Chapter 1.

Introduction

1.1. Objectives

This thesis addresses the development of a high-resolution long-range millimetre-wave imaging radar for autonomous airborne or land-vehicle navigation and guidance applications.

- The term *resolution* refers to both the angular resolution of the radar as determined by the antenna beamwidth and the range resolution as determined by the transmitted bandwidth. The definition of *high-resolution* depends on the context, and in this instance is deemed to be the resolution that is required to identify natural and manmade features useful for autonomous applications.

- Existing radar systems are generally configured either for extremely short-range indoor applications (typically less than 20m), or for medium-range outdoor applications (up to 300m). However, under certain circumstances, operational ranges of up to a few kilometres are required, which are defined in this thesis as *long-range*.

- The term *millimetre-wave* in the radar context, refers to electromagnetic radiation with a wavelength (in free space) of between 1cm and 1mm. This corresponds to frequencies of between 30 and 300GHz. For this thesis, the term refers to a specific band of frequencies within an atmospheric window centred at 94GHz.

- The definition of an *autonomous vehicle* is one that can navigate, without human intervention, through a completely unstructured and unpredictable environment to achieve a specific objective. In most cases, however, so-called autonomous systems operate under highly constrained conditions and fulfil only a small fraction of the tasks required for true autonomy. An example of this is intelligent cruise-control (ICC) which monitors the region in front of a moving vehicle, and controls the accelerator and brakes to maintain a safe following distance.

To achieve true autonomy, a vehicle must be able to sense its environment comprehensively over a broad range of scales. Objects in the immediate vicinity of the vehicle must be classified at high resolution to ensure that the vehicle can traverse the terrain. At this scale, individual rocks and small shrubs are important. At slightly longer ranges, individual trees and low branches must be resolved to allow for short-range path planning to occur. At long range, general terrain characteristics must be known so that the vehicle can negotiate around difficult or impassable obstructions. Finally, at the largest scale, the vehicle must be aware of the direction to its objective.

In this example of a ground-based vehicle, the ultimate resolution scale will probably be different from that of an autonomous aircraft which is not concerned with fine structure on the ground. However, because aircraft generally operate over longer ranges, the spread of feature sizes that must be resolved is similar.

The ultimate objective of this thesis is to develop an imaging radar that can operate at sufficiently long range with sufficient resolution to fulfil the obstruction detection and path-planning roles in a typical autonomous ground vehicle (AGV) and the navigation role in an unmanned air vehicle (UAV) with a payload capacity of less than 50kg.

In order to reach this objective, the thesis must address the following:

- Determine the appropriate operational ranges that must be achieved under all conditions including dust, rain, mist and operation during the day or at night for both the airborne and ground-based autonomous applications.

- Establish the range and angular resolutions required to identify important natural or manmade features that may be visible to the radar. It should be noted that these features may be different for airborne and ground applications and that they may also vary depending on what specific role the radar is fulfilling.

- Investigate different types of features to determine which are most suitable for these applications. It is important to consider these features in the context of their surroundings insofar as they are observable by the radar. For example, a corner reflector may be an appropriate feature when placed on the perimeter of a runway, but it is completely inappropriate in the middle of a forest.

- Features may be classified by their temporal and spatial characteristics as well as by Doppler and polarisation signatures in the electromagnetic spectrum. These should all be considered during this analysis.

- For a given operational range and required resolution, it is possible to determine both the antenna beamwidth and the radar bandwidth that are required. In addition, once the Doppler and polarisation properties that must be measured have been identified, then an appropriate radar front-end can be designed.

- Given both cost and space/mass constraints, this thesis must examine the available options and develop the appropriate radar and antenna hardware to perform the imaging function successfully.

- Because the vehicle will probably be moving, albeit slowly in the case of the AGV, this motion will have a major effect on the system design and performance. Most fundamentally, any imaging system must scan its beam over a known region, and to achieve this it must be able to compensate for, or at least measure the base motion that is perturbing this scan. In addition, many radar systems exhibit range-Doppler coupling which must be compensated for if the vehicle is moving at speed.

- As the vehicle moves, each feature will be visible for a limited period due to occlusion by other objects and the limited range and angular fields-of-view of the sensor. To ensure that the features remain in focus and are reproduced in full, the measurement and processing speed must be sufficiently high.

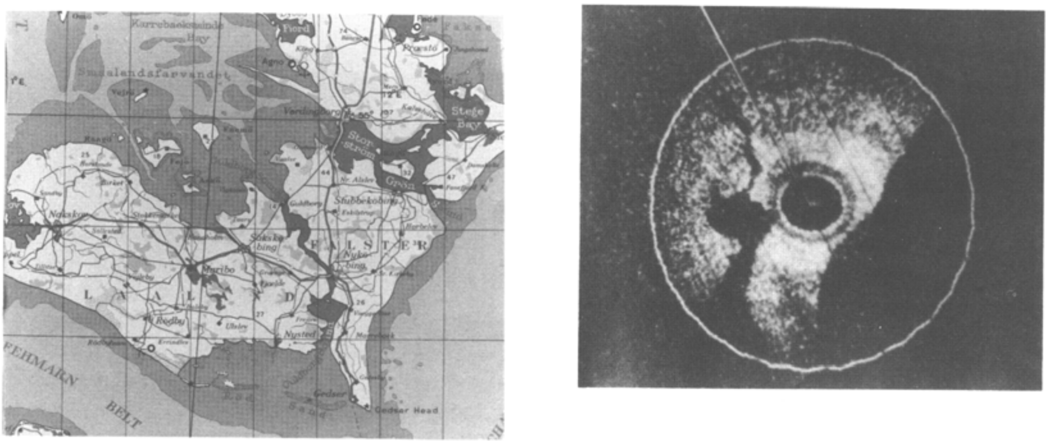
In summary, a radar front-end and its associated scanning, stabilisation and signal processing modules has been developed that is capable of generating long-range high-resolution multi-dimensional "images" under all weather conditions. The system is capable of operating from a moving airborne or ground-based platform. Finally, its outputs can be configured to satisfy the specific requirements of autonomous navigation and guidance.

1.2. Radar Imaging

The generation of a two-dimensional (2-D) reflectivity image involves the transmission of a narrow beam of electromagnetic radiation and measuring the time taken for each of the echoes to return. If the beam is displaced in angle by scanning, a 2-D image can be constructed by plotting the intensity of the echoes as a function of time (that equates to the range) and angle.

The use of reflectivity-based images to identify features on the ground stems directly from a World War II requirement to guide bombers accurately to their targets at night in inclement weather [45] p210. These bomb-aiming radars scanned a fan-beam antenna (one with a wide elevation beamwidth, and a narrow azimuth beamwidth) through 360° while transmitting a continuous sequence of short pulses through the antenna. The amplitude of the received echoes would then be reproduced on a cathode ray tube (CRT) with a circular scan pattern synchronised to that of the antenna, known as a P-scope or Plan-Position Indicator (PPI). The time delay from the transmission of a pulse to its reception determined the radial distance from the centre of the scan, and the amplitude of the echo, its size (no intensity modulation was available at that time). High-persistence phosphors were incorporated into the CRTs so that the laid-down image remained visible after the sweep had moved past as shown in Figure 1.1 [134].

An image is formed by this process because observations of the ground are constrained by the limited angular extent of the radar beam which determines the azimuth resolution of the image (2.3° in this case), and the short duration of the transmitted pulse which defines its range resolution. In this way a mosaic of individual snapshots of the ground-echo returns is built up in polar (R, θ) space to produce an image.



(a) (b)

Figure 1.1: Performance of the 3cm H_2S radar flying over the Guldborg Sound in December 1943 showing (a) a map of the area and (b) a photograph of the PPI display in which the interface between land and water is quite clear [134]

It is important to understand some of the differences between the image produced by a radar system and a photograph taken from the same vantage point. In the former, the image coordinates are range and azimuth (R, θ), while in the latter they are elevation and azimuth (ϕ, θ). The following nonlinear transformation maps between elevation and range,

$$R = \frac{h}{\sin\phi}, \tag{1.1}$$

where R is the slant range (m), h is the radar height above the ground (m) and ϕ the elevation angle (rad).

The result of this relationship is that a radar operating at low grazing-angles produces an image which looks like an aerial photograph taken downwards from a vantage point directly above the observation point, but illuminated by a source at the observation point [98] p29. This is well illustrated by the image in Figure 1.1 which shows a radar image made at 21000ft (6400m) and a map of the same area for comparison.

In the subsequent 60 years since this image was made, advances in antenna technology, waveform design and signal processing have produced radars with sub-metre resolutions combined with operational ranges of several hundred kilometres. Side-looking airborne radar (SLAR) was the first of the high-resolution imaging techniques, but it relied on the real aperture of an antenna running along the axis of the aircraft to obtain the required narrow azimuth beamwidth [67] pp693-701. This technology has all but been superseded by synthetic aperture radars (SAR) which are now commonplace in airborne applications [98] pp104-105 with an example of the output shown in Figure 1.2, in orbit around the earth [98] pp99-104 and even for mapping other planets and moons in our solar system [4], [7]. Interferometric techniques usually based on displaced antenna pairs can now produce 3-D images of the surface [98] pp359-380, [186].

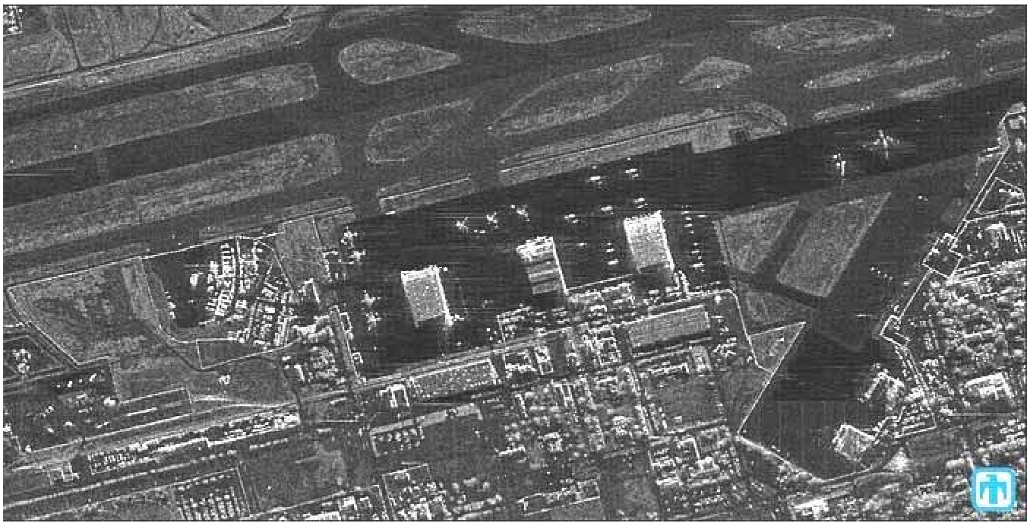


Figure 1.2: Airborne SAR image showing hangars around the Kirtland AFB, Albuquerque, New Mexico, (1m resolution) [10]

The quality of the real-aperture images produced is a function of the range and angular resolutions, the type of terrain, the altitude (grazing angle) and the direction from which the target area is viewed. It is also a function of the dynamic range of the receiver and signal-processing circuitry as well as the ability of the display to reproduce this dynamic range.

By scanning a pencil-beam radar in a spiral or by generating a raster either mechanically or electrically using a phased array, it is possible to extend the two-dimensional intensity images to three dimensions. Each return as defined by the unique coordinate (R, θ, ϕ) can be encoded by occupancy only, as shown in Figure 1.3, or by intensity, polarisation and any other measured characteristics.

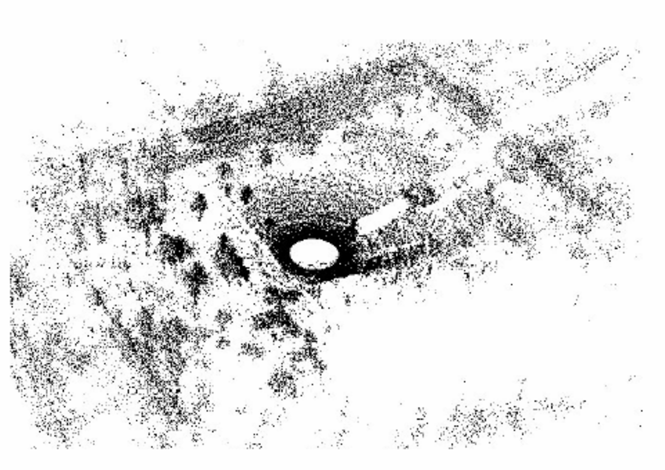


Figure 1.3: Beamwidth-limited occupancy image of a tree-lined quadrangle adjacent to the ACFR at the University of Sydney

1.3. Radar Feature Characteristics

In 2-D images the most striking and easily identifiable terrestrial feature is the interface between land and water. The smooth surface of calm water tends to reflect energy away from the radar, and so appears dark in an image [174] p26.6, [121] pp171-225, [133] pp260-306, [98] p35 while the rougher ground scatters the energy in all directions, including a significant backscatter measurable by the radar, and so appears lighter [68] pp138-152. Built-up areas are particularly good at redirecting radiation back towards the radar because the interface between a vertical building and the horizontal ground acts as a giant corner-reflector.

Depending on the ground-surface roughness and the operational wavelength, paved surfaces of runways and roads usually backscatter less energy than the surrounding terrain at shallow and even at moderate grazing angles [184] pp263-286 as shown in Figure 1.2. Even though the latter are sometimes too narrow to be directly resolvable, they are generally lined by telephone lines, fences, embankments and trees which produce a bright trail of returns on a radar display.

With the exception of the land/water interface, manmade objects are generally easier to recognise than natural features. Mountains in particular, though they produce relief-map like images generated by brightly reflecting near slopes and dark far slopes, vary greatly with the angle of incidence [98] pp37-39. Consequently the identification of a particular mountain in an extensive range is difficult. That notwithstanding, radar images were used in World War II to aid with navigation through the Alps [133].

An unbroken forest-canopy can generally be distinguished from grassland because there are variations in the scattering topology in the former that produce large variations in the amplitude of reflected signals. Grassland is usually more uniform and produces a more homogenous return [68] p154. A typical savannah in which scattered trees punctuate the otherwise uniform grassland produces returns rich in spatial features. These will show the characteristic high intensity border on the illuminated tree/grass interface generated by the corner effect of the tree trunks and the grassland, and a low intensity return due to shadowing on the opposite edge.

Trees make interesting targets because, depending on the foliage density, some penetration occurs even in the millimetre-wave band and the returns include some measure of the internal structure. This is an important consideration particularly with regard to path planning.

This discussion has considered only the mean value of the ground reflectivity whereas in reality, because each measurement consists of the vector sum of all of the scatterers within the resolution cell, the amplitude fluctuates over a wide range. This is particularly important in the millimetre-wave band where the spatial variation is best described by a log-normal distribution, as opposed to the better known Rayleigh distribution [152], [67] pp220-221 found at longer wavelengths.

1.4. Radar Systems Requirements

An autonomous vehicle that uses ground-based features for navigation or path planning must produce radar images of sufficiently good resolution and covering a sufficiently large area to be unambiguous in its operational context. For planning AGV paths at long range, tree and building sized features should be resolved. Using Table 1.1 as a guide, it can be seen that this requires a resolution of between 5m and 10m (a nominal 7m will be used as the design specification for this parameter). The required area for successful feature association as well as the time required for a UAV to align to an observed waypoint dictates that the minimum operational range of the radar be at least 3km.

Table 1.1: Radar resolution requirements for feature identification

Feature to be resolved	Required resolution
Coastlines, large cities, mountains	150m
Major highways, variations in fields	20-30m
City streets, large buildings, small airfields	10-20m
Vehicles, houses, small buildings	2-3m

For an airborne application using a small unmanned airborne vehicle (UAV) a practical operational height is about 100m, making the grazing angle 2°. For ground-based operation, it can be assumed that the rolling nature of the terrain will be such that, if visibility extends to 3km, the radar will be looking out over a valley, and that the grazing angle will be similar. A shallow grazing-angle has been selected as it accommodates a fairly narrow elevation beamwidth. This in turn allows for a high antenna gain while simultaneously minimising the backscatter volume, which is important for operation in rain. These advantages compensate for the relatively low ground reflectivity. Grassland at this grazing angle has a reflectivity which is of the order of -20dBm^2/m^2 (see Appendix B) making the radar cross section[1] about -3dBm^2 for a 7×7m resolution.

The system must be capable of operating in inclement weather: fog, mist or light rain. In addition it must be able to operate through dust or smoke and at any time of the day or night.

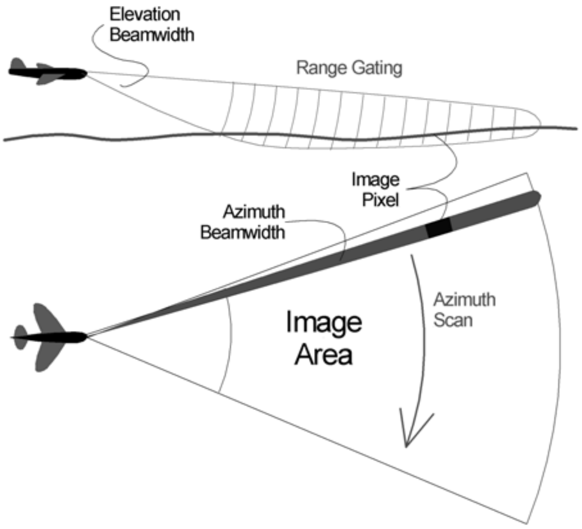


Figure 1.4: Plan and cross-section diagrams showing the long-range image generation process using a sector scan and a range-gate limited sensor

Radar systems operating using a wide variety of different modulations and scanning mechanisms have been used for imaging applications, but with the exception of those

---

[1] Radar cross section is the product of the resolution cell size and the reflectivity of the ground

used in some modern cruise missiles [5] and some UAV programmes [182], few of these have been developed specifically for autonomous navigation and guidance.

The most basic technique is to use a rotating fan-beam antenna to scan a pulsed transceiver in azimuth as shown in Figure 1.4. This technique is still used in most low and medium-cost maritime radars [166] and for airport traffic monitoring [84] where magnetrons still provide the required high peak-power pulses. However due to their relatively large mass, high-voltage power supply and poor reliability in the millimetre-wave band, magnetrons are not suitable for small, relatively low-cost applications.

Pulse-compression techniques that increase the average power transmitted without sacrificing range resolution must be utilised to achieve the required range performance. A number of these including various continuous and pulsed, chirp-based and digital compression systems are considered. It is shown that most of these would be capable of achieving the 3km and 7×7m requirements under certain circumstances, though the standard continuous wave configurations would be impractical due to antenna isolation requirements. An investigation of the applicable techniques shows that the best performance at the lowest processing cost can be obtained by an interrupted FMCW radar system.

The image generation process requires that the beam is either physically scanned using a movable antenna or mirror, or that it relies of the forward motion of the platform to perform this function. Though there are a number of advantages to the latter technique, it is only really feasible for a UAV due to base-motion considerations. For an AGV application a more conventional mechanically-scanned configuration is required.

1.5. Navigation, Guidance and Path planning

The modern history of autonomous guidance starts with the V-1 “buzz-bomb” and V-2 rockets which were developed by the Germans and deployed against allied cities with devastating effect starting in June 1944 [185], [91]. Since then a large number of ever more sophisticated systems have been developed for military applications. Truly autonomous stand-off weapons such as the Prometheus-Apache [64], [155] and the shorter range Brimstone anti-tank missile [3] are examples of the state-of-the-art in millimetre-wave radar-guided missile technology. These missiles are only autonomous insofar as they will navigate to a previously designated target and identify it using radar returns. However, the new generation of stand-off weapons under development will be able to loiter above a battle-field while scanning the ground with sophisticated radar sensors waiting for targets of opportunity to appear [91]. The radar systems under development for these applications use increasingly sophisticated hardware and signal processing to obtain the required performance. DERA and Alenia Marconi Systems have developed a radar demonstrator for a UAV with a spatial resolution of 0.5m at a range of 4km. It relies on a combination of wide-bandwidth pulse-coded modulation (PCM) and Doppler beam-sharpening (DBS) techniques to achieve this performance [28].

