that it would be possible to display very small differences in the absorption through the body, but never applied his research.

In the early 1970s CAT scanning (also called computerised tomography, or CT, scanning) was conceived, yet again, by William Oldendorf and developed by Godfrey Hounsfield. Although often employing contrast media to enhance the quality of the images obtained, this test is barely invasive.

The most basic CT measurement developed in the 1970s, involves the rotation and displacement a collimated X-ray source (pencil beam) around the patient as shown in the figure below.

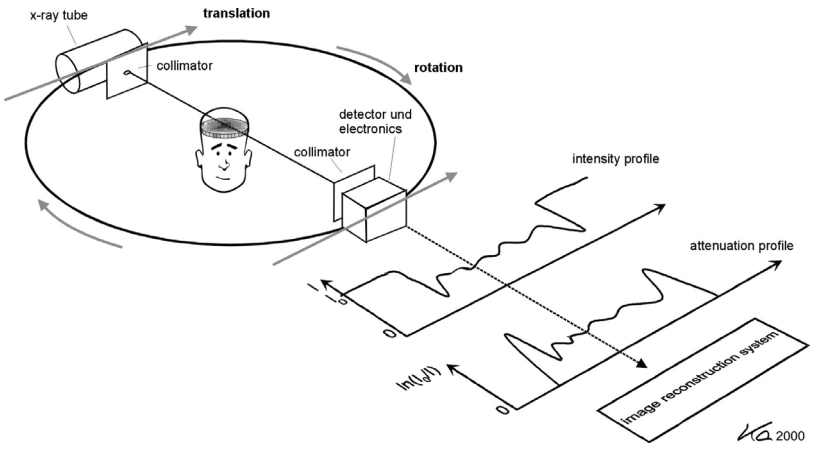


Figure 17.2: The simplest CT scan involves measuring the intensity of a pencil beam of X-rays from many different angular positions

To speed up the process, more modern CT scanners consist of an X-ray source that produces a fan-beam which penetrates the patient and impinges on a bank of detectors. This complete assembly rotates around a central core to produce a sequence of intensity measurements over 360°.

In either case, a complex image reconstruction algorithm solves this equations derived from this massive set of measurements to produce an image which represents the density of the material in cross section through the patient.

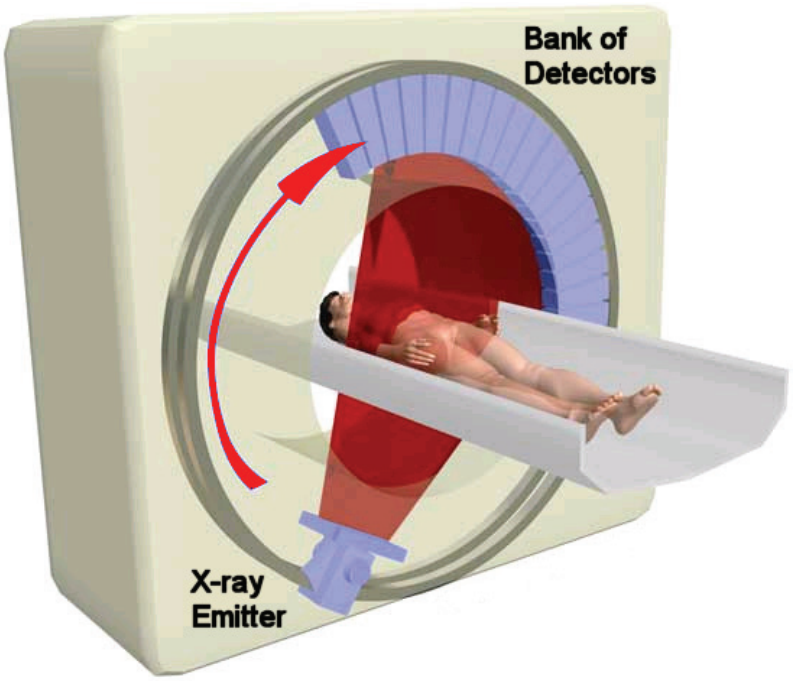


Figure 17.3: CT scanner schematic

The generation of clear images is compromised by movements of the patient during the scanning process, and so faster scan methods were developed. Continuously rotating CT systems that could image a complete slice in less than one second were first introduced in 1987. However, these systems still constructed 3D images from individual "slices" through the patient.

In the 1990's spiral scanners with multiple slice capabilities were introduced to produce even faster 3D high resolution images. The state of the art is a scan time of less than 0.5s with a slice thickness of between 0.5 and 1mm

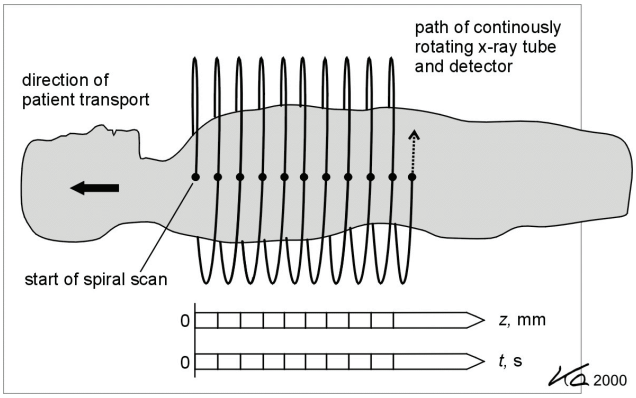


Figure 17.4: Spiral scan tomography

It allows visualisation of extremely thin sections of the brain, skull, spinal cord, and spine (as well as other parts of the body) in two dimensions and with enough clear distinction between black, grey, and white areas of the image to allow pathological diagnosis in many cases.

17.2.1. Image Reconstruction

As shown in Figure 17.2, measurements are made of the intensity of the X-ray beam and converted to a set of attenuation measurements. These are known as the "Radon Transform" of the image. An inverse transformation must be then carried out to determine the distribution of attenuations for each pixel $\mu(x,y)$ element through the target.

The easiest process to understand is one in which there are N^2 unknowns in an $N \times N$ matrix of pixels. If sufficient independent measurements are made, then it is possible to solve for all of the unknowns. In the simplest case of a target with only four elements, two measurements, from two projections will yield a system of four equations and four unknowns which can easily be solved. The extension to a 3×3 matrix with nine unknowns can also be solved easily using the twelve equations as defined schematically in the figure.

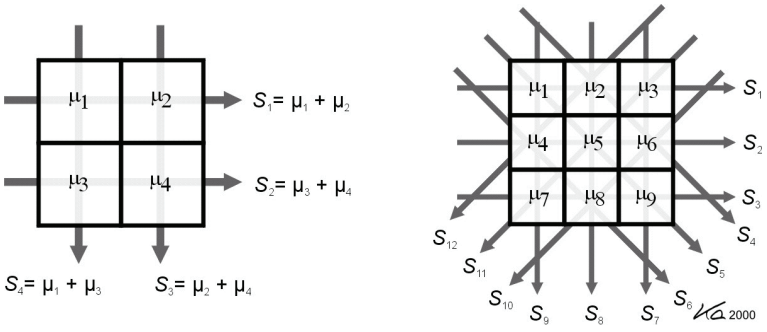


Figure 17.5: The solution of N2 unknowns for an N×N image matrix can be determined by solving a system of linear equations

This process was used by early CT scanners in which the number of elements was limited. However, with the requirement for finer resolutions and more pixels, the computational overheads became unacceptably high.