The development of AGVs, due to the more complex nature of the environment through which they have to navigate, has been less dramatic. The first vehicles followed buried wires [74] or painted lines [181] but this restriction was soon removed by the introduction of vehicles that could localise accurately by measuring their positions with respect to beacons installed at strategic locations [27], [77]. This innovation allowed trajectories to be altered easily, and could accommodate some obstacle avoidance behaviour, but it still required the installation of sophisticated infrastructure.

Development of navigation techniques based on natural features removed the requirement for any constraining infrastructure and freed the AGV to operate over a larger area. The introduction of sophisticated laser and radar-based sensors that could identify features [124], [125] and sophisticated feature-association algorithms facilitated the maturation of this technology. The main shortfall in this case was the requirement to generate *a priori* maps of the area within which the vehicle would operate, an activity that was both tedious and prone to error.

In an autonomous application, image features must be identified and their coordinates extracted and compared to those stored in a database as part of the process of localisation. Alternatively, if an aerial photograph, or another radar image of a region exists, then 2-D image correlation can be used to perform this function. It should be stressed that this process can be applied even when the radar images are obtained at very shallow grazing-angles from a ground vehicle.

The ultimate goal of autonomous navigation is to allow the vehicle, be it an AGV or a UAV, to enter completely unknown terrain and to construct its own map while simultaneously localising, a process known as Simultaneous Localisation and Mapping (SLAM). The first SLAM algorithm to provide qualification of map convergence [177] formed the basis for all subsequent feature-based algorithms. Practical confirmation of the performance of the technique in small environments followed [127], [144]. However for large environments the $O(n^2)$ computational requirement made implementation too slow for real-time operation.

In the last few years, progress has been made in addressing this problem. Culling of redundant features has been used [72], improved estimation algorithms have been developed [96] and feature-space sub-division in various ways has been tried [188], [56], [15].

The accuracy of these techniques depends on the image resolution, its distortion, and the number and relative positions of unique features that are available for correlation.

As AGVs have become more mobile, path-planning algorithms have become more sophisticated, though most are still based on occupancy grids [79], [124], [82]. These are used to generate potential fields, often on the fly [192], within which the vehicle can navigate to its destination. For outdoor applications, occupancy-grid based algorithms are not particularly effective where a trade-off between traversability, risk and reward is required. Better characterisation of the terrain is the key to solving this problem, and it is well addressed by the information-rich returns provided by millimetre-wave radar images. These are useful in identifying both terrain gradient and the surface condition (rocks, trees, grass etc.).

1.6. Contributions

This thesis forms a part of a wider project aimed at developing high-performance autonomous systems. The objective of the wider project is to produce both AGVs and UAVs that can be used to develop navigation and guidance techniques suitable for completely and partially autonomous operation.

A number of radar systems have been developed at the Australian Centre for Field Robotics (ACFR) and at other research institutions worldwide that are capable of performing both 2-D and 3-D imaging functions at short range [124], [86], [57], [42]. However, until now, no one has developed a radar system, specifically for autonomous applications, that is capable of performing these functions at long range.

This thesis aims to address this omission, by documenting the development and application of a high-resolution, long-range millimetre-wave imaging radar based on frequency modulated interrupted continuous wave (FMICW) principles.

- Unlike conventional pulsed and linear chirp pulse-compression radar systems, there is no single matched filter for the FMICW process because it is dependent on the target range. This is exacerbated in the imaging application where the range-span is extended beyond that of a point-target. One of the primary contributions of this thesis is the development and implementation of a radar configuration that accommodates both this extended range, as well as constraints to the required resolution, the extent of the transmitted chirp and processing bandwidth (see Chapter 5). This contribution is supported by a thorough analysis of both theoretical and practical aspects of FMCW radar as a background to the FMICW work (see Section 4.5 and Appendices D to F).

- In as far as this thesis is an amalgamation of both radar and autonomous guidance considerations, its primary contributions to the latter are two techniques that use radar images to generate vehicle navigation offsets to sub-pixel accuracy (see Sections 6.3 and 6.4), and a novel implementation of the gradient-descent algorithm for autonomous path planning using an imaging-radar generated terrain reflectivity map (see Section 6.5).

- Some contributions are of a more practical nature concerning the radar implementation. In isolation, the individual elements do not necessarily constitute “new” work, but as a whole they make up an implementation that is both unique and extremely effective.

- One final, important contribution of this thesis is an analysis of the predicted effects of mutual interference between various common radar types that may be used in autonomous applications (see Appendix B). Once again, the novelty of this work lies not in the individual elements of the analysis but in its whole.

1.7. Thesis Structure

Chapter 2 examines the fundamentals of radar imaging including angular and range resolution, the process of range gating and the construction of both range-gate and beamwidth-limited images. This brief introduction leads on to a description of how a range-gate-limited 2-D image may be created by scanning a fan-beam antenna over a surface to produce a radar-reflectivity image in polar space. It is shown that, if the polar image is over-sampled into Cartesian space, the image quality remains good.

The concept of a beamwidth-limited 3-D image, in which a pencil beam is scanned in both azimuth and elevation to produce occupancy images is then introduced. For navigation purposes such images sometimes take on just three states (occupied, empty or unknown), though, for radar images it is common to define the probability of occupancy using probabilistic methods such as particle filters or sums of Gaussians.

The remainder of the chapter examines the use of radar imaging in both indoor and outdoor navigation and guidance applications as well as introducing radar-based "synthetic vision" for use as an aircraft landing aid in poor visibility conditions.

Chapter 3 briefly examines some of the interactions between the radar signal and the target in an imaging context. Because there is a dearth of published target signature data in the millimetre-wave band it includes a section describing the implementation of a 94GHz reflectivity measurement system, and some of the results obtained for various terrain types. Interestingly, these data confirm the log-normal distribution of spatial variations in the reflectivity reported in the literature, but finds that the mean reflectivity of our measurements to be lower by about 10dB at low grazing-angles.

Because of their coherence and short observation times, radar images are often plagued by speckle. This chapter demonstrates, using measurements made from a helicopter, that integration of a few images from slightly different perspectives improves the image quality considerably.

The second half of the chapter begins by introducing the range equation and summarises some of the effects on propagation of millimetre-wave radiation. Foliage penetration and multipath phenomena, effects which are of critical interest to an AGV are discussed with details available in Appendix A.

Finally, in conjunction with Appendix B, the effects of mutual interference between radar systems operating in the same vicinity are discussed. It is shown that due to the limited frequency-band available, and the wide bandwidth required for high-resolution imaging, the probability of interference is high, and more importantly, certain classes of interference are difficult to overcome.

Chapter 4 examines the basic principles of a number of millimetre-wave radar systems in terms of their generalised ambiguity functions and operational range capabilities, with a view to selecting the appropriate system for the long-range imaging application. Much of the background information concerning the design and implementation of a pulsed radar operating at 94GHz is included in Appendix C.

Because FMCW is the technique of choice for most autonomous applications, it is examined in some detail (with practical aspects of implementation discussed in Appendices D to F). Of particular importance are practical methods to optimise the range resolution given specific oscillator characteristics and details of the effects of oscillator phase-noise and leakage on the radar imaging performance. Outputs from a simulation model of the radar that incorporates both the linearity and phase-noise effects are used to illustrate their effects on the range spectrum.

This chapter shows that only the pulse-compression systems would be suitable for operation out to 3km and that the most suitable of these is FMICW.

Chapter 5 develops the theoretical background underpinning a constrained “matched filter” for an FMCIW radar. The FMICW process involves transmitting a chirp signal modified by the transmitter gating sequence. This signal, after the round-trip delay, is further modified by the receiver gating sequence before it is mixed with the un-gated chirp, and the ambiguity function determined. Mathematically, this is equivalent to taking the cross response-function of the receiver and the transmitter gating functions before convolution (in the frequency domain) with the response function of the standard FMCW waveform. This produces the response function of the FMICW waveform.

It is convenient to evaluate the performance of this modulation in terms of the power spectrum of the received signal as this allows multiple targets to be introduced, and it produces a graphical output (for zero Doppler) which is easy to interpret.

A comparison is made between the classical, deterministic, FMICW modulation and one with a pseudo-random time weighting of the interrupt sequence. It is shown that in both cases, small targets are eclipsed by artefacts of the modulation mechanism. Other alternatives considered are to process each gating cycle independently but none shows any significant advantages as regards the resolution of two closely spaced targets. It is shown that by processing and integrating the results of six cycles non-coherently, the overall processing time is reduced by a factor of 2.6 with a sacrifice of only 1.2dB in SNR when compared to coherent processing of the complete waveform by a single FFT. On these grounds, this technique was chosen for implementation.

The remainder of the chapter discusses practical aspects of the implementation and testing of this FMICW radar. Testing showed that the transmitter’s power amplifier broadens the transmit spectrum with a resultant degradation in the range-resolution, but that otherwise, the performance is good, with a closure in measured echo-amplitude to within 1.3dB of that predicted.

Development of the associated signal-processing hardware and software needed to implement the range-gating technique posed quite a challenge because of the wide received signal bandwidth. An innovative coarse/fine gating technique based on a stepped down-converter followed by a complex FFT is implemented with good effect to produce a spectrum containing 384 overlapping gates spanning the range between 1.5 and 3km.

As the content of each of the gates corresponds to the received power from a specific area on the surface of the earth, defined by the extent of the range gate and the width of the antenna footprint, this signal level corresponds to the mean reflectivity of the surface at that point. By stepping the antenna azimuth over 30° in 0.23° increments to produce 128 overlapping spectra, a 2-D polar image of the ground reflectivity is constructed.

Chapter 6 opens by introducing the hardware required for airborne imaging. Stabilised 3-axis gimbals maintain the correct orientation of the fan-beam antenna as it is scanned across the surface of the earth in a controlled fashion to produce a reflectivity image of the ground. To perform this function effectively from an aerial platform requires that the base-motion be measured, and the antenna roll, pitch and heading be adjusted to compensate. An aircraft grade INS/GPS along with the gimbals performs this function.

To produce the best quality images for both machine and human interpretation, integration of 3 or 4 images spaced about 200m apart makes a good compromise between speckle reduction and blurring (caused by INS drift). This process involves the registration of each of the 384 pixel-long range spectra onto a Cartesian frame "pinned" to the surface of the earth.

In this application, autonomous navigation is achieved with the help of an *a priori* generated reference-feature map. As the helicopter approaches the target, the integrated image is processed to extract similar features before a feature-association algorithm determines the offset between the two maps. Once the helicopter has landed at a surveyed point, its known position is compared to that estimated by the navigation system. The results show that the measured variances produced over a six sortie sample is superior to those obtained from the INS/GPS.

A second process using 2-D correlation of the radar image and a template made up of a small number of blob and line features is evaluated off-line. Using this technique, offset errors as small as 0.01 pixel (≈40mm) can be measured. Scale and rotation errors can be accommodated by dilating the template-features slightly.

The final sections of this chapter discuss the overall navigation strategy for an AGV and investigate a number of novel gradient-descent algorithms that use the radar reflectivity maps for path planning. These are shown to be quite effective in guiding the vehicle to a goal along iso-reflectivity contours as far as possible, and in extracting it from local minima.

Chapter 7 provides a summary of the work done in terms of the contributions made. It goes on to summarise the work in progress and proposes a number of new radar imaging and navigation projects that should be undertaken in the future.

Chapter 2.

Radar Imaging Principles

2.1. Introduction

This chapter introduces some of the concepts needed to understand the radar imaging process. It starts by discussing the characteristics of a radar system which determine both angular and range resolution before considering the process of building different kinds of images from radar data that can be achieved by scanning in different ways. Because radar image interpretation is such an important component of the navigation process, Section 2.2.3 briefly introduces this topic. In conclusion it describes some of the work that has been undertaken by other research institutes that have developed millimetre-wave radar systems for autonomous applications in the past.

2.2. Resolution

Resolution is one of the critical parameters that determines the ultimate quality of the images produced by the radar system from the environment in which the radar is situated. In general it is described in terms of the three measurement coordinates of polar space, namely, elevation angle, azimuth angle and range.

2.2.1. Angular Resolution

The angular resolution is determined by the characteristics of the far-field pattern of the radar antenna which, for imaging applications, is usually described in terms of the half-power (3dB) beamwidth. This is in turn determined by the dimensions of the antenna aperture and the wavelength

$$\theta_{res} = 1.22\frac{\lambda}{d}, \tag{2.1}$$

where θ_{res} defines the angular resolution of a circular aperture (rad), λ the operating wavelength (m) and d the antenna aperture (m).

Symmetrical "pencil beam" patterns generated from circular apertures are most often used in beamwidth-limited imaging applications, but for most 2-D range-gate-limited applications an asymmetrical "fan beam" pattern is required. This is generally achieved by reducing the elevation aperture, shaping the primary reflector [174] p10.5 or using a combination of the two methods.

In an imaging application where constant power returns are required from the ground independent of the range, the elevation gain $G(\phi)$ of the antenna should conform to

$$G(\phi) \propto \operatorname{cosec}^2(\phi)\cos^{1/2}(\phi). \tag{2.2}$$

The second term is often neglected for low grazing-angle applications with the resultant pattern as shown in Figure 2.1.

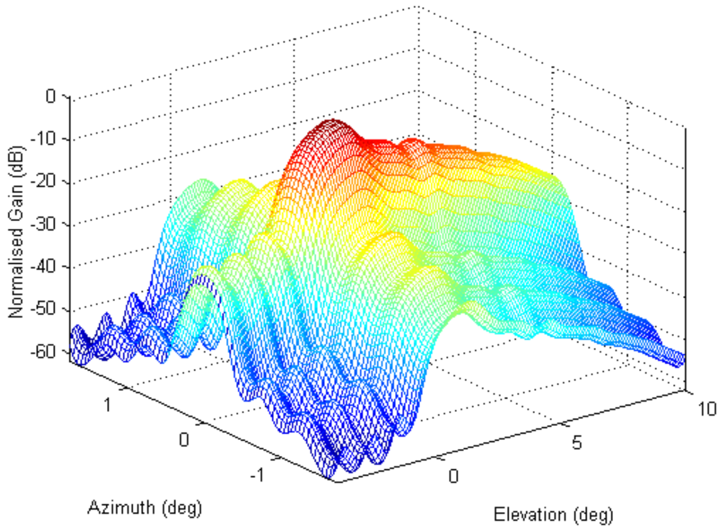


Figure 2.1: Three dimensional Cartesian plot of the gain pattern of a fan-beam antenna showing the narrow azimuth beamwidth and the cosec^2 elevation beam pattern

A uniformly illuminated circular aperture produces the narrowest possible beamwidth as described by (2.1). However, it also generates sidelobe levels that are only 17.6dB lower than the main-lobe peak, which is unacceptably high for most imaging applications as ground return variations can exceed 100dB [67] p700. Reduction of sidelobe levels can be achieved by tapering the aperture illumination as discussed in Appendix A. For Hamming illumination of a circular reflector, the theoretical sidelobe levels decrease from -17.6dB to -45dB, though in reality, most millimetre-wave radar systems do not achieve such low values because of spillover, blockage and surface tolerance effects [106].

2.2.2. Range Resolution

Range-resolution of a radar is most easily discussed in terms of the transmitted pulse-width. Figure 2.2 shows bursts of microwave energy radiating from the aircraft antenna at the speed of light until they strike an object (generally referred to as a target) that scatters the energy in all directions. A small fraction of this is scattered back toward the radar antenna which now plays the part of a receiver. Because the pulse travels to the target and back, the measured range R (m) is determined from the round-trip time, T (sec), and the speed of light, c (m/s)

$$R = \frac{cT}{2}. \qquad (2.3)$$

Similarly, the range-resolution of the radar, δR (m) is equal to half the pulse width, τ, converted to metres

$$\delta R = \frac{c\tau}{2}. \qquad (2.4)$$

If the distance between two reflectors exceeds half the pulse width, then two separate echoes will be seen by the radar receiver. If however, the separation is less than that, the echoes will merge into a single pulse which will be interpreted by the radar receiver as a single target. This is of course a gross simplification of the process and a more accurate description can be found in Section 4.3.2.

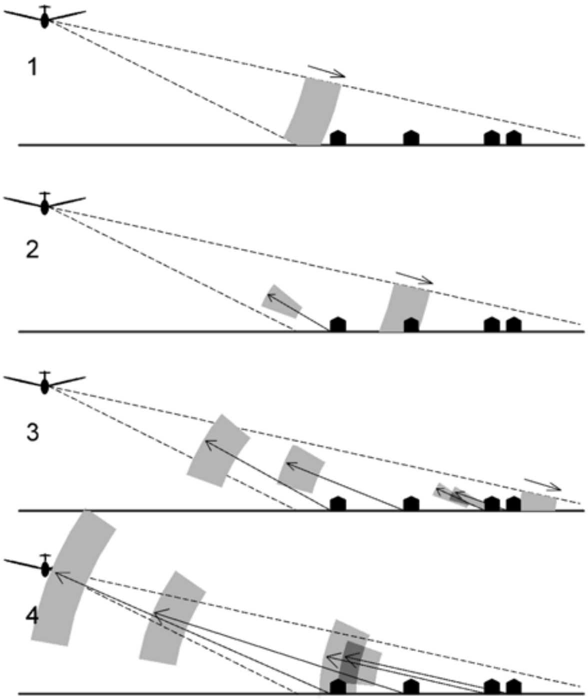


Figure 2.2: Operational principles of a pulsed time-of-flight radar illustrating the round-trip time to the target and range resolution due to overlapping pulses

In pulse-compression and FMCW radar systems, the effective range-resolution is decoupled from the pulse duration by using phase or frequency modulation to increase the transmitted bandwidth. A receiver with the appropriate matched-filter processing produces an effective (or compressed) pulse-width equal to the reciprocal of the transmitted bandwidth. The relationship between the range resolution of the radar and the bandwidth, β, is

$$\delta R = \frac{c}{2\beta}. \qquad (2.5)$$

Pulse-compression techniques are discussed in more detail in Chapter 4.

In the primitive radar systems discussed in Chapter 1, an image was created by modulating the intensity of a CRT directly using the amplitude of the echo signal. For modern applications, the received signal is first sampled and digitised in a process called range gating (see Appendix A) prior to further digital processing and display.

2.3. Building Images

2.3.1. Range Gate Limited Images

If the range-gating process is repeated continuously as the radar antenna scans in azimuth, then a 2-D mosaic of stored samples that map directly onto the reflectivity of the surface is produced.

It is clear from Figure 2.3 that though the range resolution δR is constant for the whole scan, the cross-range resolution, δXR, becomes worse as the range increases because the imaging process relies on a constant beamwidth antenna

$$\delta XR = R\theta_{res} \,. \tag{2.6}$$

As image processing in polar space is generally not practical, particularly if the frame of reference is moving, modern systems usually map the polar-image pixels into Cartesian space. This is most often achieved by selecting a Cartesian grid with a resolution equal to the cross-range resolution at the smallest operational range and then over-sampling at longer ranges.

The most simple method to allocate data to the Cartesian image plane is to transform the coordinates (R,θ) of the centroid of each polar pixel into Cartesian coordinates (x,y), and use those to load the nearest Cartesian pixel. However, this method often results in unfilled pixels in the Cartesian plane which are distracting, so the alternative, less efficient, method of clocking through all of the Cartesian pixels in the image and allocating to each the magnitude of the nearest polar pixel is performed instead. Ideally, partial pixel overlaps are integrated into their respective bins in proportion to the percentage overlap.

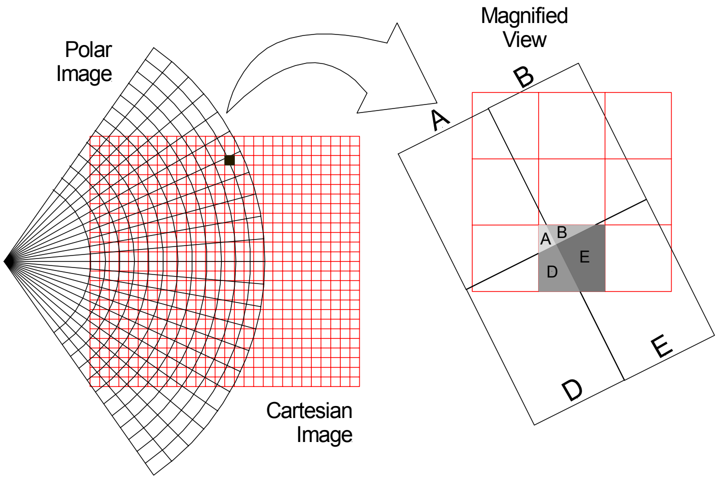


Figure 2.3: Polar to Cartesian image transform shows an example where each Cartesian pixel becomes the weighted average of the polar pixels with which it overlaps

In the example shown in Figure 2.3, the signal level in the illustrated Cartesian pixel is calculated from the weighted average of the polar pixels A,B,D and E in proportion to their respective overlapping areas.

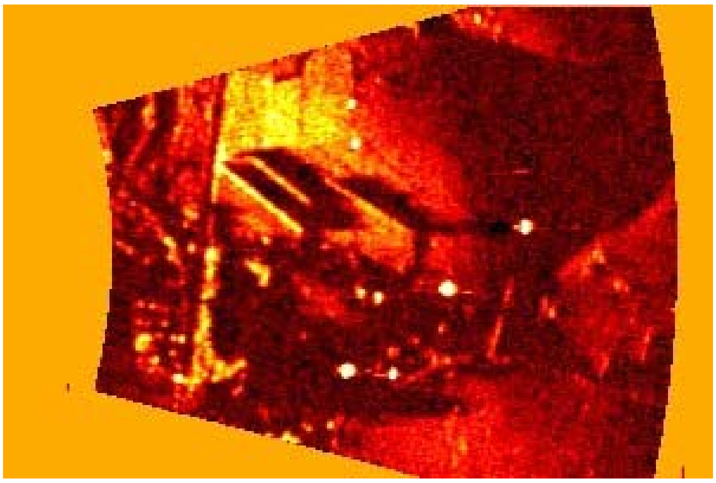


Figure 2.4: Range-gate-limited image of the ground transformed from polar to Cartesian space illustrating the good image quality obtained by over-sampling and pixel weighting

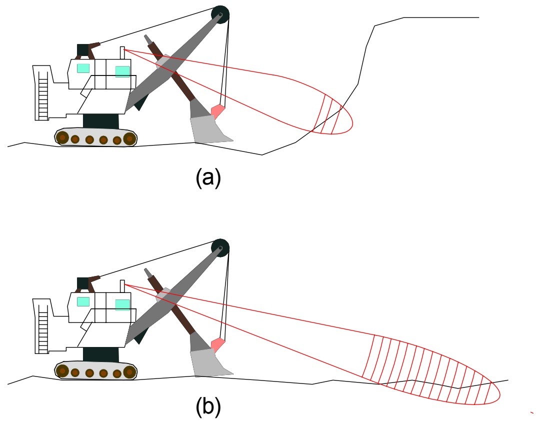
Obviously the process of over-sampling and pixel weighting cannot improve the actual resolution but it does result in a clear image with a minimum of sampling artefacts as the example in Figure 2.4 shows.

This is a 300×200 pixel image made using a millimetre-wave radar system that is described in detail later in this thesis. Each of the pixels represents an area on the ground of 8×8m which is slightly larger than the theoretical range-resolution of the radar which is close to 5m, but much smaller than the cross-range resolution. The actual resolutions are represented graphically by the thin red lines towards the bottom left and bottom right of the image, these are 20m and 40m respectively.

2.3.2. Beamwidth Limited Images

The difference between range-gate width (or pulse width) limited and beamwidth-limited imaging is formally defined in Appendix A and is illustrated in Figure 2.5.

In the beamwidth-limited case, the echo returns are mostly constrained within a single gate, and so the image resolution is determined by the angular extent of the beam and the range-gate extent. In the range-gate limited case, the image resolution is determined by the azimuth beamwidth of the radar and by the range-gate extent. The illustration shows that the slope of the target surface plays an important role in determining which of the modes would be most practical.



(a)

(b)

Figure 2.5: The difference between (a) beamwidth and (b) range-gate-limited imaging illustrated in this example of a shovel-mounted imaging radar

It can be seen that in the beamwidth-limited case, target reflectivity is not critical as long as a target is detected, because a 3-D surface can be constructed through the detected points. In the range-gate limited case, on the other hand, the surface is assumed to be flat, and all of the information is encoded in the intensity of the return.

At their most coarse, beamwidth-limited images make no assumptions about the surface being mapped, and a single point is defined in 3-D by the angular position of the centre of the beam and the range to the centre of the range gate containing the largest return as shown in the Figure 2.6.

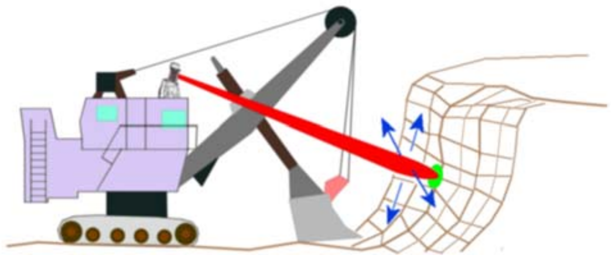


Figure 2.6: Illustration of beamwidth-limited surface quantization in which a limited number of uniformly spaced measurements is made at discrete points over the target

With a good-resolution radar system, this technique can produce acceptable images particularly if the space is over-sampled in angle. A typical image plotted as an occupancy space is shown in Figure 1.3.

A number of shortcomings are intrinsic to this measurement technique:

- A large RCS object will dominate the return if it is anywhere within the beam,
- Only a single point is created per measurement,
- The good range-resolution (compared to angular resolution) is not exploited by the process to improve the effective resolution.

A technique to overcome the last problem relies on the use of relative amplitudes of the returns in the surrounding range bins in conjunction with measurements made at angles adjacent to the angle of interest, to improve the effective radar resolution.

2.3.3. Image Interpretation

Accurate target identification using radar images is dependent on the resolution of the radar system in relation to the feature size that must be resolved and also on the reflectivity contrasts inherent in the target area. The relationship between the image resolution and the features that can be resolved is shown in Table 1.1.

Some of the characteristics that are common to most radar images made at low grazing-angles are listed below:

- Smooth horizontal surfaces (runways, lakes etc.) are specular and reflect most of the power away, so appear dark,
- Grass and cultivated fields scatter uniformly and vary due to differences in crop growth rates, local slope distributions etc., and so show texture,
- Scrub and trees are good scatterers so appear bright,
- Large trees, walls, banks etc. cast deep shadows,
- Hills can be identified by changes in reflectivity with grazing angle,
- Buildings, vehicles and other man-made objects are usually good reflectors.

For the bomb aiming capability as was required from H_2S radar discussed in Chapter 1, clusters of buildings are ideal as they produce large echo returns that are easily distinguished from those generated by natural vegetation. However, from a precision-guidance perspective, large buildings such as hangars, or closely spaced clusters of buildings make challenging targets, as it is often not possible to predict exactly which part of the distributed structure is the primary reflector. This problem is not as serious at long range where the structure is smaller than the radar resolution, or at very short range where individual reflectors within the structure can be resolved. Ideal candidates are groups of well-spaced point-targets or homogenous structures sufficiently large to cover many pixels, such as lakes or runways, that are embedded in a uniform “sea” of vegetation.

The resolution required to obtain multi-pixel features is, of course, determined by the feature size relative to the resolution. For long-range applications, natural features of interest are copses (clusters of tree growth), and wooded valleys, while manmade features comprise roadways and buildings. As discussed in Chapter 1, a resolution of at least 7×7m is required to achieve this capability. Whether this resolution is sufficiently good under all conditions is a matter for debate, though this thesis demonstrate that it is sufficiently good for accurate navigation over an airfield.

An examination of Figure 2.4 illustrates some of the problems encountered when trying to analyse radar images. The features that draw the attention first are three bright and dark bands towards the centre of the image. To the uninitiated, these would be identified as large sheds, however, they are in fact just avenues of closely planted trees (the bright bands) that cast deep shadows because of the low grazing-angle (dark bands). A second feature is the dark stripe that runs through the image from the top to the bottom. In this case it is dark not because of shadowing but because it is a sealed road with a low reflectivity. Finally, other bright spots such as the three brightest that lie on a diagonal across the image are high reflectivity manmade objects which are too small to cast shadows. In this case they are electricity pylons, but they could easily be water tanks or windmills.

Image interpretation may be facilitated if additional information is gathered. For example the use of polarisation can be used to differentiate between manmade and natural objects even if their reflectivities are similar. Doppler signatures are useful in determining whether a return is generated by a rock outcrop, which is static or a tree canopy which would move slightly.

The detailed interaction of a radar pulse with manmade objects and the ground as a function of the grazing angle, surface roughness and texture (surface covering) is discussed in Chapter 3 and Appendix A.

2.3.4. Alternative Representations

For navigation or guidance purposes, range-gate-limited images can be correlated with aerial photographs or specific unambiguous features can be extracted from the image and associated with similar features extracted from the aerial photograph. The feature association technique is faster than correlation because only a few features are required, but it requires some skill to examine an aerial photograph and predict those features that are likely to be visible on a radar image.

For scenes that contain height and reflectivity information, a representation called IHS (intensity-hue-saturation) where the height determines the colour (hue) and the brightness (intensity) is derived from the reflectivity can be used [6].

For short-range navigation and obstacle avoidance applications in buildings or underground in mines, various representations have been tried. The most simple is a volume cell with only two states (occupied or empty) to indicate the presence of a target [124]. Most systems now use statistical representations that allocate a target-presence probability to each cell. The evidence or occupancy grid method [138], [79] is one of the most widely used representations for several reasons: it is easy to implement, has fixed memory and computational requirements and it allows a statistical representation of the environment, albeit at spatially discrete intervals.

More recently, other probabilistic methods such as particle filters [126], [88] and sums of Gaussians [137] have been used that allow probabilistic mapping in a non-spatially discrete way.

One difficulty with occupancy grids is data association. Cross-correlation methods are computationally expensive if the search area is large and problems of ambiguity can occur in the association process (see Section 6.4). Their main problem is one of size. For large environments, a trade off between grid resolution and computational complexity must be made. For accurate navigation, the grid size should be a fine as possible, whereas from a computational perspective the grid size should be coarse [15].

An alternative to probabilistic mapping is the use of feature primitives. This is where a set of primitives (edge, corner, wall, building, etc.) are defined, extracted from the sensor data and located in Cartesian space. This method suffers from the fact that primitives can be difficult to extract, and may be ambiguous. Combinations of the various representations are possible. An example of using feature based association for navigation can be found in Section 6.3.

2.4. Radar Applications for Autonomous Guidance and Navigation

2.4.1. Indoor Applications

Many research institutes world-wide are investigating indoor automation using video imaging, sonar and laser scanners. A few have used millimetre-wave radar of which the most advanced is the Technische Universität München. In the early 1990s [124], [70] their research centred around a demonstrator vehicle that contained a radar providing 3-D real-time information up to a range of 50m. The vehicle also employed a short-range laser scanner, ultrasonic sensors and shaft encoders on the wheels.

The Pulsed-Doppler radar operating at 94GHz produced a 1.7ns pulse by gating a continuous-wave signal using a high-speed PIN switch. It operated with a 168mm diameter antenna to obtain range and angular resolution to measure indoor topology to a volume pixel (voxel) resolution of 0.2×0.2×0.2m. The beam could be directed through 360° in azimuth and ±20° in elevation using the classical mirror-scanner design. In addition, a combination of Doppler-based velocity measurement and split-gate techniques gave the sensor the unique ability for precise range tracking.

Pulsed technology was utilised in preference to FMCW because, at the time, processing speeds were not sufficiently fast to handle the requirements of a real-time FFT to perform the range-gating function. Operating at a pulse-repetition frequency of 1MHz, the received signal was sampled after a programmable delay and filtered, then digitised at 10kHz. This data-rate was too low for real-time vehicle guidance.

Problems that are common to all active systems included specular scattering from targets that resulted in large fluctuations in target return amplitude. Ghost targets generated by multiple bounces were dealt with by using a constant false alarm rate (CFAR) technique for the former and by using only short-range data for the latter.

The volume-cell representation performed adequately, as shown in Figure 2.7, given the specular nature of most of the targets and the fact that only a single observation was made of each voxel.

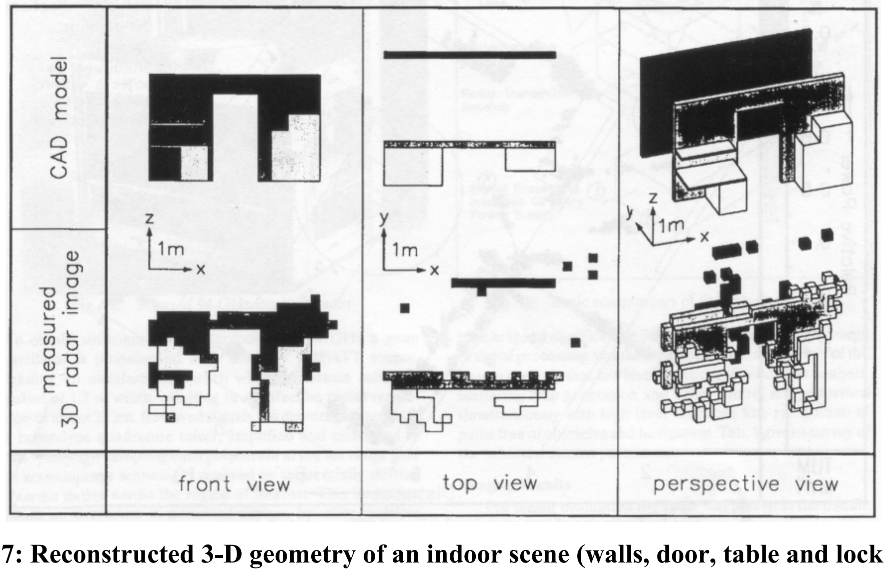


Figure 2.7: Reconstructed 3-D geometry of an indoor scene (walls, door, table and locker) based on 20×20×20cm volume cells

The capability of the radar as a collision avoidance sensor was investigated by non-coherently integrating 15 scans produced at 2m intervals to produce a single image, and then convolving this with the cylindrical outline of the vehicle to determine a boundary line within which the vehicle should remain as it moved.

Low-cost non-mechanical scanning technology is the holy grail for millimetre-wave radar systems. This technology is being investigated by the Radiophysics Department of the St. Petersburg State Technical University [8]. A 35GHz FMCW system has been demonstrated that uses a fixed horn to illuminate the scene and an electronically controlled linear ferrite integrated phased-array (IPA) receive antenna. The IPA can produce a scan angle of 50° (-5° to +45° from normal) without switching the feed port, and this can be extended to 90° if the feed port is switched.

The phenomenal increase in computer processing power in the last decade has allowed researchers in both indoor and field robotics to use FMCW radars [59], [85], [179], [8] for their systems. The reason for selecting this carrier modulation is because it offers some unique advantages over the other techniques as discussed in Chapter 4.

2.4.2. Ground-based Outdoor Applications

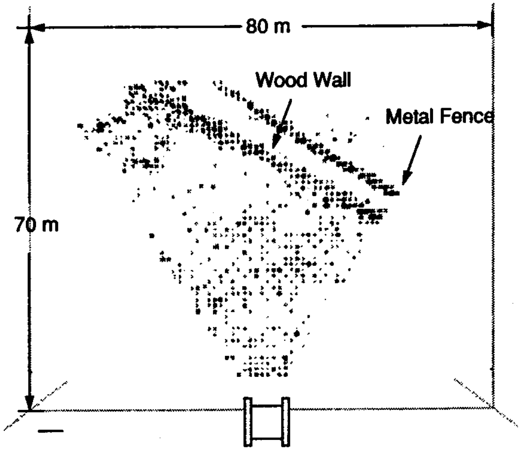
The three primary players investigating non-automotive radar-based outdoor navigation and guidance applications are Carnegie Mellon University (CMU), the Australian Centre for Field Robotics (ACFR) at the University of Sydney and to a lesser extent the Helsinki University of Technology (HUT).

At CMU [125], an unscanned FMCW radar operating at 77GHz was developed to investigate autonomous navigation on highways. This system has a range resolution of 0.5m produced by a swept bandwidth of 300MHz. It uses an array of four horns and a cylindrical lens to produce an antenna pattern with an elevation beamwidth of 3° and an azimuth beamwidth of 12° made up of four overlapping beams. The bearing to a point target is determined by using the echo phase difference in adjacent channels. This has the advantage of being non-scanning, but because the spacing of the horns exceeds $\lambda/2$, grating lobes are generated in azimuth and ghost targets appear.

3-D imaging developments have included a mirror-scanned pulsed system [22], followed by more recent work [86] the original aim of which was to produce a non-mechanical imaging radar for field robotic applications. Techniques to produce low-cost phased array-type scanners and to synthesise narrow beams using relative motion between the radar and the target are being investigated. The program has not been particularly successful and conventional scanning radars are still being used [85] to produce images of the quality of those shown in Figure 2.8.

One of the sensors used for research work is a dual antenna 77GHz FMCW unit with a 2° elevation and a 1° azimuth beamwidth, that scans over 64° in azimuth and over 8° in elevation by selecting from a sequence of four stacked beams. Operational range is limited to between 1 and 64m. The theoretical range resolution is 0.5m obtained by a swept bandwidth of 300MHz over a period of 256μs. Linearization is accomplished by lookup table that provides adequate performance but it cannot achieve a linearity to match that produced using closed-loop techniques (see Appendix E).

(a) Bird's Eye View



(b) Oblique View

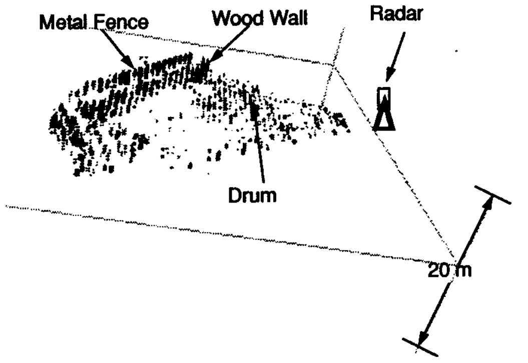


Figure 2.8: Two perspectives of occupancy grid based 3-D representation of outdoor data made at CMU. In these images the radius of the opaque spheres is proportional to the log of the reflectivity value at that point

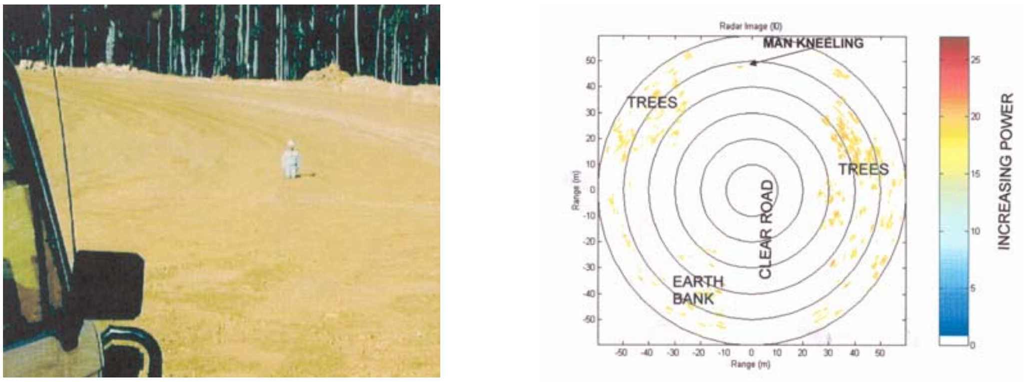
A similar radar operating at 93GHz [179], was developed by HUT. It used a pair of 150mm diameter horn-lens antennas to produce a symmetrical beamwidth of 1.5°. Because mirror scanning with the dual-beam configuration can be mechanically difficult, azimuth scanning was achieved by rotating the whole radar. A range resolution of 0.2m for an operational range of 50m was achieved using wide-band FMCW techniques.