In modern CT scanners, the convolution- backprojection procedure is usually applied. It starts with a matrix loaded with zeros, and as each measurement is made, the projection value is added to all of the elements in the array along the direction of measurement. However, because each component does not contribute only to the value at the desired point, but to the whole image, it is clear that an unsharp image will result.

To minimise the magnitude of this unsharpening, each projection is convolved with a convolution kernel prior to backprojection. This convolution kernel represents a high pass filter.

Because convolution in the spatial domain is equivalent to multiplication in the frequency domain, it is possible to perform this convolution-backprojection process in that domain.

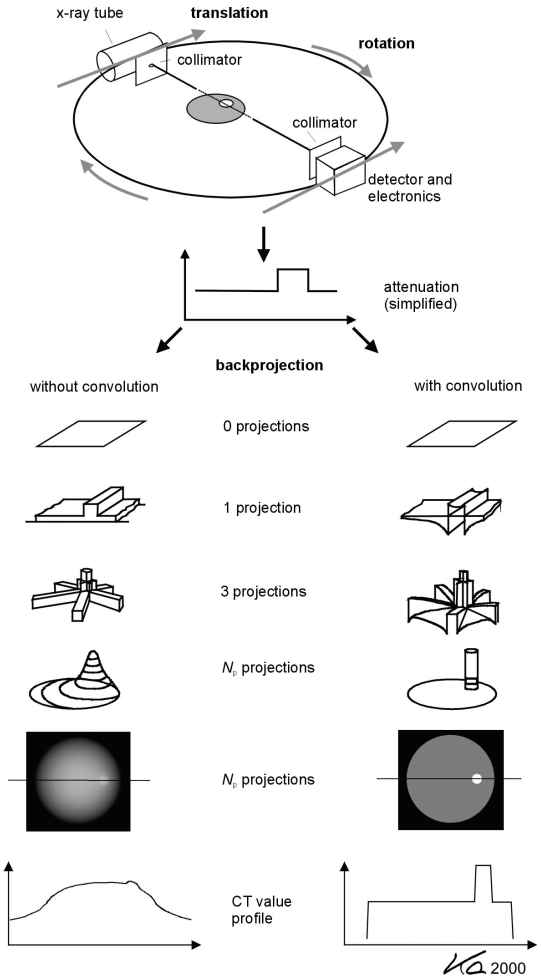


Figure 17.6: Image reconstruction by backprojection shows that convolution with a high pass kernel is required to ensure a sharp image

17.2.2. What is displayed in CT images

As explained earlier, the CT computes the spatial distribution of the linear attenuation coefficient $\mu(x,y)$. However, because μ is strongly dependent on the energy of the X-ray photons, it would make a direct comparison between images made with different systems impossible. Therefore, the image displayed as an attenuation relative to that of water. These are referred to as CT units, or sometimes Hounsfield Units (HU) in honour of the inventor.

$$CT_{units} = \frac{\mu_T - \mu_{water}}{\mu_{water}}.1000 \qquad (21.1)$$

On this scale, water, and consequently any water equivalent tissue with $\mu_T = \mu_{water}$ has a value of 0 HU, by definition. Air corresponds to a value of -1000 HU because μ_{air} is close to zero. Bone and other calcifications with high atomic numbers and high density offer increased attenuation and therefore have higher CT values – typically up to 2000.

Most medical scanners cater for a range of CT values from -1024 to +3071, a total of 4096 values, as these can be represented by a 12bit number. Because, neither the imaging software, nor the eye can discern 4096 different shades of grey, CT ranges of interest for a specific imaging task, are windowed and expanded to fill all shades from black to white.

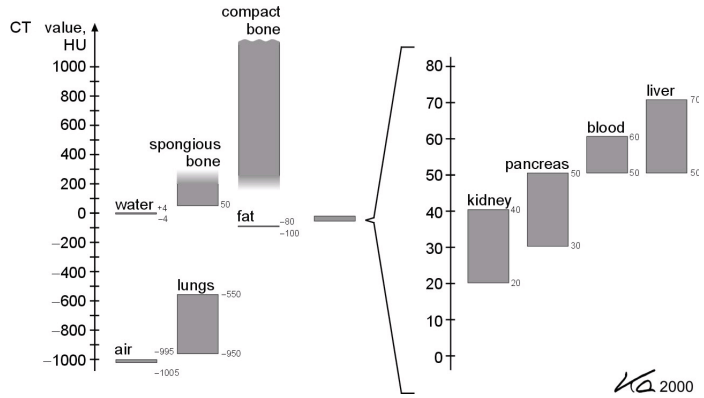


Figure 17.7: CT values for normalised attenuation of different materials

The contrast of a CT image is far superior to that of a conventional X-ray image, not because of the higher powers used, but because of the differences in the way that the images are generated.

In the figure below, a comparison is made between a CT image and an X-ray image. For the former a contrast between two arbitrary adjacent pixels is shown to be 28HU which corresponds to 50% in this scaling. The X-ray intensity will be determined by the sum of the attenuations along the whole path through which the beam travels. It can be seen in this example that the contrast between two adjacent paths through the cranium is only 0.23%

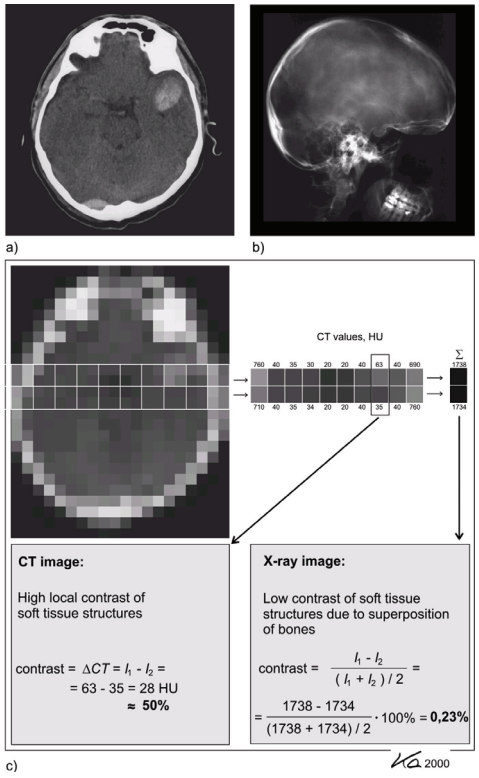
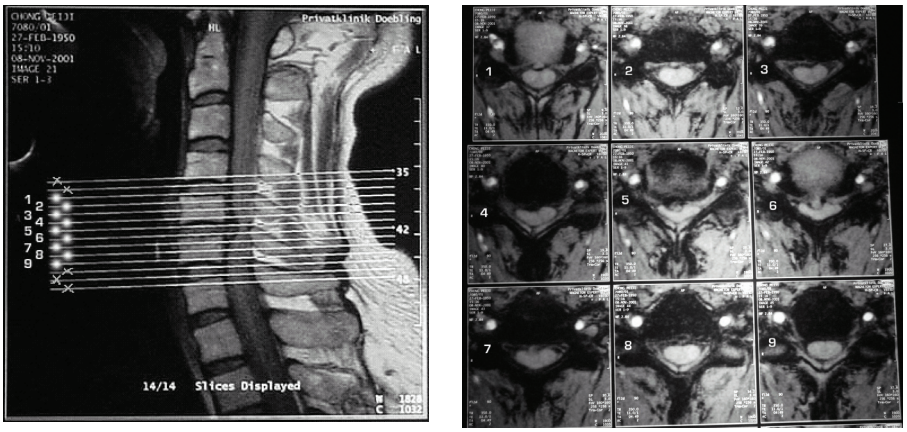


Figure 17.8: Contrast of CT and X-ray images

17.2.3. Two Dimensional Displays

CT Scanning is primarily limited to the transverse plane because of the mechanics of the scanning system. All of the other image planes are synthesized from the volume image constructed from these transverse slices.



(a) **(b)**

Figure 17.9:CT Image of the spine showing (a) longitudinal section indicating the slices and (b) individual images of each of the slices

17.2.4. Three Dimensional Displays

3D displays represent a scan volume in a single image which has generally been manipulated to enhance a specific characteristic. This can only be done successfully when one high-contrast structure (like the skeleton) is to de displayed. Shaded surface displays (SSD), maximum intensity projections (MIP) and volume rendering (VR) or perspective volume renderings (pVR) are some of the common manipulations

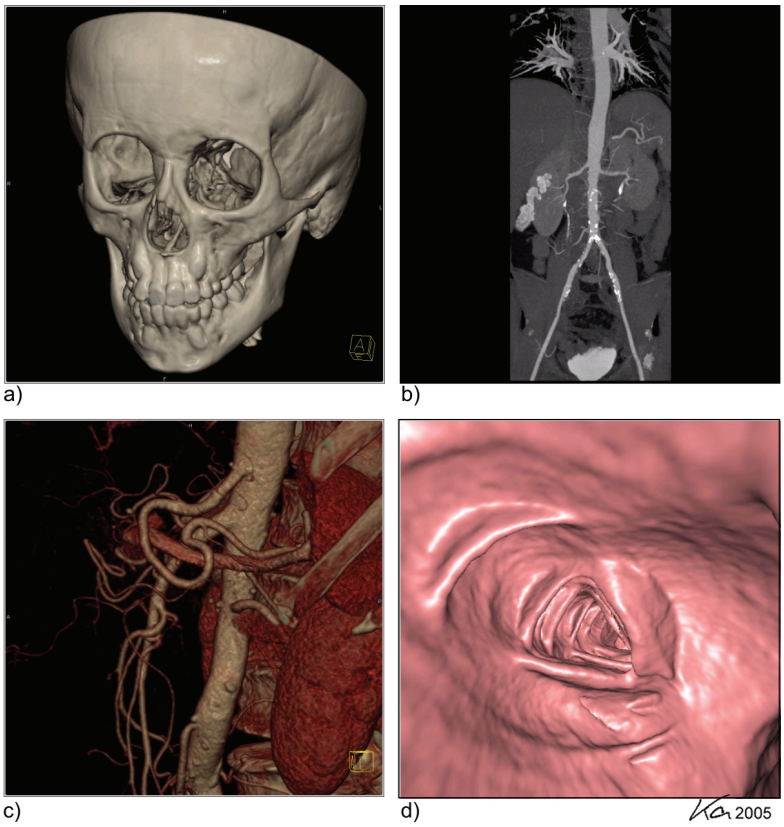


Figure 17.10: Different 3D display methods (a) SSD for skeletal structures, (b) MIP for CT angiography, (c) VR for abdominal structures and (d) pVR for a virtual colonoscopy

17.3.MRI-Magnetic Resonance Imaging

MRI imaging uses the principle of Nuclear magnetic Resonance (NMR) and an assembly of powerful magnets to produce images generated by various elements within the patient. It is most commonly used to image hydrogen as the human body consists predominantly of that element.

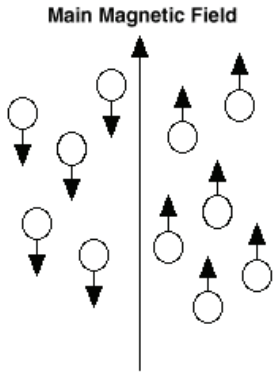
17.3.1. Nuclear Magnetic Resonance (NMR)

Atomic nuclei have an angular momentum arising from their inherent property of rotation, or spin. Since the nuclei are electrically charged, the spin corresponds to a current flowing around the spin axis which in turn generates a small magnetic field



(a) (b)

Figure 17.11: Comparison between (a) spinning top precession and (b) hydrogen atom precession



Each nucleus with an odd number of nucleons, and hence a non-zero net spin therefore has a magnetic moment, or dipole, associated with it. In general the orientation of the dipoles is random, but if they are placed in a magnetic field, they will become aligned to it.

Protons, or hydrogen nuclei, with a spin of ½ can align parallel to the field or antiparallel. The two orientations have slightly different energy (Zeeman splitting) with the spin up state (parallel) having the lower energy. Their effects almost cancel, but the lower state has a slight excess that is exploited by the MRI.

Figure 17.12: Hydrogen atoms align to an external magnetic field

When a magnetic field is applied, the whole population of nuclei has a bulk magnetisation vector M aligned to it. This is defined as the z-direction and by applying a small rotating magnetic field in the x-y plane, the nuclei can be tipped away from the z-direction. The field rotation rate is tuned to the natural precession frequency of the nuclei (hence magnetic resonance). For Hydrogen nuclei in a magnetic field of 1Tesla (10^4 Gauss), the Larmor relationship which defines resonant frequency as a function of the applied field predicts the resonance to be 42.57MHz.

Other materials will have different resonant frequencies in the same field (^{31}P) has a resonance at 17.24MHz and (^{23}Na) at 11.26MHz.

It should be possible to "tune in" to any specific species to observe their response in isolation. However, for imaging purposes, only Hydrogen is used in practise because its concentration is high and it is more sensitive.

As the power in the rotating field pulse is increased, the bulk magnetisation vector M will continue to rotate away from the direction of the static magnetic vector until it is orthogonal to it, rotating in the x-y plane only. When the pulse ends, the magnetisation vector continues to rotate for a time and in so doing, generates a small EMF in the field coils.

Interactions between adjacent nuclei that begin to spin at slightly different rates results in a loss of energy and causes the total spin to decay until it is once again aligned with the static field.