The first radar that was developed at the ACFR operated at 77GHz using a closed-loop-linearised FM sweep of 600MHz giving a theoretical range resolution of 0.25m [78]. A 2-D mirror scanner with an antenna aperture of 150mm produced a beamwidth of 1.8° that could be scanned over 360° in azimuth. To allow for beacon identification in highly cluttered environments, the unit had a dual-polarisation receiver that could discriminate between specular returns from unwanted clutter (typically containers in this application) and those from strategically placed polarisation-rotating beacon reflectors.

A curved mirror was used to spread the elevation beam to compensate for irregularities in the road surface that caused the platform to pitch or roll.

For vehicle navigation [60], an extended Kalman filter running at 20Hz was implemented using suitably extrapolated observations made of the radar beacons within the range of the radar to constrain the vehicle position uncertainty.

This radar technology has been extended at the ACFR to produce an extremely sensitive collision-warning device that is capable of detecting a person kneeling in the path of the vehicle at a range of 50m as shown in Figure 2.9.



(a)

(b)

Figure 2.9: Collision avoidance radar application for human detection at the ACFR comparing (a) a photographic image and (b) a radar image with the return from a person indicated

These radar techniques and applications have been extended in the past five years to include a miniature 77GHz mirror-scan radar on gimbals mounted in the nose of a UAV (see Section 7.2.3) and a 94GHz unit (under construction) that uses a robust spinning-grating antenna for an AGV. Technical details of these and other scanning radars are referred to from time to time in the remainder of this thesis.

2.4.3. Airborne Applications

Though in the strictest sense, "synthetic vision" is not a navigation application, a brief summary of developments in this field is included here. One of the early programmes by Lear Astronics [46], [47] was successful in developing long-range 94GHz solid-state FMCW radar with a maximum operational range of 6000m in ground tests. This bistatic configuration was not particularly effective during flight trials when the radome-induced leakage (discussed in detail later in this thesis) reduced the sensitivity of the system considerably.

More common were systems that use high-power pulsed sources. A demonstrator built by Norden Systems [118] produces 1.5kW pulses at 95GHz from an Extended Interaction Oscillator (EIO). Tower tests in 1992 at the Wright Patterson Air Force Base confirmed good radar performance under various adverse weather conditions.

Honeywell has successfully demonstrated the capabilities of a 35GHz pulsed-magnetron based imaging radar driving a head-up display (HUD). This radar produces camera-like images at over 10 frames per second [48], [83], [167].

These programmes were evaluated by the Georgia Institute of Technology [102] as part of the Synthetic Vision Technology Demonstration (SVTD), the main goals of which were to document and demonstrate aircraft sensor and system performance using millimetre-wave and IR sensors. The primary conclusions drawn by the evaluation determined that the contrast achieved for the 35GHz radars was superior to that of the 94GHz contenders, but that the resolution was hardly adequate from the 750mm aperture specified for the systems.

Another 35GHz system that should be mentioned is the Longbow fire-control radar for the AH64/D Apache helicopter. However, as its specifications are not available in the open literature, very little can be said about its performance compared to the other systems discussed in this section.

MBDA in conjunction with BAE Systems developed what is known as the Automatic Landing Guidance (ALG) system which has been widely trialled in the USA [94]. The radar is an extension of the 94GHz FMCW Brimstone seeker development, and has proven itself on numerous flight trials to the extent that the Safety Review Board has approved its use for landing in 0,0 visibility conditions.

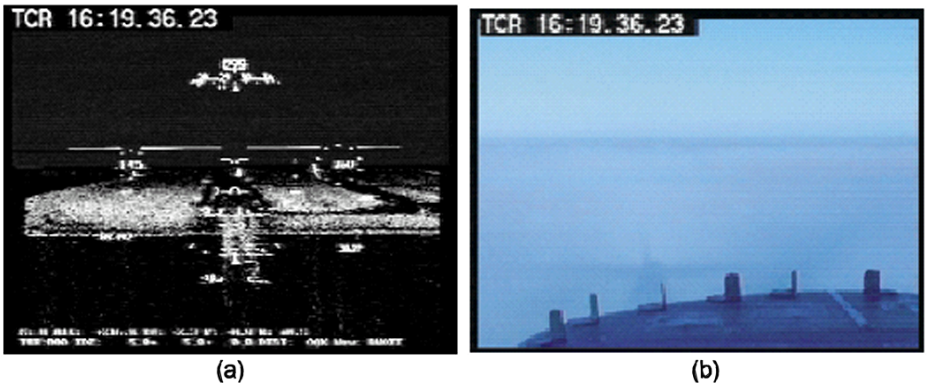


Figure 2.10: Synthetic Vision using ALG showing (a) a radar Image of a runway and (b) a visible cockpit-camera image of the same view

A number of more recent German programmes have been reported. DLR's Institute of Flight Guidance has developed a pilot assistance system of which the radar-based synthetic vision is one part [119]. Their technique involves using *a priori* knowledge of the expected airport position as well as the negative-contrast image of the runway and positions of landing lights to generate multiple hypotheses from which the correct runway position can be extracted.

Finally, DLR's Institute of High Frequency Techniques and Radar Systems has developed the Sector Imaging Radar for Enhanced Vision which uses a real aperture high-resolution millimetre-wave system in conjunction with some extremely sophisticated signal processing to produce SAR quality images [120].

2.5. Conclusions

This chapter has shown that radar images can be generated by scanning a narrow antenna beam across the surface of the earth and gating the returns in range. The images are then formed either by plotting the return intensity as a function of angle and range, or in 3-D by plotting an occupancy volume (the presence of a target in a specific cell). Combining the two techniques, and even displaying additional Doppler or polarisation information is possible, and can lead to improved situational awareness.

An analysis of feature size requirements included in the chapter has determined that a nominal 7 metre range and cross-range resolution would be the minimum required to identify both natural and manmade features for long-range operation.

A summary of the work that has been undertaken to date shows that all of the sensors developed, with the exception of those developed for “synthetic vision” have been used either for extremely short-range indoor applications, or moderately short-range outdoor applications. The problem of obtaining the “bigger picture” by using a long-range radar system for autonomous navigation or guidance has yet to be addressed.

Chapter 3.

Target and Propagation Effects

3.1. Introduction

This chapter begins by examining the characteristics of the targets that would be used by a radar mounted on an autonomous vehicle. These targets are in fact the ground returns which are more commonly considered to be clutter. Measurements made to determine the spatial statistics of various clutter types as well as the temporal statistics of typical millimetre-wave clutter are included as they are critical in determining the performance requirements of a radar system. The section also includes the effects of target motion on the measurement statistics, and hence on the apparent image quality.

The second half of the chapter considers some of the propagation effects that alter the ability of a radar to perform its primary imaging function. This includes a brief overview of the radar range-equation leading into a discussion of the attenuation and backscatter effects of clear air and various hydro-meteors on the propagation of millimetre-waves. A final discussion on multipath effects and electromagnetic interference concludes the chapter.

For more information on these and other target and propagation effects, Appendix A and B should be read.

3.2. Target Reflectivity Measurements

Because of the lack of reliable published information on millimetre-wave clutter, it was decided that a series of measurements should be undertaken.

A conventional closed-loop FMCW radar manufactured by Philips Microwave and fitted with a 250mm diameter Cassegrain antenna manufactured by Hughes Microwave was used in conjunction with a Hewlett Packard HP8562A spectrum analyser. The latter was used to perform the range gating function, and was configured as follows:

• Frequency Span	1kHz – 500kHz,
• Bandwidth	3kHz.

The 3dB (-6dB two way) beamwidth of the antenna and the spectrum analyser bandwidth determine the extent of the gate in cross-range and range. At 300m this equates to an area of 5×3m when the correction factor of 1.33, which is used to correct for illumination by a Gaussian beam, is applied [67] pp192-195. More information on the relationship between the received spectrum, and the range resolution can be found in Chapter 4 and Appendices D, E and F.

A set of precision triangular trihedral corner reflectors spanning the RCS range from 1 to 1000m^2 at 94GHz is used for calibration. A telescope attached to an alignment jig that fits over the face of the reflector is used to align it to within a fraction of its 12.5° beamwidth (1dB) [66] p40.

3.2.1. The Noise Floor

Much of the noise within a radar system is generated by the thermal motion of the conduction electrons in the ohmic portions of the radar-receiver input stages. This is known as thermal or Johnson noise, P_N (W), and can be expressed in terms of the system temperature, T_o (Kelvin), of a matched resistor at the input of the receiver, and the bandwidth, β (Hz),

$$P_N = kT_o\beta, \tag{3.1}$$

where k, Boltzmann's constant (1.38×10^{-23} J/K), is the constant of proportionality.

To account for the fact that the noise power in practical receivers is always greater than that of thermal noise alone, the total noise at the output of the receiver, N (W), can be considered to be equal to the noise power output from an ideal receiver multiplied by a factor called the Noise Figure, NF,

$$N = P_N NF = kT_o\beta.NF\,. \tag{3.2}$$

Figure 3.1 and Figure 3.2 show the radar noise floor as a pseudo image with range (frequency) and time and also as a set of line graphs showing the temporally processed statistics.

The graphs exhibit the standard $1/f$ noise characteristic of an FMCW radar. The noise variation extends to +/-4dB around the mean with a standard deviation of between 1.5 and 2dB, and a mean level of –79dBm in the region between 250 and 350kHz. Of particular note is the decrease in noise floor as the frequency increases, and the spurs generated by the linearization process at 180 and 370kHz.

It is very seldom that the signal level exceeds -30dBm even for large targets, so the effective dynamic-range is about 50dB for these measurements. However, due to the phase-noise effects surrounding each return, this does not imply a contrast of anything approaching this magnitude (see Appendix F).

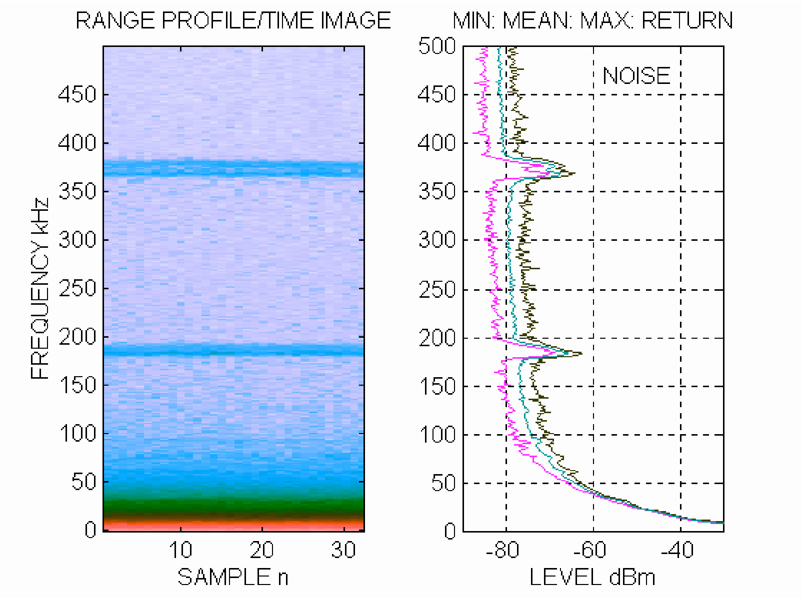


Figure 3.1: FMCW radar system noise-floor characteristics showing the minimum, mean and maximum values measured over a 500kHz frequency span

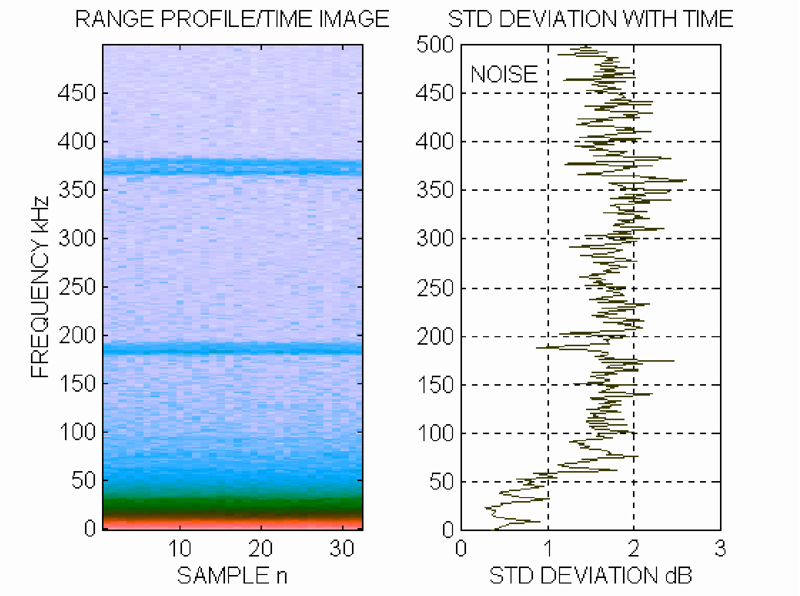


Figure 3.2: FMCW radar system noise-floor characteristics showing the standard deviation of the noise over a 500kHz frequency span

3.2.2. Spatial Variations

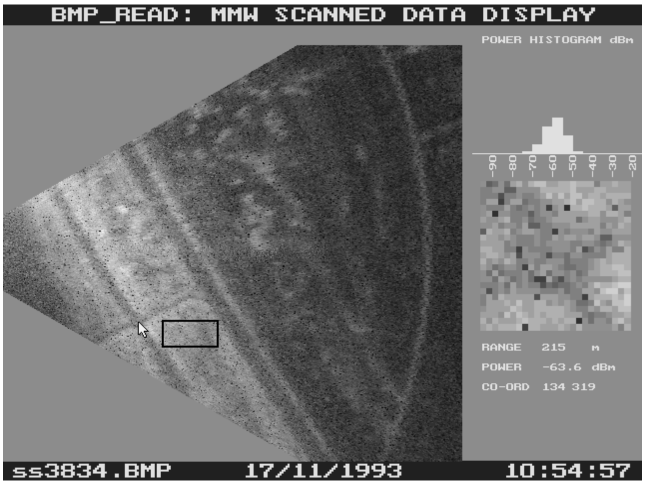
From the perspective of robotic applications, spatial variations in ground and foliage reflectivity are important during the search for specific features. Consider the example of an AGV trying to navigate across an undulating field of grass and rocks cut by the occasional road. Spatial variations in the reflectivity of the terrain could be used to determine a safe path between the rocks if there is a reliable method to discriminate between the two terrain types, but if there is a possibility that a high reflectivity grass tussock could look like a rock, and visa versa, then it is important that the signal processing and path-planning algorithms are aware of this (an example is shown in Section 6.6).

At microwave frequencies, the spatial distributions of reflectivity within uniform areas are Rayleigh distributed except for elevation angles of less than 5° [24]. Where the statistics of the radar return from a number of terrain types are considered, then the distribution appears more nearly log-normal and the standard deviation is greater.

To confirm this hypothesis, a custom suite of programs was written to examine the images in minute detail with the specific goal of determining the differences in reflectivity of various forms of clutter at low grazing-angles.

The following measurements were made from a radar system mounted at a height of 3m operating across relatively flat ground near an airfield. This equates to a grazing angle of about 1.1° at a range of 150m. Figure 3.3 shows a section of runway and taxiway being analysed.

A crude representation of the raw statistics of the highlighted rectangular section of the image are shown in the histogram at the top right of the figure. It can be seen that the return levels from this particular piece of clutter are roughly Gaussian with a mean of -60dBm. The cursor arrow marks a single image pixel on the tarmac intersection of two taxiways. The small image on the right shows a magnified view of this area, and the details below it are for the individual pixel being measured.


BMP_READ: MMW SCANNED DATA DISPLAY

RANGE 215 m
POWER -63.6 dBm
CO-ORD 134 319

ss3834.BMP 17/11/1993 10:54:57

Figure 3.3: Image reflectivity analysis tool output showing the captured file ss3834.bmp which consists of a section of runway with two intersecting taxi ways and the surrounding scrub. The circular arc is an artefact of the radar linearisation process used to make the measurement

Using another program, the outputs of which are shown in Figure 3.4, selected areas of the image can be highlighted and then stored (along with their range) for further processing in MATLAB. This allows for a more comprehensive contrast analysis between the reflectivities of different target types to be made.

The received power is a function of the target reflectivity, the antenna gain, the beamwidth and the receiver bandwidth (gate size). To calibrate the system, a reference target reflector with an RCS of 10m^2 was placed at a range of 140m, aligned perfectly using a telescopic sight, and the magnitude of the return logged. Because the terrain between the radar and the reflector was rough grass and scrub, it was assumed that multipath interference would be insignificant and so no additional precautions were taken. From the received-power level of -26dBm it is simple to produce the appropriate formula to convert from received power to reflectivity.

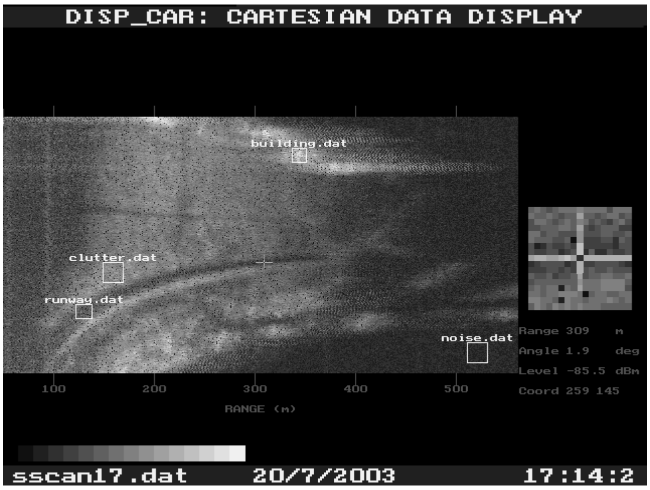


Figure 3.4: Image reflectivity analysis tool output showing the captured file, sscan17.dat with various sections marked for storage and further processing

For a range-gate size of 3m, corresponding to the 3kHz RF bandwidth of the spectrum analyser, and a 3dB beamwidth, as stated previously, of 0.85°, the conversion is

$$10\log_{10}\sigma^{o} = 10\log_{10}P_r - 36 + 30\log_{10}R\,, \qquad (3.3)$$

where σ^{o} (m^2/m^2) is the reflectivity, P_r (dBm) is the received power and R (m) is the range.

The reflectivity data PDFs for these boxed areas are shown corrected for range in Figure 3.5.

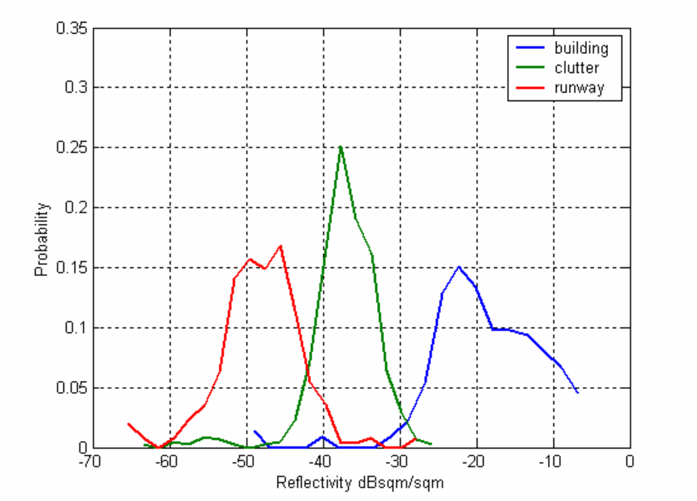


Figure 3.5: Range corrected histograms of reflectivity distribution for the different target areas show the typical log-normal distributions obtained for millimetre-wave clutter

In this example the median value for the runway reflectivity is -47dB with the clutter some 10dB higher at -37dB and the building higher still at -22dB. The standard deviations of the runway and clutter reflectivity are similar at about 5dB with the building a bit larger at 7dB.