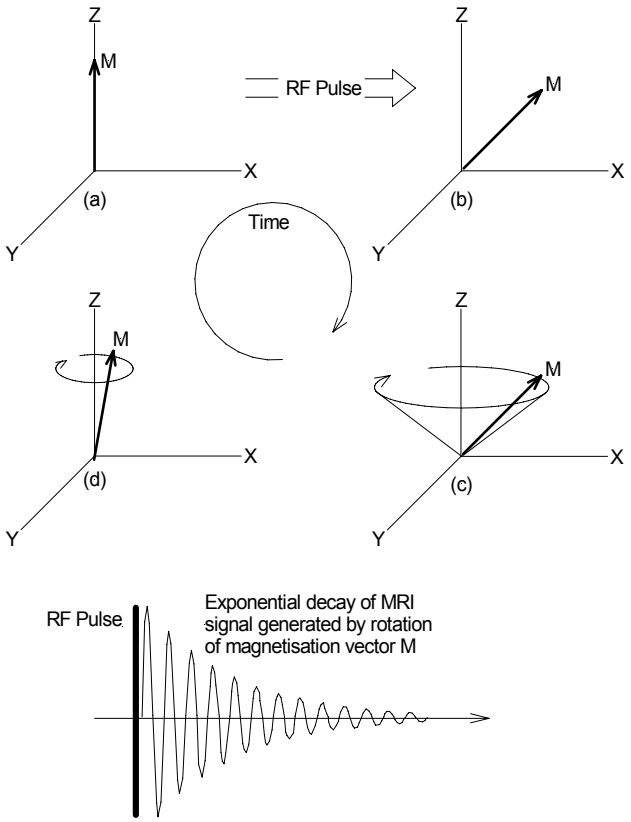
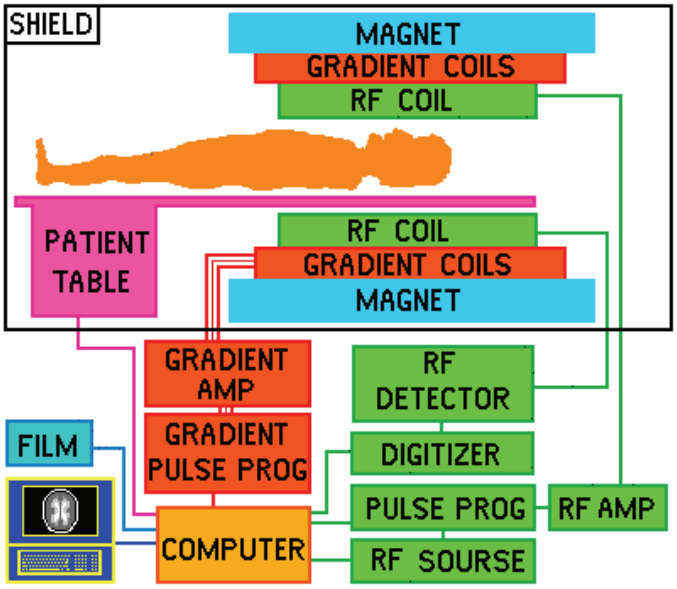


Figure 17.13: Principles of nuclear magnetic resonance showing (a) the unexcited magnetisation vector aligned with the external field (b) result of the application of an RF pulse to rotate the magnetisation vector (c)

17.3.2. Imaging Process



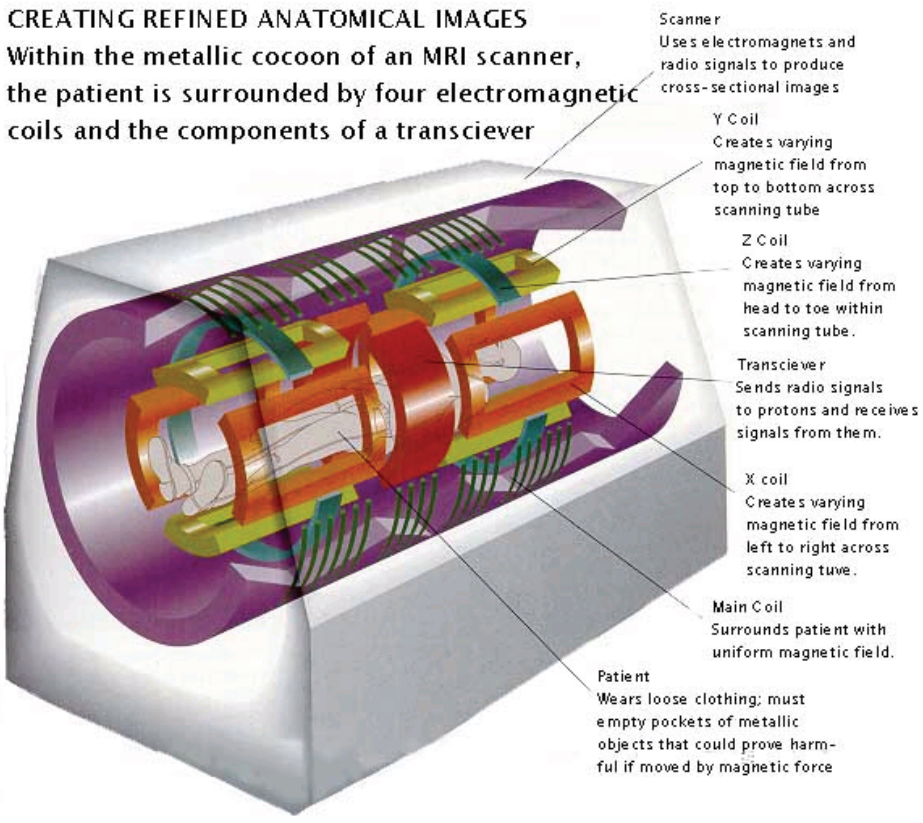


Figure 17.14: Schematic of a MRI machine

The imaging process exploits the time taken for the spin to decay, as this gives an indication of the proton, and hence water, density in the body

As can be seen in the figure above, in addition to the fixed magnetic field generated by the main coil, three gradient coils can apply varying fields in the x,y and z directions. Manipulation of these fields is used to identify the spatial position of the hydrogen concentration

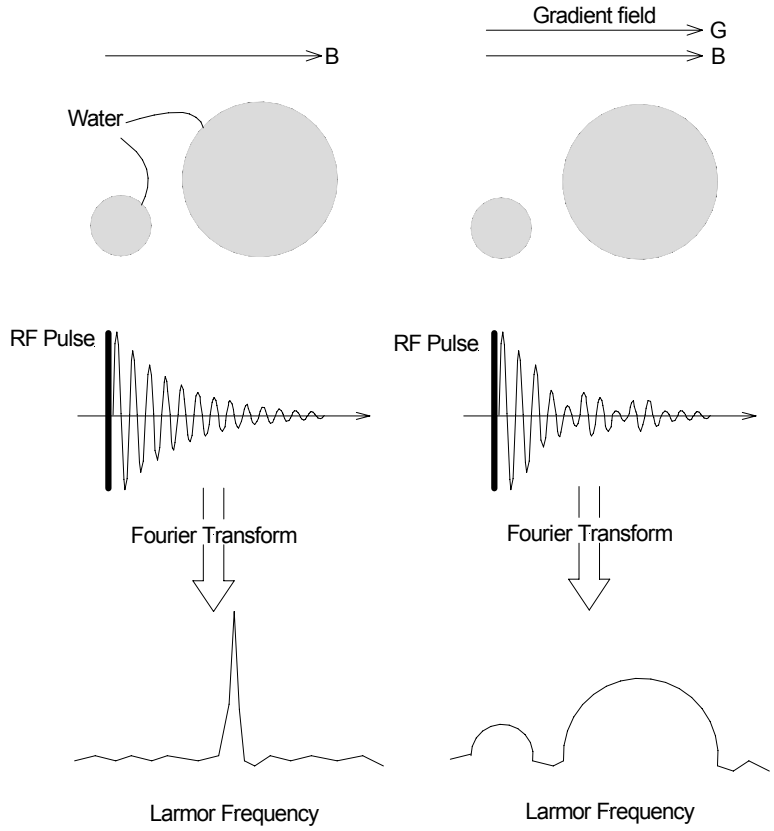


Figure 17.15: Spatial encoding of the MRI signal to produce an image slice

As shown if the figure, if no gradient field is applied along the magnetic axis, then the Fourier transform of the decaying Larmor signal shows a single broadened spectrum.

If a magnetic field gradient is applied, then the Larmor spectrum will have characteristics of the variations in proton density profile in the direction of the gradient. By rotating the gradient field and producing proton density profiles from various angles, a computer can perform the inversion function to recreate in 2D the internal structure of the body. This process (using the projection method) is illustrated in the figure below.

This use of proton density makes it possible to produce images of tissues that are comparable, and in some cases superior, in resolution and contrast to those obtained with

CT scanning. Moreover, since macroscopic movement affects NMR signals, the method can be adapted to measure blood flow.

Because the direction of imaging is determined by the magnetic gradient, it is possible to produce images in any plane; axial, coronal or sagittal.

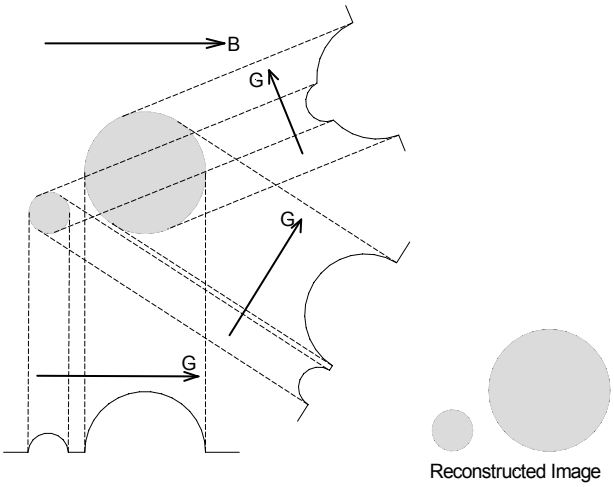


Figure 17.16: Image reconstruction using the projection method

17.4.MRI Images

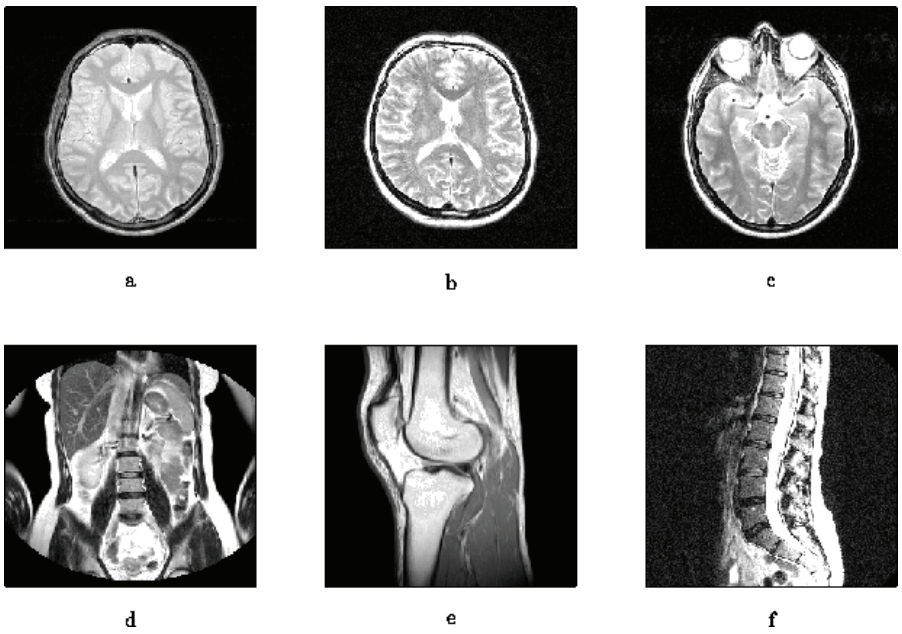


Figure 17.17: MRI images of various parts of the body showing its capability to produce images of soft tissue

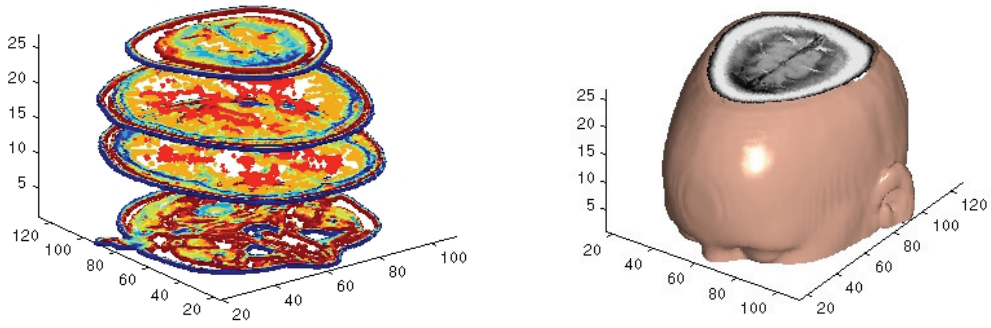


Figure 17.18: 3D MRI image recreation

17.5.Functional MRI (fMRI) Investigations of Brain Function

fMRI investigations of brain function rely on the fact that oxygenated and deoxygenated haemoglobin molecules behave slightly differently in a magnetic field. MRI images can then be made to show the oxygen use in the brain using the Blood-Oxygen-Level-Dependent (BOLD) signal

Images made using this technique, while the subject is observing a specific scene or thinking specific thoughts, are compared to those made by a control to determine where a specific brain function occurs

The image below shows the difference between the subject looking at a face and looking at a blank screen

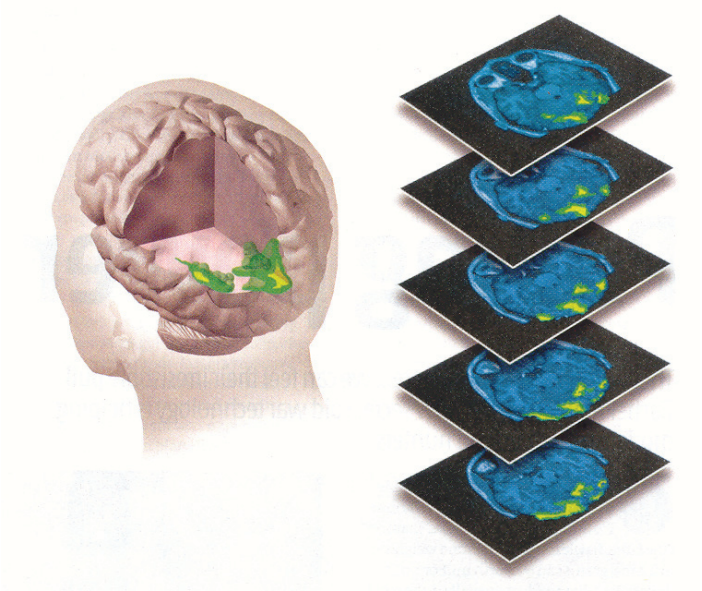


Figure 17.19: Functional MRI image of the areas of the brain used to recognise a face

17.6.Positron Emission Tomography (PET)

Radioactive molecules made from radioactive isotopes with short half-lives such as ^{11}C, ^{15}O or ^{13}N are injected into the body where radioactive decay takes place releasing a positron.

PET produces images of the body by detecting the gamma radiation – two photons each with an energy of 511keV that are created by the annihilation of a positron and an electron that occurs after this decay as shown in the figure below. These photons are emitted at very nearly 180° from one another.