These measurements should be considered to be approximations as there are a number of potential sources of error in the measurement and in the subsequent determination of the reflectivity:

- Error in the true RCS of the reference reflector,
- Uncompensated multipath effects during calibration with the reference reflector,
- Errors in the conversion from spectrum analyser beamwidth to range gate span due to uncertainties in both the radar sweep and the filter shape,
- Uncertainty in the true beam shape of the antenna used to estimate the footprint.

The spatial distributions of the target reflectivity for both the runway (tarmac) and the clutter (rough grass and low scrub) are log-normal in shape, an observation which has been confirmed by other researchers [86]. The return from the building which is characterised as a reflectivity value for purposes of comparison even though it is more likely to be the return from a few high RCS points, has a less well defined distribution which appears to be made up from two distinct reflectors with different cross sections.

The statistical data from [68] shown in Figure 3.6 for grass and crops confirm the variation in reflectivity measured above, but models the mean reflectivity to be more than 10dB too high. This is probably due to the fact that the data for low grazing-angles are extrapolated from measurements made above 10°, whereas the data presented in the previous section are measured at a grazing angle of about 1°.

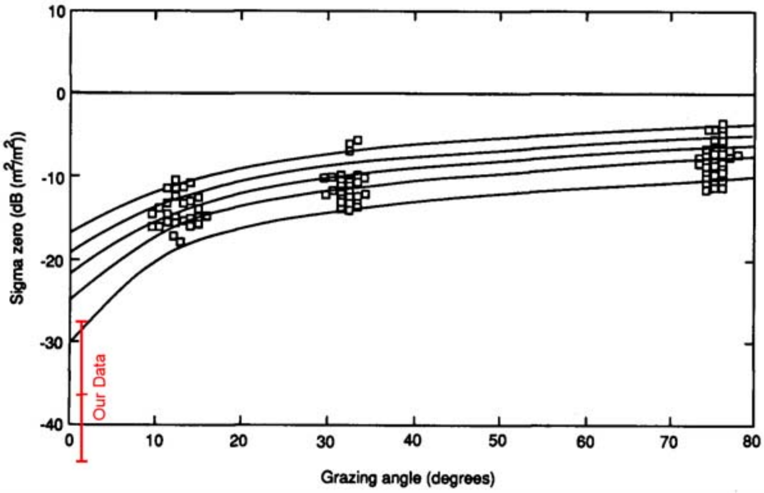


Figure 3.6: Comparison between the measured reflectivity of grass and crops at 95GHz with the Georgia Tech statistical model predictions for 10%, 30%, 50%, 70% and 90% distribution levels

The noise floor is not a function of the range and so it cannot be plotted on the reflectivity data axes. As it defines the minimum level below which measurements cannot be made its PDF is plotted along with that of the runway. It can be seen from Figure 3.7 that the returns from the runway are still more than 10dB above the noise floor and so the reflectivity measured for this and all of the other target types is valid.

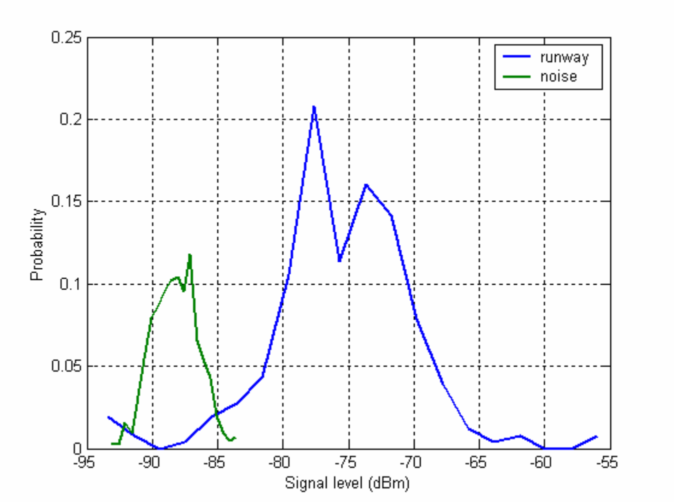


Figure 3.7: Histograms of runway signal level compared with the noise floor confirm that the runway reflectivity statistics are not noise limited in Figure 3.4

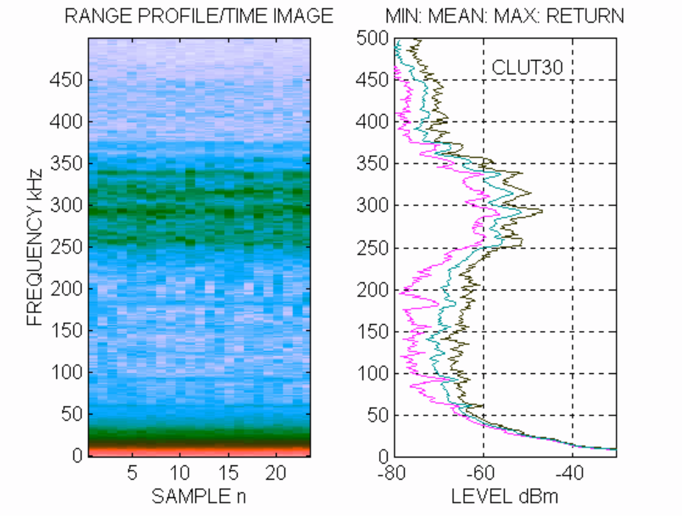
The narrow elevation beamwidth and hence the small footprint for anything but extremely low grazing-angles required some attempt to be made to integrate the outputs of multiple images made while the radar was in motion. This was not very successful because, though the radar displacement could be measured in the direction of travel using a calibrated odometer, the orientation and lateral displacement were not measured for the ground based measurements. Airborne measurements discussed in Chapters 5 and 6 used a highly accurate GPS/INS to determine the position and pose of the platform during the imaging process to allow the successful integration of multiple images.

3.2.3. Temporal Variations

Temporal fluctuations are those variations that occur within a single resolution-cell with time. They are described by amplitude statistics and frequency spectra or decorrelation times. Measurements made on the same areas at 10, 16, 35 and 95GHz show that both the standard deviation and the shape of the distribution changes with increasing frequency. Whereas at 10GHz the distribution is generally Rayleigh in shape, up at 95GHz it more closely approximates log-normal with its associated longer tails [67] p371. It should be remembered that these statistics are approximations only and in reality the data seldom fits a specific distribution exactly. Variations will be a function of a number of parameters including processing bandwidth [67] pp 221-222, radar resolution, scene content and geometry etc.

Temporal analysis of a typical area that might be traversed by an autonomous vehicle was undertaken. The clutter area chosen was predominantly grassland with the occasional bush. The pattern of the grass growth was irregular with some clumps and strips reaching a height of 500mm and other areas much shorter (100mm).

Though there was no apparent correlation between the visual appearance of the clutter and the variation in the observed reflectivity, after a brief search, a section of clutter was found with a reasonably small spatial variation over the range between 250 and 350m. This was logged from a stationary radar overlooking the scene and it is displayed in Figure 3.8.

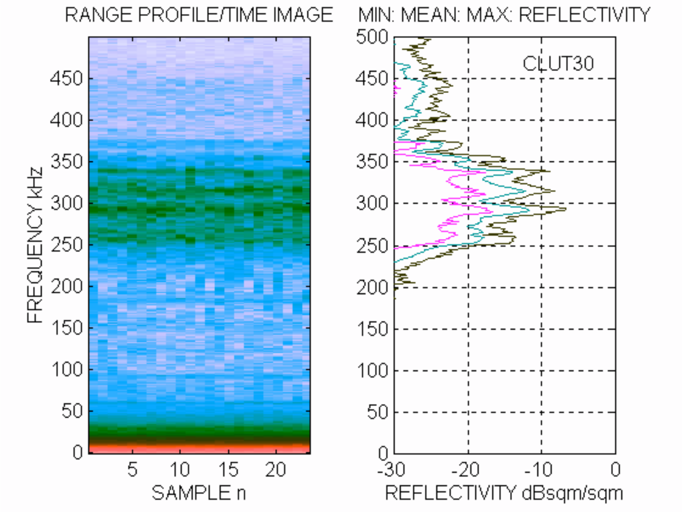


(a) (b)

Figure 3.8: Measured clutter-return power from grassland and scrub showing (a) the time-frequency pseudo-image and (b) the variation in signal level measured during a 60s period

The pseudo image in the figure is a representation of the same clutter slice measured at intervals of just over 2s for 1 minute. These data, converted to a received power in mW, is averaged to produce the mean value at each range which is displayed in dBm. Minimum and maximum values are also determined and displayed in the adjacent graph.

In the region of interest, a spatial variation in reflectivity of about 10dB occurs. This is typical for clutter of this nature as discussed in Section 3.2.2.



(a) (b)

Figure 3.9: Measured clutter reflectivity from grassland and scrub showing (a) the time-range pseudo-image and (b) the variation in reflectivity levels

The temporal variation is measured to have a standard deviation of between 2 and 3dB with a peak to peak variation on either side of the mean of between -6dB and +4dB. Because the variations in the clutter level are very similar to those measured for the noise and also for the signal as a whole, most of the variation is probably generated by the measurement process. Furthermore, because the data were only sampled every 2.5s, the temporal clutter-correlation period cannot be determined from these measurements.

Though clutter spectral-analysis has not yet been used during the feature characterisation process, it will become an important technique to help discriminate between rocks and bushes. At low frequencies Doppler-shift spectra due to clutter movement [68] p152 can be represented by a Gaussian function of the form

$$W(f) = W_o \exp\left(-af^2 / f_t^2\right), \tag{3.4}$$

where: W – Power density (W/Hz),
W_o – Power density at zero velocity (W/Hz),
f_t – Cutoff frequency (Hz),
a – Experimentally determined coefficient.

For many clutter types, particularly trees, a physical mechanism exists that generates two different spectral signatures (a low frequency component due to limb movement and a high frequency component due to leaf flutter) as shown in Figure 3.10.

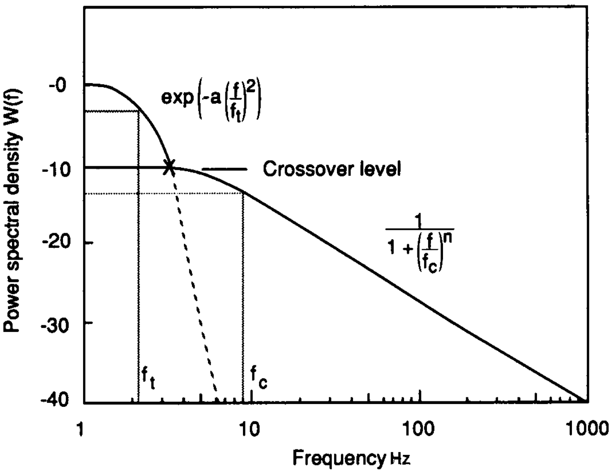


Figure 3.10: Low and high frequency components of the power spectral density returns for wind-blown trees showing differences in the spectra that can be attributed to leaves and branches

The low frequency distribution is modelled by the Gaussian function described in (3.4), while the high frequency component is modelled using the following Lorentzian function

$$W(f) = \frac{A}{1 + \left(f / f_c\right)^n}, \tag{3.5}$$

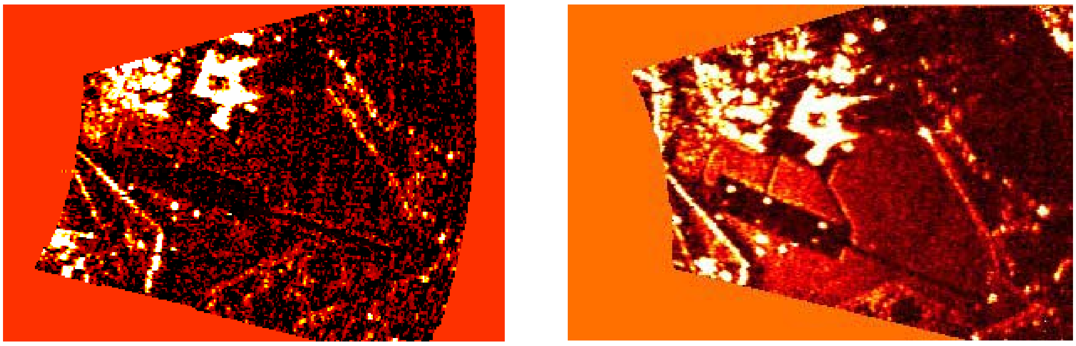
where A – Empirically derived constant,
f_c – Cutoff frequency (Hz),
n – Empirically derived factor (dependent on the frequency).

At 94GHz typical values for f_t = 3.5Hz, f_c = 35Hz, n = 2 and the breakpoint is 18dB for high wind speeds. At low wind speeds the low-frequency component remains unchanged but the high-frequency cutoff, f_c reduces to 6Hz.

3.2.4. Effects of Radar Movement

If the radar system is moving during the observation period, then sequential views of the same target area will occur from different view points. It has been shown that averaging these images results in a marked reduction in the standard deviation of the reflectivity of the observed terrain and an improved clarity of image.

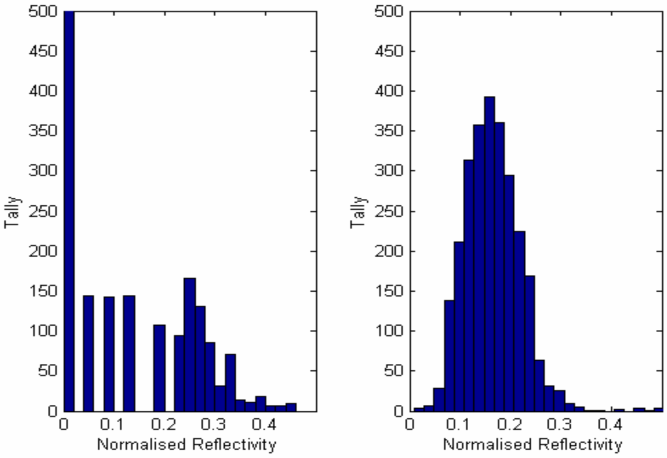
Reflections from the individual scatterers in a range cell can combine constructively or destructively to produce measured reflectivities with large standard-deviations. These variations result in an image that is mottled and difficult to interpret by eye. This is known as speckle and can be minimised by either wide-bandwidth processing, or by integrating the image from slightly different perspectives. The latter effect is clearly shown in Figure 3.11 where the image on the left is unintegrated and the image on the right is produced by the integration of a sequence of four images each displaced in range by about 200m.



(a) (b)

Figure 3.11: The effect of moving the observation point on image clarity is illustrated by comparing the image (a) produced by a single scan and (b) four integrated scans

The reflectivity statistics for the same homogenous region of clutter in the two cases is shown in the received power histograms in Figure 3.12.



(a) (b)

Figure 3.12: Histograms of the grass verge surrounding the runway for (a) a single image and (b) an integrated sequence of images obtained from slightly different perspectives

For the unintegrated image, the normalised reflectivity distribution is largely bimodal with a large number (1500) of the pixels having very low reflectivity, a few sparsely distributed through the intermediate region and then another concentration around 0.25. For this distribution, the median is 0 and the standard deviation is 0.13. For the integrated image, the distribution is almost Gaussian with a median value of 0.16 and a standard deviation of 0.06. Both of the histograms have been truncated to eliminate the sparse tail of high-reflectivity returns that extend up to a normalised reflectivity of one and also to eliminate tallies above 500.

Though it is possible that some of the improvement in the image quality is due to the higher SNR as a result of the integration, given the fact that only four samples were integrated (<6dB of non-coherent gain), it is believed that most of the improvement is due to a reduction in the speckle.

3.3. Target Feature Representations

Section 2.3.4 has considered a number of different ways of representing the presence of a target echo. However, in this case, as two different measurement regimes are used to generate 2-D and 3-D images for navigation purposes, it may be justifiable to produce two different target feature representations.

3.3.1. 2-D Representations

Two different models to determine the probability of occupancy from signal and noise distributions are considered [86]. In the first a uniform distribution for the echo amplitude PDF is used and in the second a lognormal distribution is assumed. The latter appears to produce a much more abrupt transition in the probability of occupancy than the former. The model also makes assumptions about the occupancy of cells in front of and behind an occupied cell. These assumptions rely on the fact that the beam dimensions are often larger than those of an illuminated object and so may pass around, or even through it and illuminate objects further away.

Representation in the 2-D case starts with a basic reflectivity map in which each cell on a rectangular grid is given a mean reflectivity value determined by the received signal level in that cell. If a number of samples have been integrated to obtain that value, then additional information on the statistics of that pixel (such as the variance) can also be stored. For more complex radars, additional information will be available. For example, if the full scattering matrix is measured then the four mean reflectivity values and their associated statistics can be stored.

The next phase of the process would be to construct an occupancy grid based on the P_d and P_{fa} which would serve different functions depending on the application. In the path-planning case, the reflectivity information could be processed to define go and no-go areas for an AGV which would then be used to select a safe path. This would be a dynamic map which would change as the vehicle traversed the terrain and obtained additional information about the surface conditions from the sensors.

The use of additional statistical information, discussed above, based on temporal fluctuations of the reflected signal level should be used to add to the information about the surface, as should the polarisation information. However, because of the speckle inherent in any of these observations, the occupancy grid would be patchy unless some spatial integration was introduced.

Finally, one of the most useful characteristics of radar images and their translation into a useful representation that has not, thus far, been considered are the spatial variations in reflectivity that impart texture to the surface. The use of such an approach for path planning is discussed in Section 6.6.

3.3.2. 3-D Representations

A basic 3-D occupancy grid is developed [124] by sampling the space using a polar scanner, transforming to a Cartesian coordinate system aligned parallel to the physical scene geometry and then identifying occupied cells using a volume CFAR process with a threshold set 7dB above the mean noise level. These results are shown in Figure 2.7. In this model, no assumptions are made about the occupancy of the cells either in front of or behind those which contain targets.

The 2-D process outlined in Section 3.3.1 is extended to 3-D in which it is shown that though the evidence of occupancy and emptiness are low after a single observation, with the accumulation of observations, a dramatic increase in the levels of evidence occurs along with an improvement in the surface smoothness. Simulated models of road surfaces that contain interesting features like ditches, bumps and posts show a similar pattern. In this case the physical surface is defined as an iso-surface with a 90% probability of occupancy.

Differences between the methods used to specify the confidence levels set in these two examples can be reconciled by considering the probability of occupancy to be equivalent to the P_d for a specific SNR and a specific P_{fa} (established by the CFAR process).

3.4. Propagation

This section examines and quantifies the propagation effects that may alter the performance of a radar system. As shown schematically in Figure 3.13, these include atmospheric effects, multipath propagation, natural and manmade interference and interactions around the target area.

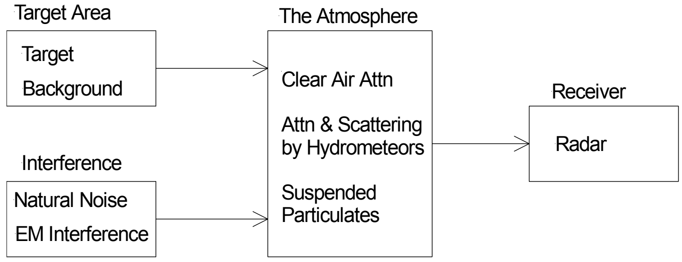


Figure 3.13: Block diagram of the radar environment

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 Figure 3.14.

In the context of this thesis, the surface clutter and manmade objects constitute the targets of interest with the volume clutter producing returns that interfere with the imaging process.