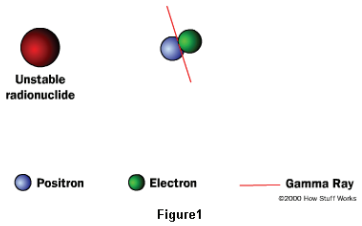


Figure 17.20: Emission of a gamma ray following annihilation of a positron and electron

The gamma ray detector used in a PET scanner consists of an annulus of scintillation crystals each connected to a photo-multiplier tube as shown in the figure below. If a gamma ray strikes one of the crystals, it produces a photon of light that is converted to an electron and amplified by the tube to produce a measurable signal. A coincidence detector ensures that only pairs of hits that occur within about 25ns are recorded

A coarse image is made of an axial slice through the body by combining the directions and relative time delays all of the measured events.

The scanner is moved and the process repeated until a 3D image has been built up.

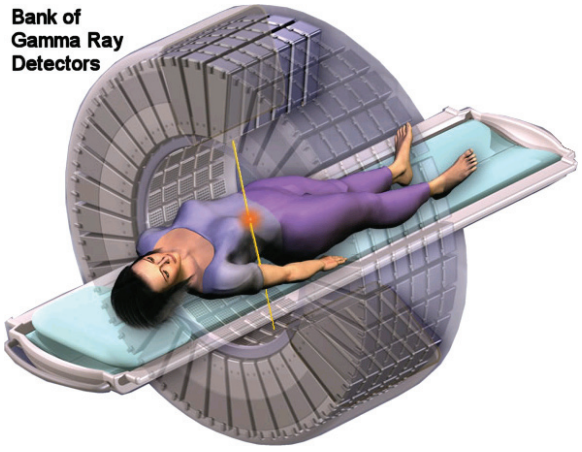
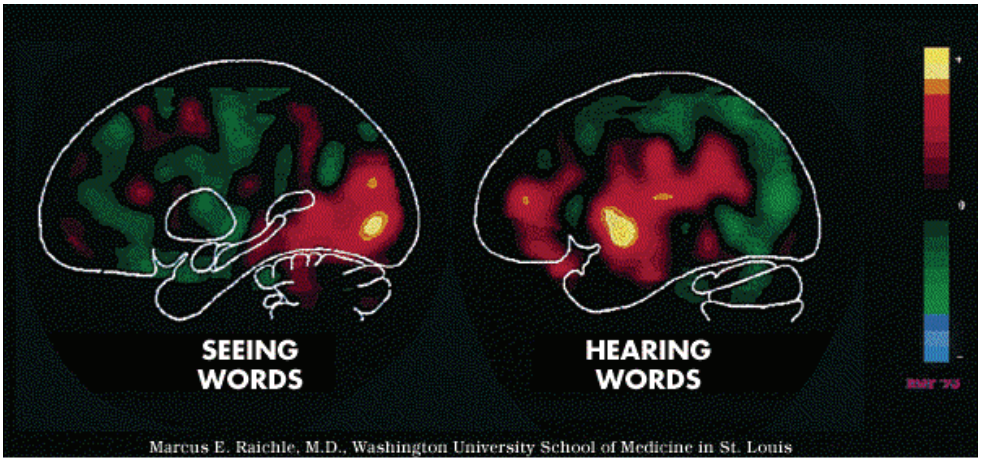


Figure 17.21: Schematic diagram of a PET scanner

Depending on the type of molecule injected, PET can provide information on different biochemical functions.

For example, if the molecule that is radioactively tagged is glucose, then the PET scan will show an image of glucose metabolism, or how much energy the body is using in a specific area. This has been useful in studies of brain function.

PET scans do not have the same high resolution as that of fMRI shown in the previous section, but they are more sensitive in identifying physiological processes that use chemicals that do not have a good MRI signature. For example, abnormal bone growth that can indicate the presence of a tumour can be pinpointed using radioactive Phosphorus.



The PET scan on the left shows two areas of the brain (red and yellow) that become particularly active when volunteers read words on a video screen: the primary visual cortex and an additional part of the visual system, both in the back of the left hemisphere. Other brain regions become especially active when subjects hear words through ear-phones, as seen in the PET scan on the right.

Figure 17.22: PET images showing brain function while words are processed in different ways

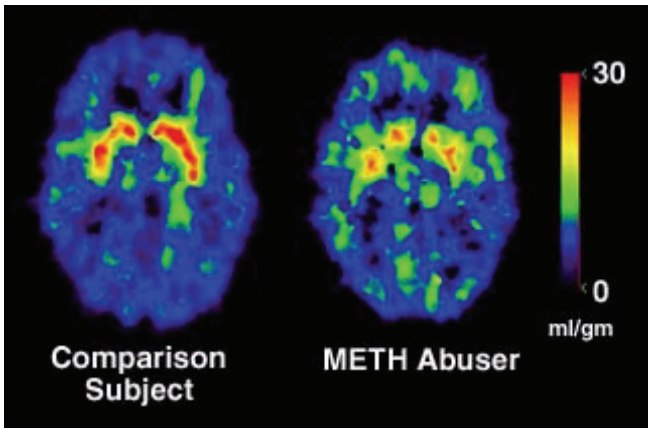


Figure 17.23: PET scan showing abnormal brain function of a METH user

17.7.3D Ultrasound Imaging



Figure 17.24: 3D ultrasound system

If the relative positions of the images made by a 2D-ultrasound scanner are known, then it is possible to build up a 3D image. In general, the transducer is still moved by hand as skilled Sonographers perform the function better and are less intimidating than mechanical positioners.

Some of the original work was done at UCSD on a suitably modified Acuson 128XP/10 with a C3 transducer. However, in the last few years a number of manufacturers have produced commercial versions of the 3D imager.

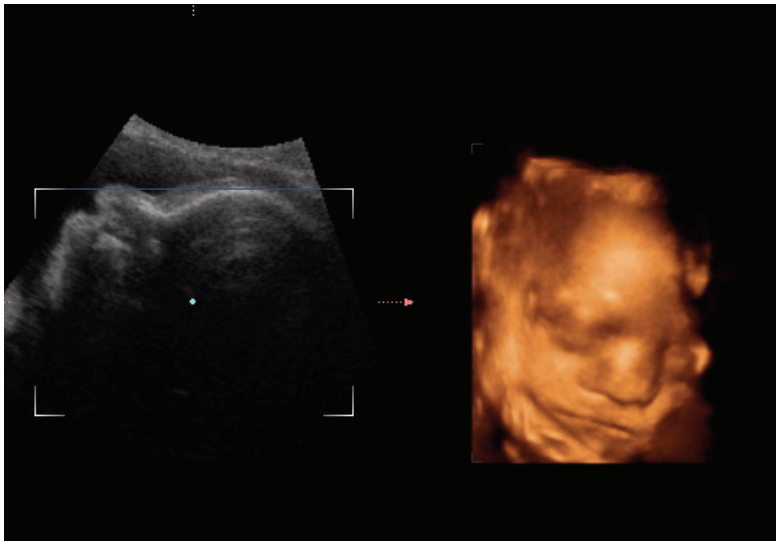


Figure 17.25:Process of converting a 2D ultrasound scan to a 3D image

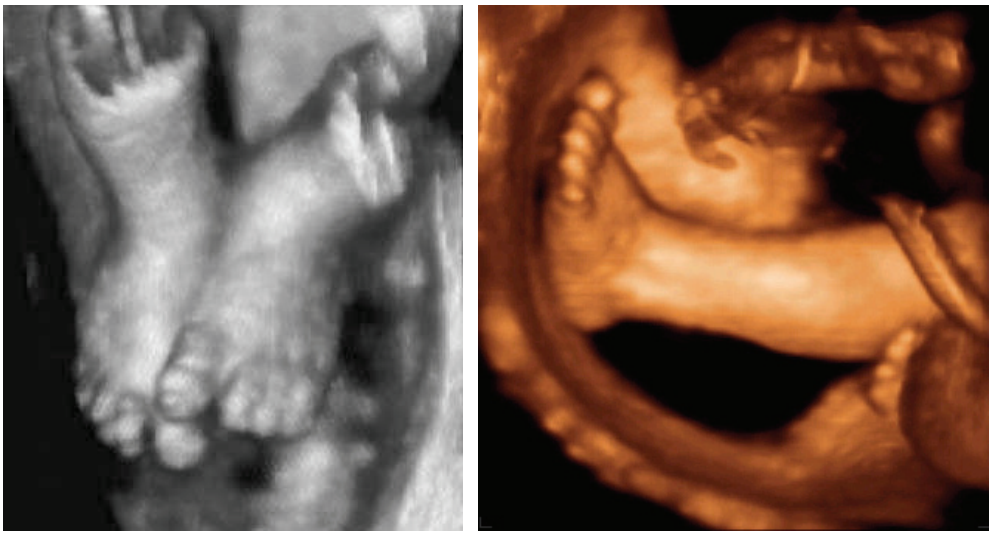


Figure 17.26: Feet



Figure 17.27: Hands and faces



Daily telegraph - Tuesday Feb 1 2001

Watching your baby smile in the womb

By TIM HILFERTY

NO longer do parents have to wait until their baby is born before they decide whom it takes after.

Three-dimensional ultrasound images, first tested in Australia, have improved at such a rate that parents can see their baby smile months before birth.

These amazing pictures show a baby just before and after birth.

The before picture, taken by the latest in 3-D ultrasound machines — the Siemens Omni — is as detailed and as clear as any baby snap in a family photo album.

But it may be some time before NSW parents have the option of viewing the face of their unborn child.

While the technology is slowly being introduced into hospitals in the UK, the US and Europe, so far only one 3-D machine is in operation in NSW and only on a trial basis.

Dr Henry Murray, an obstetrician at Nepean Hospital, said a machine was being tested there to study its potential uses.

The machine costs between $200,000 and $400,000.

They work by taking two-dimensional images from two different angles, and then creating one image using a sophisticated computer program.

Dr Murray admitted he felt a sense of wonder when he looked at the face of an unborn baby.

"You still feel that way, it is one of the joys of obstetrics," he said.

"You can get some pretty amazing images with the 2-D machines but this goes one better than that."

Where the machine does excel is in picking up facial abnormalities, and deformities in the arms and legs.

"It can show the surgeons what they are up against," Dr Murray said.

He said hospitals in some countries were using 3-D ultrasounds as a "novelty" for excited parents-to-be.

"That's not something we would encourage. Although there is no evidence that ultrasounds are harmful, we don't like doing them unless they are absolutely necessary."

Andrew Hartmann, managing director of Acuson Australia, which distributes ultrasound machines, said crystal-clear images of babies' faces were rare.

"Three things have to happen; the baby has to be reasonably still, it has to be facing the right way, and it has to be smiling," Mr Hartmann said.

He said the Omni machine, which took the US images, would probably not be exported to Australia.

Before . . . a scan of a baby in the womb before birth.

After . . . the same baby shortly after being born.

Figure 17.28: Image comparison: before and after

The scanning and processing speed of modern ultrasound machines is now so fast that manufacturers are offering movies. This is referred to as 4D ultrasound because it includes a time axis.

17.7.1. Ultrasonic Computed Tomography

Because the pulsed echo ultrasound process can only see tissue interfaces, an alternative is ultrasonic computed tomography which uses both the attenuation, and the propagation time to estimate both the attenuation coefficient, and the refractive index of the object

17.8.3D Sonar Imaging

Sophisticated 2D sonar arrays such as the one developed by Thomson Marconi Sonar (TMS) in Sydney can produce short range 3D images with voxol resolutions down to 1×1×1mm. In this system a group of 3 uniformly spaced transmitters illuminates the target with high frequency (>1MHz) sound pulses.

A sparse phased array made up of 84 tiles each made up of a random pattern of 32 hydrophone receivers receives the echo.

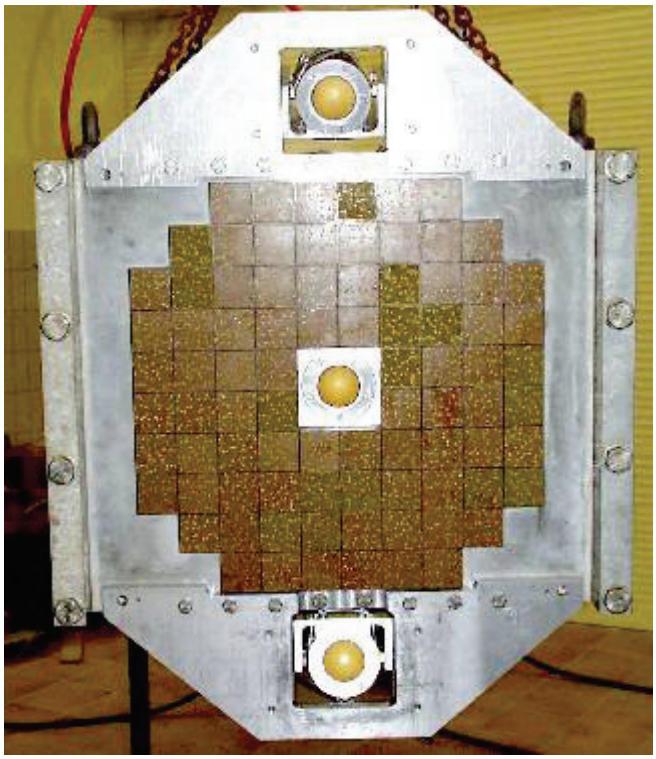


Figure 17.29: Complete non-scanned imaging sonar

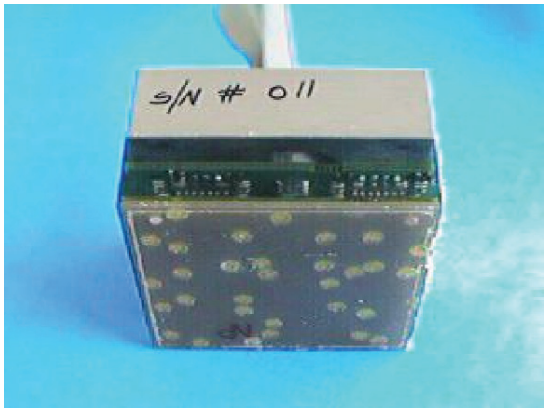


Figure 17.30: Individual 32 transducer receiver tile

After a single pulse, the received amplitude and relative phase information from each of the 2688 receivers is processed to produce high resolution 3D images such as those shown below.