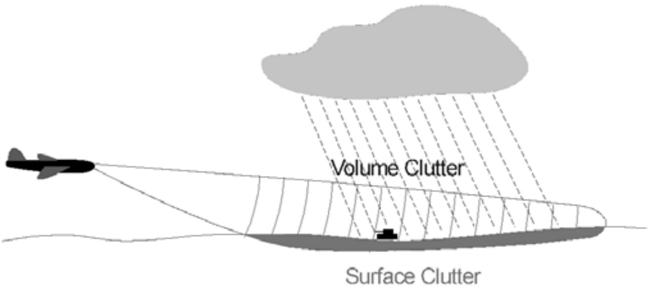


Figure 3.14: Schematic diagram illustrating the surface and volume clutter returns that compete with the target echo within each range gate

3.4.1. 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 power, P_r, and the transmitted power, P_t (W), can be written

$$P_r = P_t \exp(-2\alpha R). \tag{3.6}$$

Because the radar range equation is often evaluated in dB, the change of base theorem can be used to express the attenuation as $\alpha_{dB} = 4342.9\alpha$ in dB/km (one way), making the relationship

$$10\log_{10} P_r = 10\log_{10} P_t - 2\alpha_{dB} R_{km}, \tag{3.7}$$

where R_{km} is the range (km).

This non-linear relationship with range makes the equation difficult to solve in a closed form, and numerical or graphical methods are normally used. For example, if the radar range equation for a monostatic system with an antenna gain, G, operating at a wavelength, λ (m), with losses, L, is written in dB to include the attenuation term

$$\begin{aligned} 10\log_{10} P_r = {} & 10\log_{10} P_t + 10\log_{10}\left(\frac{\lambda^2}{(4\pi)^3}\right) + 20\log_{10} G + 10\log_{10}\sigma \\ & -10\log_{10} L - 40\log_{10} R - 2\alpha_{dB} R_{km} \end{aligned} \tag{3.8}$$

then the received echo power from a point target with an RCS of σ (m^2) can easily be plotted along with the minimum detectable signal to determine the maximum range as shown in Figure 3.15.

The intersection of the dashed line which defines the minimum detectable signal for a specific P_d and P_{fa}, and the received power curves determines the operational range of the radar system. This is the model of a typical FMCW automotive radar, and it can be seen that depending on the weather conditions, the detection range for a $1m^2$ target drops from a maximum of 500m in clear air down to 300m in heavy rain.

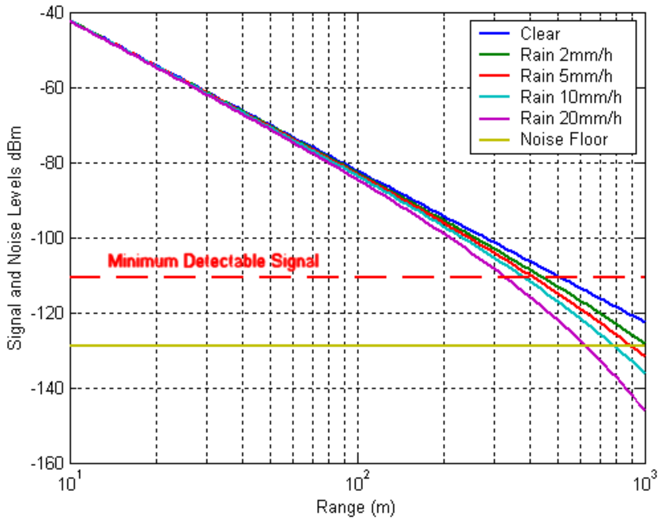


Figure 3.15: Graphical solution to the radar range equation showing the effects of atmospheric attenuation on the detection range for a point target

In an imaging application, the ground RCS is a function of the footprint which (because of beam divergence) is proportional to range. This results in a range equation where the detection range is a function of R^3 for range-gate-limited imaging and R^2 for beamwidth-limited imaging.

For an appropriately controlled elevation gain, as discussed in Section 2.2.1, a constant signal-level is maintained independent of the range for returns from the earth's surface.

3.4.2. Atmospheric Attenuation

A summary of the attenuation that can be expected at frequencies extending from 10GHz through to 1000THz in clear air as well as through fog and rain is shown in Figure 3.16.

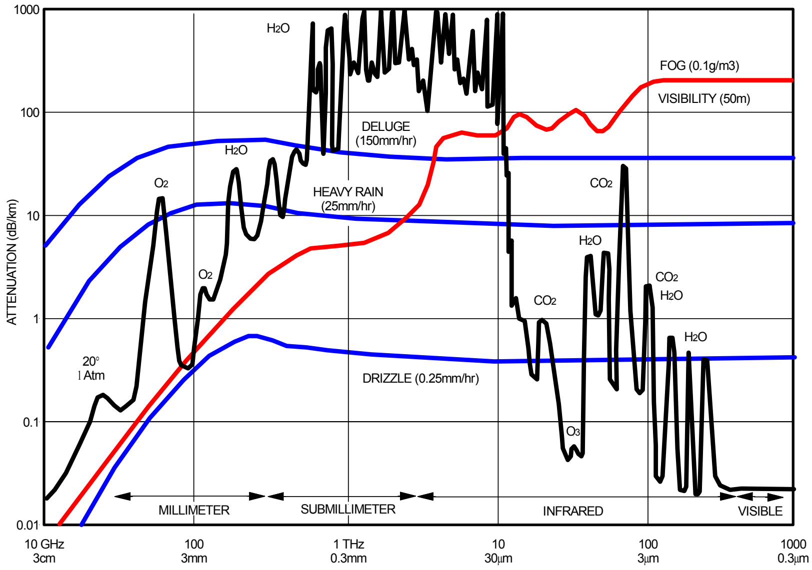


Figure 3.16: Summary of atmospheric attenuation effects for clear air, fog and rain at wavelengths ranging from 3cm to 0.3μm

It can be seen that the atmosphere becomes progressively more opaque as the frequency increases. This trend is punctuated by peaks in the absorption caused by interactions of the electromagnetic waves with molecules of oxygen and water. Bands between these resonant peaks, known as atmospheric windows, occur at 35, 94, 140, 220 and 360GHz [157], [20].

Between 1THz and 10THz the atmosphere is almost completely opaque as far as a radar system is concerned, with typical attenuations of between 100 and 1000dB/km. At even higher frequencies, 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. For this reason if the available volume is restricted, most high-resolution, long-range sensors operate either in the millimetre-wave or in the infrared/visible region of the spectrum.

The total attenuation at any frequency is determined by summing the different components. For example, if it were both misty and drizzling, then the total attenuation at 94GHz would be 0.2 + 0.2 + 0.2 = 0.6dB/km. More detail concerning the background to this figure including the effects of rain and fog can be found in Appendix A.

3.4.3. Attenuation due to Dust and Smoke

The absorption of EM waves by dust and smoke is directly proportional to mass loading as determined using Rayleigh or Mie scattering theory. In the example shown in Figure 3.17, the magnitude of dust attenuation for a loading of 2.6g/m^3 at millimetre-wavelengths as a function of particle size and frequency is calculated [61].

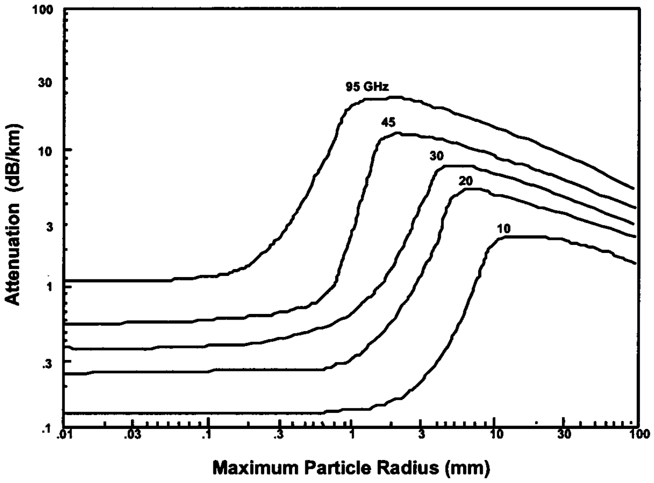


Figure 3.17: Signal attenuation (one way) at different frequencies due to dust, as a function of maximum particle radius for a dust concentration of 2.6g/m^3

As linear relationship between dust concentration and the attenuation exists, if the dust concentration is increased to 10g/m^3, then the attenuation in dB/km would increase by a factor of 10/2.6. Further work in regard to attenuation through dust-storms has been undertaken [93], [90], [173] in which dust particle size distributions are used to quantify the attenuation more accurately.

The attenuation, α_{dust} (dB/km), of radar signals propagating through a dust laden atmosphere using the Rayleigh approximation is given by

$$\alpha_{dust} = \frac{2.317\times 10^{-3}.\varepsilon^{*}_{dust}}{\left[\left(\varepsilon_{dust}+2\right)^{2}+\varepsilon^{*\,2}_{dust}\right]\lambda}\cdot\frac{1}{K^{1.07}}, \tag{3.9}$$

where ε_{dust} is the real component of the permittivity of the dust particle, ε^{*}_{dust} is the imaginary component and K is the visibility (km).

The measured permittivity of dry dust at 10GHz varies slightly from 5.23 - j0.26 [93] to 5.52 - j0.16 [173] depending on how it is measured. For crushed limestone the value is 7.41 - j0.063 [148]. Because of its hygroscopic nature dust is seldom completely dry, and so the water contribution, as a proportion of the total mass, must be considered when the attenuation is calculated. The permittivity of water is 8.35 - j15.45 at 89GHz.

There is some discrepancy between the data presented by [61] and the attenuation calculated using (3.9) which shows a much lower attenuation. Depending on the reference, the total attenuation due to dust in the most severe sand-storm[2] will be between 1.2 and 4dB/km. For most autonomous applications this attenuation is insignificant and can be ignored.

3.4.4. Foliage Penetration

Data obtained for foliage penetration [20] at low grazing-angles (<3°) is presented in Figure 3.18. It can be seen that the attenuation coefficient is not constant as it is with most homogenous materials, but increases with increasing foliage depth.

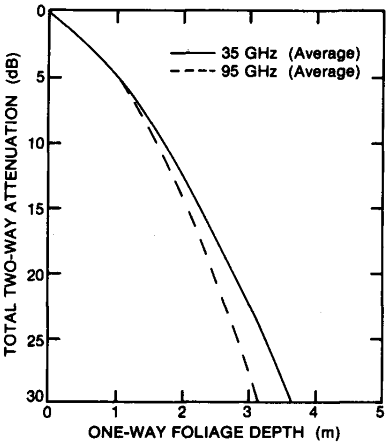


Figure 3.18: Total attenuation of a radar signal as a function of one-way foliage depth

There is little statistical difference between horizontal and vertical polarisation through trees [67] p171, but horizontally polarised signals are less attenuated through tall grass.

---

[2] The mass loading is 10g/m^3 which equates to an optical visibility of less than 5m

The attenuation is higher for wet foliage than for dry foliage, and tends to follow a power law relationship with frequency as opposed to the logarithmic law for dry foliage [67] p171

$$\alpha_{fol} = 1.102 + 1.48\log_{10} f_c, \tag{3.10}$$

where f_c is the frequency (GHz).

A more detailed model considers a tree to be made up of randomly aligned dielectric slabs that represent the individual leaves. An empirical relationship that relates the number of leaves N, to the diameter of the tree trunk D, at a height of 1.4m above the ground is shown. In this case the coefficients A and B are dependent on the tree type

$$\log_{10} N = A + B\log_{10} D. \tag{3.11}$$

The effective dielectric constant of the tree as a whole is then determined from the volume ratio of the two components (leaf and air), and the attenuation coefficient, α_{tot}, is calculated from the imaginary component

$$\alpha_{tot} = 4.343\frac{2\pi\varepsilon_{leaf}^{*}}{\lambda}\frac{V_{leaf}}{V_{tree}}, \tag{3.12}$$

where ε_{leaf}^{*} is the imaginary component of the relative dielectric constant of a leaf, V_{leaf} the leaf volume (m^3), V_{tree} the tree volume (m^3) and λ the wavelength (m).

In addition to the leaves there is also scattering and absorption by the branches and trunks of trees. Measured attenuation values of defoliated trees are used to represent these losses, to which are added the calculated foliage losses to produce the results shown in Figure 3.19.

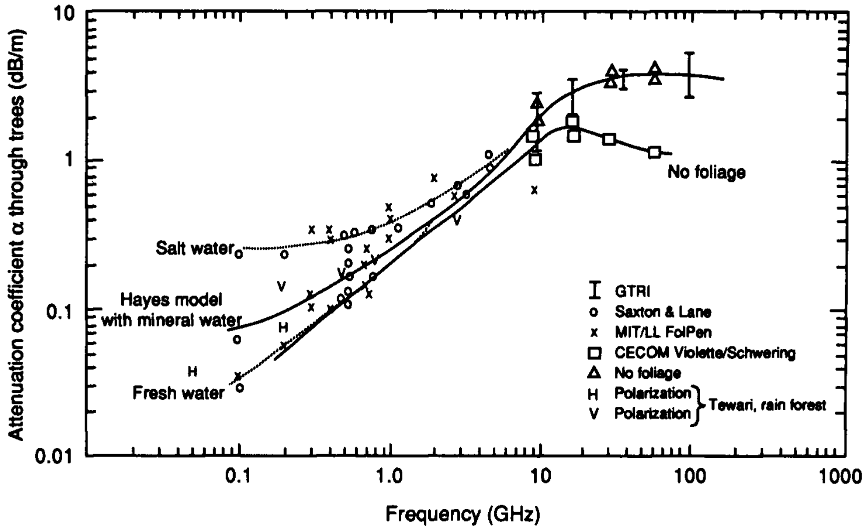


Figure 3.19: Measured and calculated attenuation coefficients through trees as a function of frequency for different water-salinity models

This model considers the average properties of the tree as a whole and does not consider the increasing foliage density with depth, as illustrated in Figure 3.18.

In comparison to dust attenuation where losses are generally of the order of a few dB/km, it can be seen that attenuation through foliage plays a much more significant role in restricting the propagation of millimetre wave radiation at low grazing angles.

3.5. Other Important Effects

Two other important effects can influence the generation of the radar images. They are multipath (described in detail in Appendix A) and electromagnetic interference.

3.5.1. Multipath Images

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, because the radar in an AGV is close to the ground, 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 as shown in the example presented in Figure 3.20.

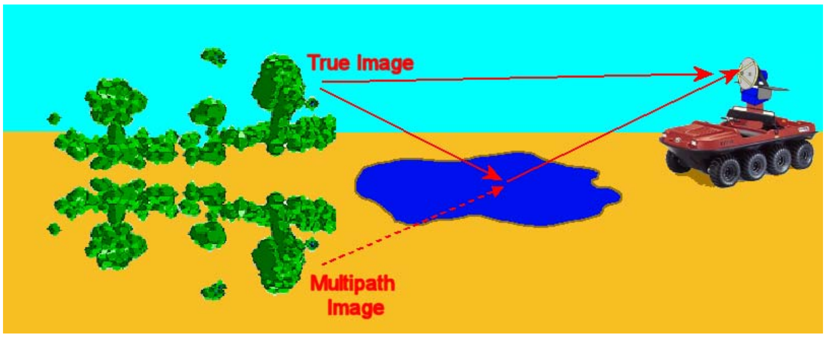


Figure 3.20: An illustration of the confusion that can arise in the interpretation of the ground level by an AGV fitted with a radar sensor, due to the effect of multipath

In addition to such specular returns from water, at very shallow grazing angles, earth and heavily grazed pasture would provide sufficiently smooth surfaces to appear Lloyds mirror like and multipath effects including severe target fading could occur. This mechanism is discussed in Appendix A and an example of this effect is shown in Figure 3.21.

In this figure it can be seen that fading occurs at 1.8km and 3.3km for this specific configuration of radar and target heights, and that the fading is so severe that the target is undetectable for more than 500m at one stage. The primary effect of this phenomenon would be to make apparently strong returns appear to fade in and out as the AGV approached.

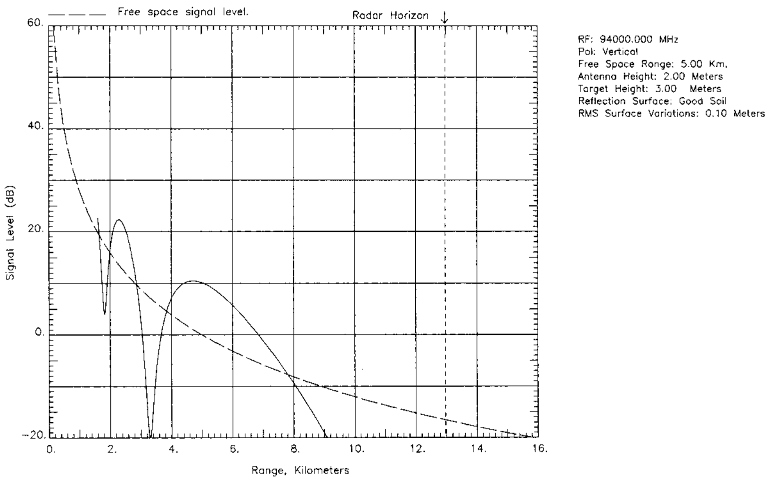


Figure 3.21: Multipath fading over flat earth (RMS surface variation 0.1m) for an antenna height of 2m and a target height of 3m

Multipath effects can also occur in bearing as shown in Figure 3.22 where an outdoor radar was used to produce a point-cloud image in a built-up environment.

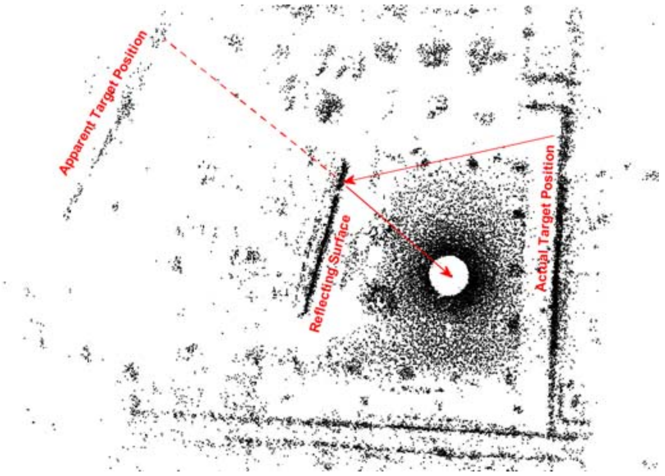


Figure 3.22: Effect of multipath on the bearing of target features as seen in this point-cloud radar image of the ACFR quad and adjacent car-park.

At shallow grazing-angles, in particular, the walls of the building can appear smooth to millimetre-wave radiation, it therefore reflects the radar signal off at an angle to illuminate an adjacent feature. The reflection of this feature appears to pass through the wall, so it would be easy for the vehicle to infer that the wall either did not exist or that there were holes through it (through which an AGV might pass). In addition, if the vehicle was using the reflected features for navigation it would estimate its position incorrectly.

3.5.2. Electromagnetic Interference

Electromagnetic Interference (EMI) is of concern insofar as any radar unit on an autonomous vehicle will only be one of a number of sensors or electrical systems. Though it is unlikely that any interference will have significant power in the millimetre-wave band where it is envisaged that any radar would operate, it would be capable of interfering with the IF or baseband signal if the appropriate precautions are not taken.

A second and more significant cause for concern is the possibility of mutual interference between the imaging radars of fleets of AGVs and UAVs operating in the same vicinity. A typical high range-resolution imaging radar would need a bandwidth of 1GHz to obtain a range resolution of 15cm. To avoid the possibility of mutual interference each unit operating within the same region would have to have its own unique frequency allocation. As this cannot always be achieved, a study has been undertaken (see Appendix B) to determine what the effects of the frequency overlap would be.

3.6. Conclusions

This chapter has examined the different methods of constructing radar images depending whether 2-D or 3-D information is required. The most common imaging technique discussed relies on a pulse-width-limited resolution cell to produce a 2-D image encoded for variations in reflectivity.

Measurements made by us and by others prove categorically that conventional FMCW radar systems are sufficiently sensitive to differentiate between differences in target reflectivity to produce high quality radar images of the ground even from low grazing-angles. These measurements confirm most of the differences in reflectivity are due to the differences in the surface roughness of the different terrain and vegetation types. They also show that even low-reflectivity surfaces like paved roads, runways and water can be identified by their higher reflectivity outlines.

It has been found that, because of the coherent nature of the ground reflectivity measurements made, speckle can reduce image clarity considerably. To overcome this problem, transmitted bandwidths in excess of the resolution requirements can be used to decorrelate the returns. An alternative solution is to integrate a number of images from different vantage points.

Additional information including polarisation or Doppler should be used if possible to augment the image quality. This would be particularly useful in discriminating between high-reflectivity natural and manmade objects in the case of the polarisation, and between rocky outcrops and trees in the Doppler case.

As regards propagation, it has been shown that the effects of dust and mist/fog will have a small effect on the performance of the radar if the appropriate band (94GHz) is used, but that rainfall and foliage will have a major effect on performance. Of primary concern are both the attenuation effects that reduce the amplitude of the ground return, and backscatter effects that generate a competing return. These two effects together must be considered when the transmitter power and the antenna beam shape are determined for long-range operations.

Finally it was shown that multipath effects will play a large part in determining the image quality, and could easily result in spurious returns both in elevation to confuse the radar interpretation of the slope of the land, and in azimuth which could displace features and generate errors in bearing.

Chapter 4.

Radar Systems

4.1. Introduction

The selection of an appropriate radar front-end for long-range operation requires a good understanding of both the basic principles of millimetre-wave radar in general and a knowledge of the systems that have been developed in the past for autonomous and semi-autonomous applications.

This chapter starts by introducing the concept of an ambiguity function which provides a quantitative measure of the radar resolution if the source (or target) is moving. This introduction is followed by a brief overview of some of the more common radar configurations that may be suitable for long-range imaging operations.

The conventional pulse time-of-flight system introduces the discussion before pulse-compression systems including SAW pulse-compression, phase coding and linear FM are considered. These systems are analysed briefly in terms of performance and complexity to determine which should be short-listed for further consideration.

A final selection of the long-range radar configuration is made at the conclusion of the chapter.

4.2. Modulation and the Ambiguity Function

A fixed-frequency continuous wave (CW) radar cannot resolve range because there is no unambiguous reference to determine the time of arrival of closely-spaced received echoes. If range is to be measured, as it must if images are to be generated, then some form of modulation must be applied to mark the CW carrier to allow for this. Intuitively it is obvious that that the briefer this modulation period is, the closer the spacing that can be resolved. However it can be shown from the properties of the Fourier transform that as this occurs, the transmitted bandwidth increases proportionally.

In a simple pulsed-AM radar, the spectral broadening is obtained by on/off modulation to produce a short pulse and then measure the round-trip time accurately. As the pulse width is decreased, the resolution with which range can be measured improves. At the same time, the accuracy with which the Doppler frequency can be measured degrades proportionally because of the short observation time.

In a CW radar, the spectral broadening is generally achieved by frequency or phase modulation of which the most common form is linear frequency-modulation commonly implemented as FMCW. In this case too, there is a direct link between the measured range and the target Doppler.

Though the interaction between the measurement of range and Doppler for pulsed and FMCW systems is easily understood, an analytical method to quantify the ability of a modulated waveform and its associated processing filter to resolve two or more targets at different ranges and velocities is required. A method to perform this exists and produces as an output the response function for the radar.

In Figure 4.1, the radar transmitter receives a complex modulation function $s(t)$ which is applied to the carrier and then radiated from the antenna. It strikes a target at range R moving with radial velocity v_r and the echo is received by the same antenna and downconverted to a convenient IF where it passes through a matched filter with frequency response $H(f)$ before envelope detection.

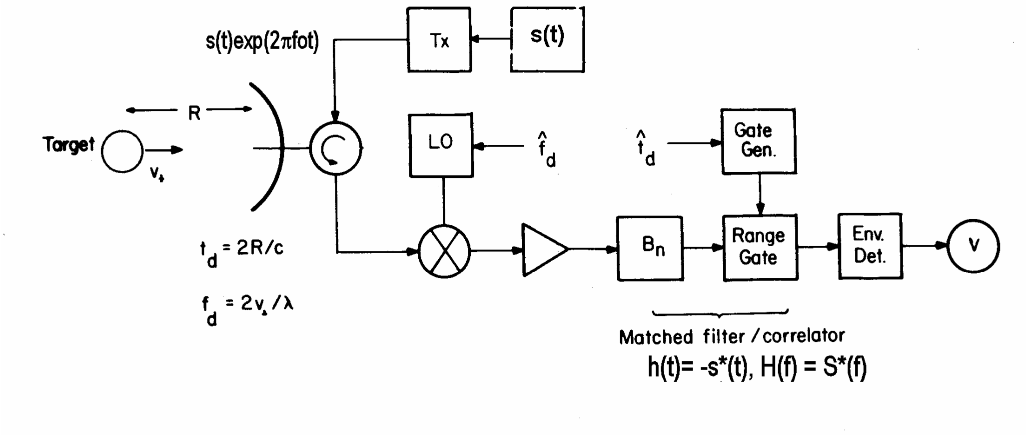


Figure 4.1: Block diagram illustrating the standard processing of a radar waveform

The receiving system is tuned for peak response to a target with specific values of delay and frequency, and the response of the system is measured in the delay-frequency plane with respect to this peak. This avoids problems associated with the realisability of filters having negative delay and permits the waveform and filter functions to be written without including the carrier term.

Ignoring propagation losses and assuming that $s(t)$ corresponds to the RMS level of the modulated carrier, then the receiver output will depend on the waveform and the filter properties, $\chi(t_d,f_d)$, where

$$\chi(t_d,f_d)=\int_{-\infty}^{\infty}H(f)S(f-f_d)\exp(j2\pi f_d t)df. \quad (4.1)$$

In this case, the two resolution coordinates are connected by the Fourier transform of time to frequency and vice versa.

As the matched-filter response is the complex conjugate, $S^*(f)$, of the original transmitted envelope voltage spectrum and G_a is a constant of proportionality with dimensions $1/(v.s)$ to keep the expression dimensionally correct, then

$$H(f) = G_a S^*(f). \tag{4.2}$$

When a radar processes the reflected echo through a matched filter the output response-function is

$$\chi(t_d, f_d) = \int_{-\infty}^{\infty} S^*(f).S(f - f_d)\exp(j2\pi f_d.t)df\,, \tag{4.3}$$

or in the time domain,

$$\chi_o(t_d, f_d) = \int_{-\infty}^{\infty} s(t - t_d).s(t)\exp(j2\pi f_d.t)dt\,. \tag{4.4}$$

From (4.4) it is clear that the response function, $\chi_o(t_d,f_d)$, is just the cross-correlation of a Doppler shifted version, $s(t)\exp(j2\pi f_d.t)$, with the unshifted waveform. The ambiguity diagram, as plotted in the following sections of the thesis, is the square of this response function $|\chi_o(t_d,f_d)|^2$ [189].

It can also be shown, though it is beyond the scope of this thesis to do so, that the volume under this surface remains constant for any matched filter [44] pp163-169. An extension of this derivation [180] shows that, in theory, a waveform can have as large a bandwidth as desired and so the range and velocity resolutions can be increased simultaneously without limit for a given signal to noise ratio.

This relationship is critically important to the successful development of radar systems that rely on pulse-compression techniques to achieve long-range performance without radiating astronomically high peak-powers, as it allows for the generation of “thumbtack” ambiguity functions in which all of the energy from the received pulse is compressed into an arbitrarily narrow, large amplitude peak.

Later sections of this and subsequent chapters examine different pulse-compression techniques, their implementation and some advantages and short comings of each as they relate to the autonomous guidance problem. However, prior to this, it is necessary to introduce the classical uncompressed pulse time-of-flight radar technology as a baseline reference by which the subsequent techniques can be judged.

4.3. Pulsed Radar

With the advent of solid-state components capable of operating in the millimetre-wave band, a number of development programmes were undertaken to produce small, high-resolution radar systems for various military and a few commercial applications [139], [191], [55].

A common radar configuration uses a single-stage non-coherent pulsed IMPATT (Impact Avalanche and Transit Time) transmitter operating at 94GHz with a second-harmonic Gunn LO and a conventional super-heterodyne receiver as shown in Figure 4.2. Because of the high cost and poor performance of amplifiers at 94GHz RF amplification is not often used and the mixer is fed directly from the antenna via a circulator.

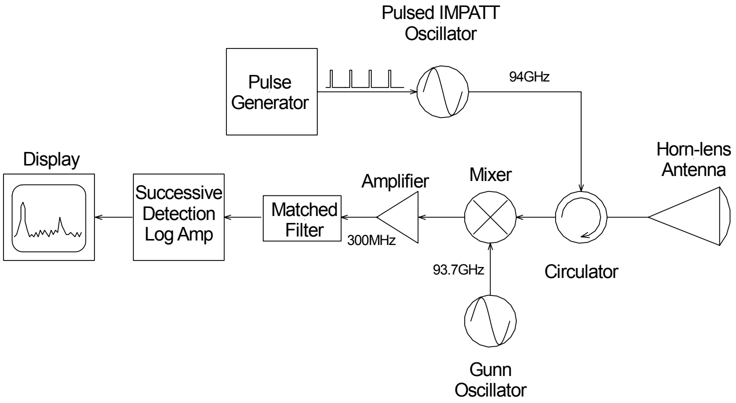


Figure 4.2: Schematic block diagram showing the major components of a typical non-coherent pulsed millimetre-wave radar system

For coherent operation, it is possible to implement the standard master oscillator power amplifier (MOPA) configuration that uses a chain of injection-locked IMPATT oscillators to carry the phase (and hence frequency) of a high-stability reference oscillator to the final pulsed oscillator to produce a fixed frequency pulse [73]. This configuration also allows for a synthesized low frequency chirp to be up-converted to 94GHz for pulse-compression applications [50].

In most applications, however, the standard free-running IMPATT power oscillator configuration is used because of its simplicity and low component count.

4.3.1. Transmitter Options

There are a number of different methods of generating pulsed signals in the millimetre-wave band. These include common tubes such as magnetrons and extended interaction Klystrons (EIK), both of which can produce pulsed output powers up to 6kW at 94GHz. However if the range is short and the power requirements are not too high, then solid state devices like pulsed IMPATT diodes which can produce a maximum of 22W per device are often used instead [67]. For even shorter ranges, pulsed Gunn oscillators or gated CW IMPATT and Gunn oscillators are alternatives.

One of the unique characteristics of pulsed IMPATT diodes is the generation of a chirp which is a consequence of the dramatic[3] junction temperature rise during the pulse. Models of this process are described by a number of authors [160], [100], and a number of others describe how this chirp can be compensated for by providing an upward ramp on the current pulse with the slope adjusted to balance the thermal effects [122], [87].

---

[3] A typical pulsed IMPATT diode with a junction diameter of 100μm will dissipate 200W during a 100ns pulse resulting in a typical rise in junction temperature of 200°C

Even with fairly sophisticated programmable pulsed current sources, it is not possible to eliminate this chirp completely, and this leads to compromises in the design and hence performance of the matched filter because the bandwidth is not determined solely from the pulse width but a combination of that and the residual chirp as discussed in Appendix C.

4.3.2. The Matched Filter

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 which is shown truncated in Figure 4.3.

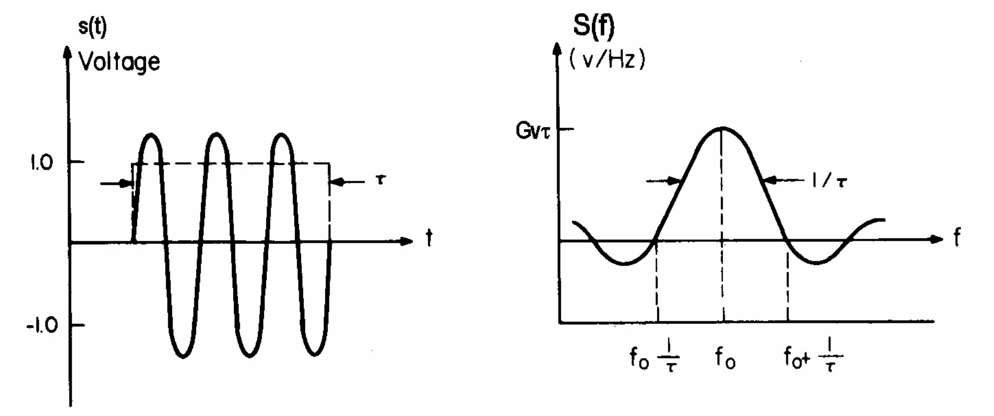


Figure 4.3: The relationship between the waveform and spectrum of a rectangular pulse

The spectrum $S(f)$ in V/Hz is given by the following formula where G_V = 1V is included to make the expression dimensionally correct

$$S(f) = G_V \tau \frac{\sin \pi\tau f}{\pi\tau f}. \tag{4.5}$$

The energy in Joules (W-s) in the pulse, assuming that the system is normalised to an impedance of 1Ω is then

$$E = \int_{-\infty}^{\infty} s(t)^2 dt = \int_{-\infty}^{\infty} S(f)^2 df = \tau. \tag{4.6}$$

The matched filter for this pulse has a rectangular impulse-response of duration τ and a frequency response identical to $S(f)$

$$H(f) = \frac{\sin \pi\tau f}{\pi\tau f}, \tag{4.7}$$

which makes the output spectrum in V/Hz

$$S_o(f) = S(f)H(f) = G_v \tau \frac{\sin^2 \pi\tau f}{(\pi\tau f)^2}. \tag{4.8}$$

The peak output (V) can then be calculated from the response function at zero time shift and zero Doppler to be

$$\chi_o(0,0) = \int_{-\infty}^{\infty} G_v \tau \frac{\sin^2 \pi \tau f}{(\pi \tau f)^2} = G_v = 1. \tag{4.9}$$

For the normalised impedance of 1Ω, the peak output power will be 1W, so if white noise is present with a uniform density N_o (W/Hz), the noise output power is

$$N = N_o \int_{-\infty}^{\infty} |H(f)|^2 df = N_o B_n, \tag{4.10}$$

where B_n is the effective noise bandwidth of the filter (Hz).

The envelope of the output waveform obtained by convolution of $s(t)h(t)$ or from the Fourier transform of $S(f)H(f)$ is a triangular function shown in Figure 4.4.

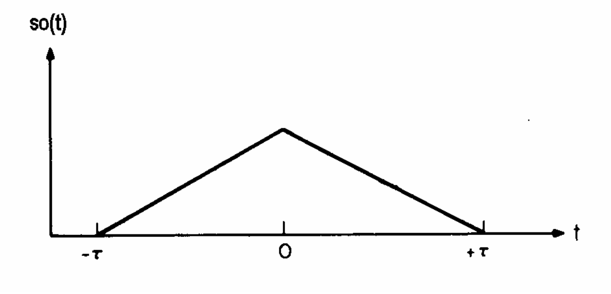


Figure 4.4: Matched-filter output for a generic rectangular pulse of duration τ

As a target is moved to other values of range and Doppler frequency, the peak output signal from the matched filter will decrease from the maximum. If these values are plotted on the delay-frequency surface for a 100ns long rectangular pulse, the full ambiguity function can be obtained as shown in Figure 4.5.

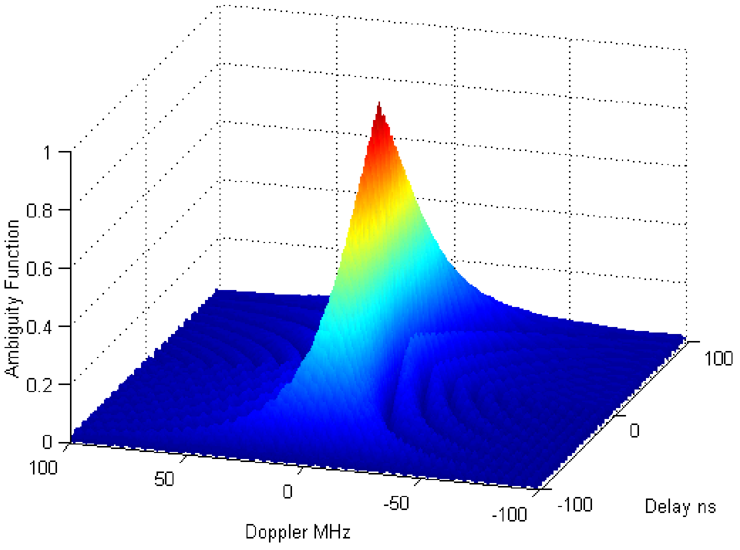


Figure 4.5: Ambiguity function for the matched-filter response to a 100ns long rectangular pulse

A convenient method of representing this function is as a contour plot which is generated by cutting the surface parallel to the time-frequency plane at various heights to generate a series of concentric ellipses. The cut computed at a height of 0.75 is referred to as Helstrom's uncertainty ellipse and is often used as a graphical representation of the ambiguity function as shown in Figure 4.6.

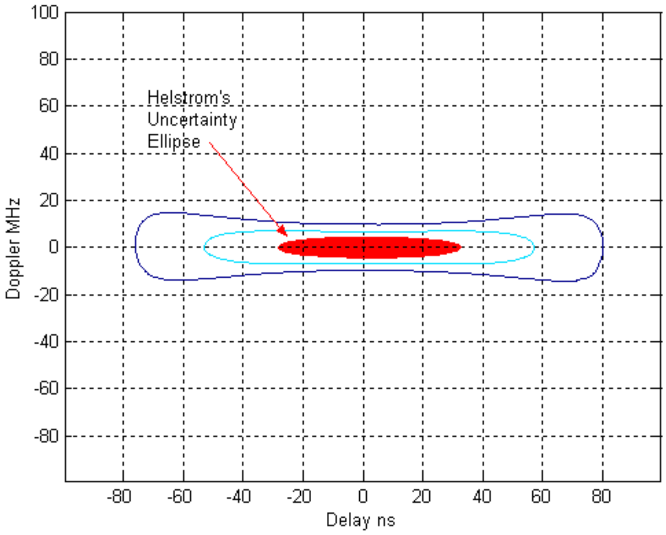


Figure 4.6: Contour plot and Helstrom's uncertainty ellipse of the ambiguity function for the matched-filter response to a 100ns rectangular pulse

When a rectangular pulse is processed through a matched filter it produces a triangular output envelope 2τ wide at the base and with a well defined central peak as shown in Figure 4.4. 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 in Figure 4.7. For all phase angles the second peak is still identifiable and the two targets are said to be resolved in range.

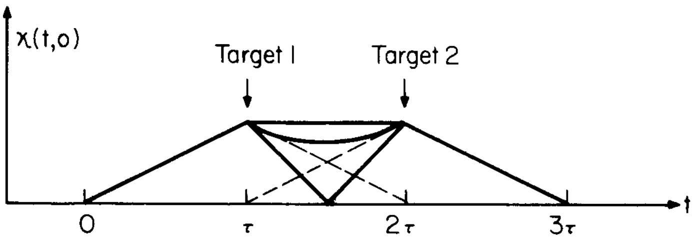


Figure 4.7: Matched-filter output of a pair of closely spaced targets showing the limits to the range resolution at zero Doppler

4.3.3. Potential for Autonomous Applications

There is no doubt that high-resolution millimetre-wave radar systems like the one described in this chapter (and in Appendix C) offer a number of extremely useful functions to facilitate the difficult process of guiding a UAV or AGV.

Target tracking of prominent features on the ground can be used to estimate the radar position with good accuracy using SLAM [149] or any of the more conventional feature-association methods discussed in Section 1.5. This ability is particularly well supported by radar systems when a single or a limited number of features is used.

Images made at very low grazing-angles cannot rely on the radars tracking the ground, so the image shown in Figure 4.8 was produced by controlling both the position of the range gate and the beam angle to produce a raster in range and azimuth.