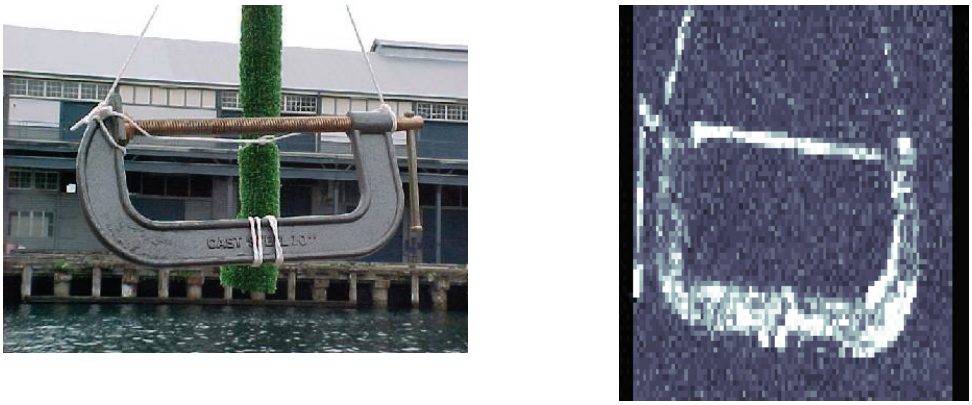


Figure 17.31: 2D projection of a 3D image of a G clamp

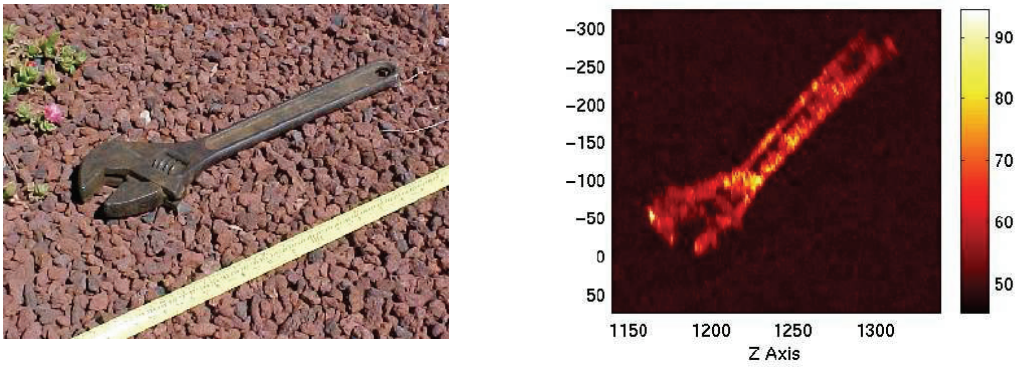


Figure 17.32: 2D projection of a 3D image of a monkey wrench

17.9.3D Ground Penetrating Radar

A 2D-image slice is made by dragging a standard GPR across the ground and recording amplitude and phase information from each position and then synthesising the image using SAR techniques discussed earlier

If a number of slices are produced side-by-side as with the axial tomography method, then a composite 3D image can be constructed

This requires that the position of the GPR be well registered using differential GPS.

The following images made by Dean Goodman in Japan show a circular tomb and contents.

The use of GPR for archaeological surveys eliminates the damage caused by exploratory trenches that were employed in the past.

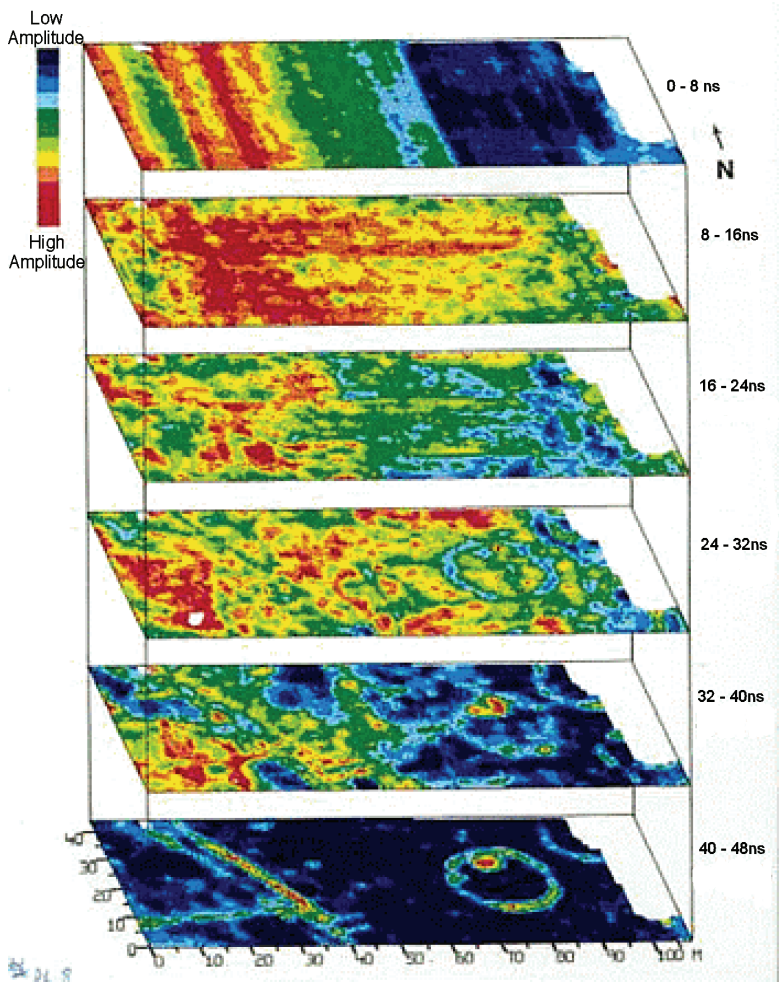


Figure 17.33: A pseudo 3D GPR image showing a 1100 year old circular burial moat adjacent to a more modern (600 year old) fence line

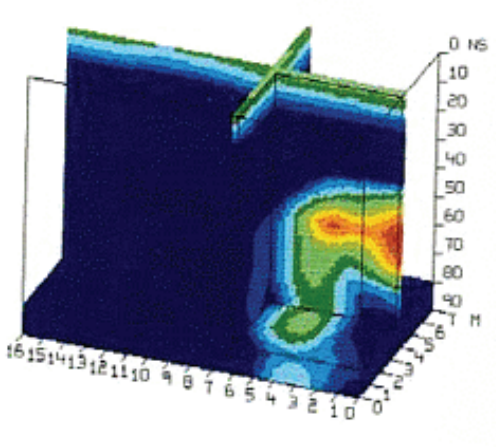


Figure 17.34: 3D cutaway of a burial chamber shows the main chamber and a vertical shaft that leads to an offering below the burial

17.10. References

[1] Burial Mound GPR, http://sciam.com/exhibit/062298radar/japan.html, 22/02/2000.

[2] Private Correspondence from Thomson Marconi Sonar (TMS)

[3] UCSD 3D Imaging Ultrasound, http://tanya.ucsd.edu/gallery.html, 23/01/2000.

[4] How Ultrasound Works, http://www.howstuffworks.com/ultrasound.html, 04/04/2000.

[5] How Nuclear Medicine Works, http://www.howstuffworks.com/nuclear-medicine.html, 04/04/2001.

[6] I. Paykett, NMR Imaging in Medicine, Scientific American, may 1982.

[7] Pet/CT Scanner, Time, December 04, 2000.

[8] The Mind Readers, New Scientist, September 21, 2002

[9] Griff Watson Art and Illustration, www.griffwason.com/gw_pages/glw-ct_scanner1.htm

[10] Philips medical Systems, www.newscenter.philips.com/.../ Index.html

[11] Medison 3d Ultrasound, www.hospitalmanagement.net/. ../imaging/medison/

[12] Kalender, W. Computed Tomography – Fundamentals, System Technology, Image Quality, Applications. Publicis, 2005.

[13] Kak, A and Slaney, M. Principles of Computerized Tomographic Imaging. Siam, 2001