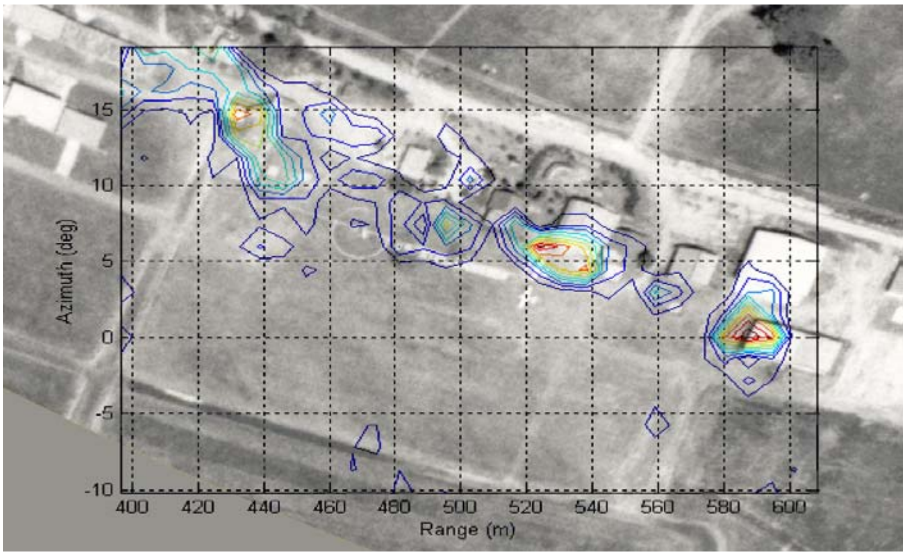


Figure 4.8: Pulsed millimetre-wave radar reflectivity contour map superimposed on an aerial photograph of the same airfield scene showing high intensity returns produced by buildings

In this image, it can be seen that buildings generally reflect well, with the largest signals being returned by corrugated-iron hangars (as one would expect). These strong and stable echoes are ideal candidates for the feature association process if the required accuracy is not high. However, because of the uncertainty in identifying which buildings will reflect well, and even which parts of those buildings are the actual reflectors, accurate association is not possible. From an autonomous navigation perspective, the ability to produce and interpret radar images is of paramount importance insofar as they can supply information not available from any other source.

The ability of even the simple pulsed radar system, discussed in Appendix C, to measure the slope of the surrounding terrain is advantageous to the autonomous navigation process where an AGV may be required to traverse unknown unstructured terrain. This process is an extension of the standard occupancy grid mechanism introduced in Section 2.3.4 [138], [79] that identifies potential obstacles but does not characterise the difficulty of traversing them. In conjunction with the clues offered by changes in surface reflectivity and shadowing, this slope-estimation algorithm could become a powerful path-planning tool.

The use of a pulsed IMPATT radar system for autonomous applications is unlikely for a number of reasons, the primary one being that pulsed IMPATT oscillators are temperamental and still require manual tuning to obtain good performance. A second consideration, also concerned with hardware, is that a pulsed radar requires two millimetre-wave oscillators; one transmitter and one local oscillator. This makes them expensive compared to the homodyne continuous wave options that are discussed later.

The performance of a pulsed system can easily be extrapolated to long-range operation as specified from the measurements reported in Appendix C. It is assumed that a state-of-the-art IMPATT transmitter with an output power of 20W, and a fan-beam antenna with a gain of 44dB are used. The range and the expected RCS are as specified in Chapter 1 for long-range operation.

Table 4.1: Pulsed radar performance at long range

Contributor	Existing Short Range	Feasible Long Range	Signal Difference
Tx Power	7dBW	13dBW	+6dB
Antenna Gain	37.5dB	44dB	+13dB
RCS	+10dBm^2	-3dBm^2	-13dB
Range	146m	3000m	-52dB
SNR	32dB	-14dB	-46dB

The SNR would thus be reduced from 32dB to 32 - 46 = -14dB even without considering the effects of atmospheric attenuation. For a 10dB SNR that is adequate for imaging applications, the transmitter power would have to be increased by 24dB, from 13dBW to 37dBW (5kW). Such high power could be produced by a magnetron or EIK, but both of these tubes require heavy and expensive power supplies and other ancillary components which are impractical for a small low-cost radar.

The only practical method to achieve the specified range and resolution while still transmitting a lower power is to use one of the pulse-compression techniques that are discussed below.

4.4. Pulse Compression

From a more formal perspective, the use of a real waveform $s(t)$ imposes two serious restrictions on the shape of the ambiguity function:

- The product of the delay and the Doppler resolution remains close to unity as shown earlier.
- The pulse energy is directly proportional to the pulse width, and hence, to the resolution.

These problems can be avoided by combining the powers of many small transmitters spatially in a phased array, and by introducing complex modulation schemes which are most often accomplished by applying a process called pulse-compression.

Pulse-compression techniques are employed to divorce the usual signal-bandwidth (range-resolution) effects from the transmitted pulse-length. The bandwidth to be transmitted is increased by modulating the signal within the transmitted pulse in amplitude, phase or frequency. For a transmitted pulse length τ_t the returned echo pulse can be processed through a matched filter and compressed into a pulse with a duration τ_c which is approximately equal to the reciprocal of the transmitted bandwidth β_t. The ratio of the transmitted to compressed pulse-widths is called the pulse-compression ratio ρ and is given by

$$\rho = \beta_t \tau_t = \frac{\tau_t}{\tau_c}. \tag{4.11}$$

Values of between 1000 [67] p628 and 10,000 [44] p173 are available now using standard surface acoustic wave (SAW) devices, but this will be increased significantly in the future. A radar with a pulse-compression ratio of 2,000,000 is described in [44] pp175-180.

The way in which pulse compression affects the radar range for a matched system is that the maximum detection range is proportional to the product of the transmitter power P_t and the pulse width τ_t

$$R \propto \left(P_t \tau_t\right)^{1/4}. \tag{4.12}$$

There are advantages to producing a system with the longest real pulse-width possible within the constraints of the minimum operational range, while maintaining good range resolution equal to the compressed pulse width τ_c. Thus, the ideal system would transmit a signal with a duration equal to the round-trip time to the nearest target, and then listen for echoes during a similar period.

The process of compressing the pulse using a matched filter produces range sidelobes in the form of ripples that precede and follow the main pulse. These range sidelobes can hide smaller target returns in the vicinity of a large echo. For this reason many systems apply a mismatch stage to reduce the range sidelobes, but in the process the SNR is reduced and the range resolution is degraded.

Many different approaches to pulse compression have been developed [129]. The better known codes include the following, of which some are discussed in detail in this thesis:

- Barker binary phase
- Pseudo-random binary phase
- Random binary phase
- Step frequency modulation
- Linear frequency modulation
- Nonlinear frequency modulation
- Step frequency derived polyphase (Frank and P1 codes)
- Butler matrix derived polyphase (P2 code)
- Linear frequency derived polyphase (P3 and P4 codes)
- Huffman codes
- Complementary codes

It can be seen from this list that there are two basic approaches to pulse compression. The first involves subdividing the transmitted pulse into short chips to which a pseudorandom-phase code is applied, or to introduce over the transmitted pulse a continuous phase modulation, such as a quadratic function that produces linear FM.

The binary phase codes exhibit a thumbtack ambiguity diagram in which the main-lobe cell-width is reduced to $\tau_c = 1/\beta$ in delay while it remains $\beta_o = 1/\tau$ in Doppler. This reduces the resolution area in the main lobe of the ambiguity diagram, from unity to $1/\beta\tau$. The ambiguity diagram consists of a single narrow spike surrounded by a plateau of time and Doppler sidelobes.

The Barker codes have been called the perfect codes because the highest normalised time sidelobes at zero Doppler are only one element high. Unfortunately the longest Barker code consists of 13 chips with a pulse-compression ratio of 13. Other phase codes that have similar qualities to the Barker codes such as the pseudo-random codes can provide very large pulse-compression ratios, but their peak sidelobe powers are not as low.

4.4.1. Matched-filter Definition

A pulse-compression radar is the practical implementation of a matched-filter system [176] p10.2 as shown schematically in Figure 4.9 below.

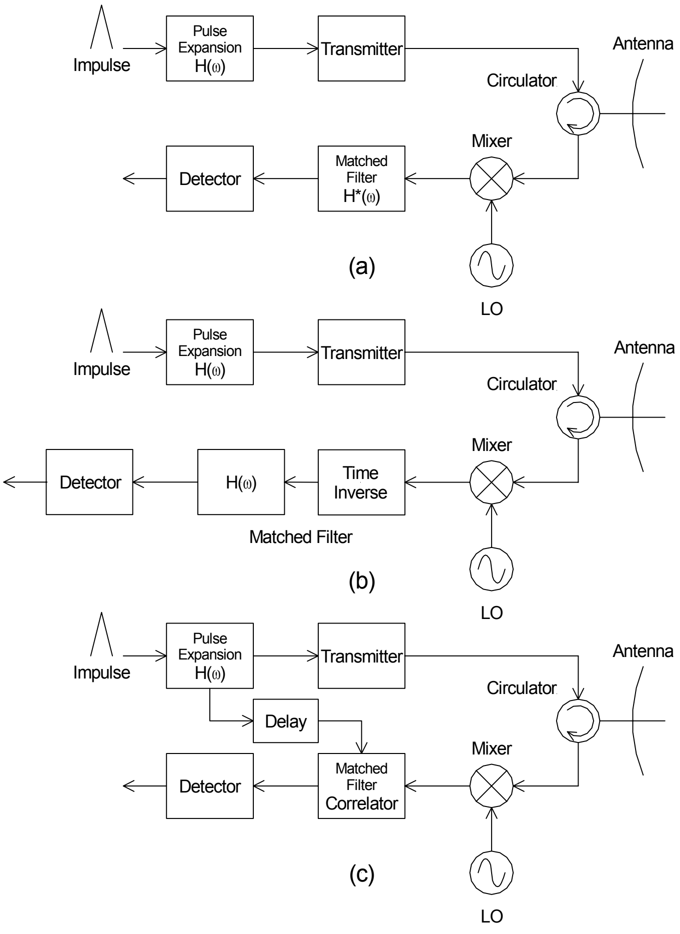


Figure 4.9: Matched-filter configurations for pulse compression using (a) conjugate filters, (b) time inversion and (c) correlation

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

$$y(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} |H(\omega)|^2 \exp(j\omega t)d\omega . \tag{4.13}$$

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

$$y(t) = \int_{-\infty}^{\infty} h(\tau)h^*(t-\tau)d\tau . \tag{4.14}$$

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 4.9c.

4.4.2. Phase-Coded Pulse Compression

The simplest way to understand pulse compression with a matched filter is to consider one of the shorter Barker codes in a binary phase-shift keying (BPSK) modulation technique. 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. 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 Figure 4.3. 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 in Figure 4.10.

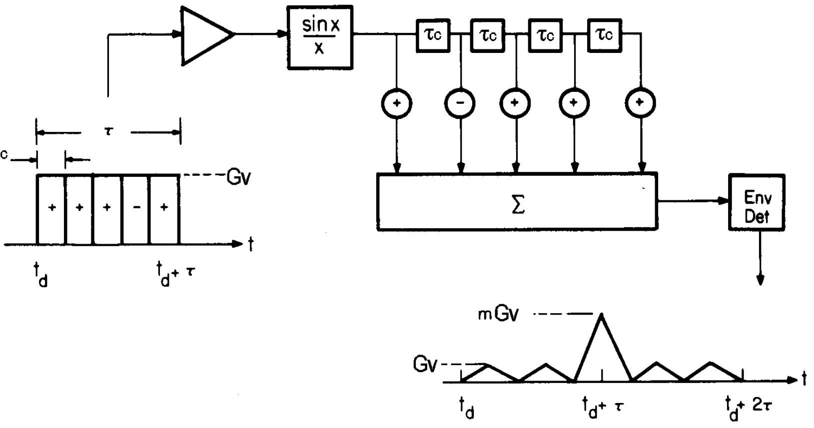


Figure 4.10: Diagram to illustrate the concept of phase-coded pulse compression for a five bit Barker code

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 $1/m^2 = -13$dB.

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 in Figure 4.11 shows that the main-lobe decays quickly and sidelobes increase rapidly with increasing Doppler.

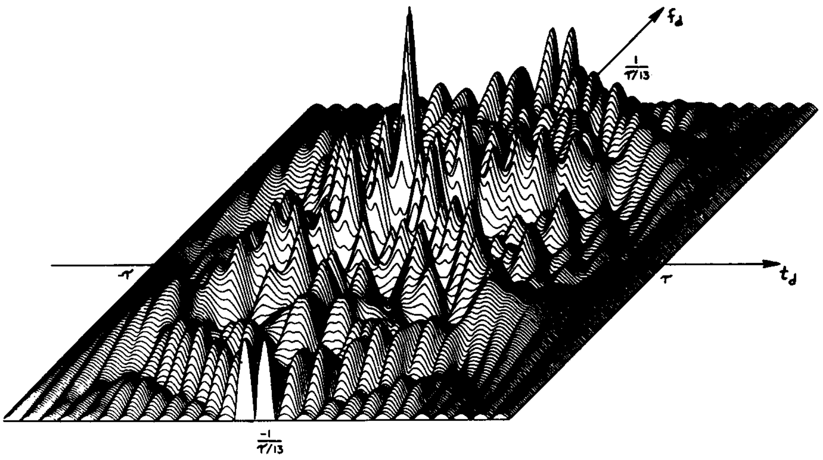


Figure 4.11: Ambiguity diagram for a 13-bit Barker code showing the "thumbtack" main lobe decaying into a sea of increasing delay and Doppler sidelobes

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.

Another fundamental problem with modulation for this application is Doppler sensitivity. Due to the high operational-frequency, Doppler shifts will be large even for an imaging radar mounted on a slowly moving AGV, and so the radar resolution will be degraded over most of the image.

4.4.3. SAW Based Pulse Compression

In a pulse-compression system such as that shown in Figure 4.12, 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. 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. 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.

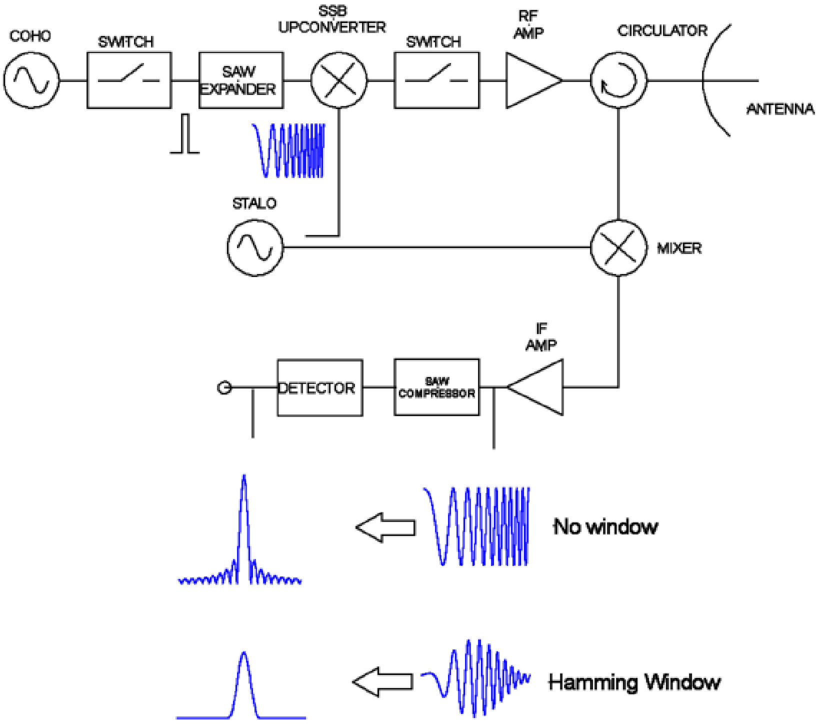


Figure 4.12: Conceptual diagram of a linear-chirp pulse-compression radar

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) 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.

4.4.4. 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

$$\tau_t = \frac{N}{\Delta f} = N\tau_f, \tag{4.15}$$

and the transmitted bandwidth β_t is

$$\beta_t = N\Delta f = \frac{1}{\tau_c}, \tag{4.16}$$

making the pulse-compression ratio

$$\rho = \frac{\tau_t}{\tau_c} = \tau_t \beta_t. \tag{4.17}$$

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 [128].

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 as discussed in Appendix F.

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.

4.5. Frequency Modulated Continuous Wave Radar

4.5.1. Operational Principles

The schematic block diagram in Figure 4.13 shows the structure of a homodyne radar[4]. In this case, the CW signal is modulated in frequency to produce a linear chirp which is radiated toward a target through an antenna.

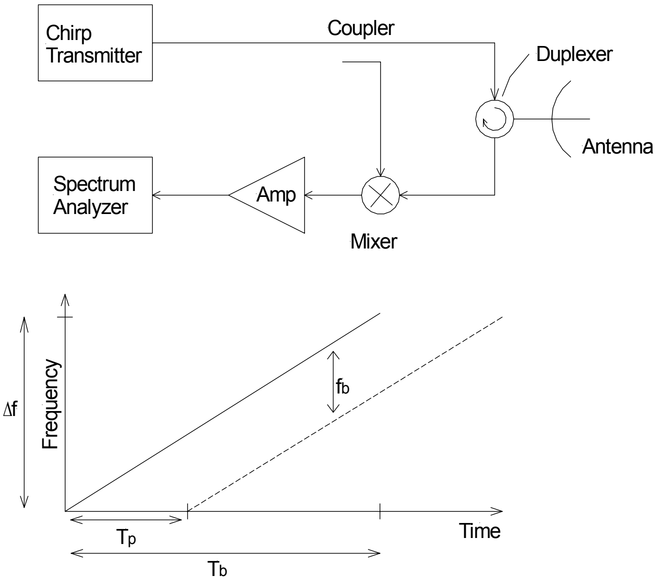


Figure 4.13: Schematic diagram illustrating the FMCW concept

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

$$\omega_b = A_b t \,, \tag{4.18}$$

substituting into the standard equation for FM results in

$$v_{fm}(t) = A_c \cos\left[\omega_c t + A_b \int_{-\infty}^{t} t\,dt\right], \tag{4.19a}$$

$$v_{fm}(t) = A_c \cos\left[\omega_c t + \frac{A_b}{2} t^2\right]. \tag{4.19b}$$

---

[4] a CW radar in which the microwave oscillator serves as both the transmitter and local oscillator

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.

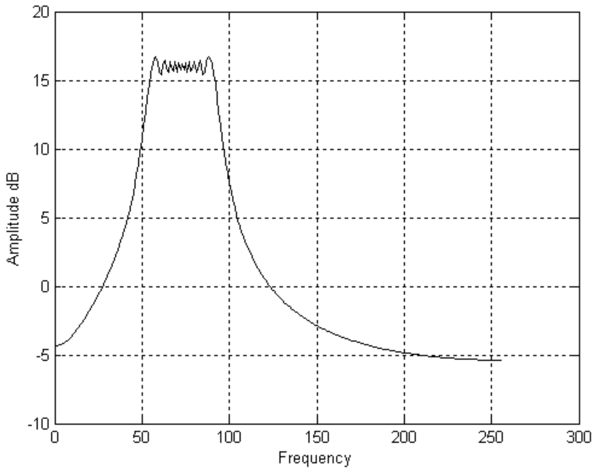


Figure 4.14: Frequency domain representation of a linear FM chirp

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

$$v_{fm}\left(t-T_p\right)=A_C\cos\left[\omega_c\left(t-T_p\right)+\frac{A_b}{2}\left(t-T_p\right)^2\right]. \tag{4.20}$$

Calculating the product of (4.19) and (4.20),

$$v_{fm}\left(t-T_p\right)v_{fm}\left(t\right)=A_c^2\cos\left[\omega_c t+\frac{A_b}{2}t^2\right]\cos\left[\omega_c\left(t-T_p\right)+\frac{A_b}{2}\left(t-T_P\right)^2\right]. \tag{4.21}$$

Equating using the trigonometric identity for the product of two sines,

$$\cos A\cos B=0.5\left[\cos\left(A+B\right)+\cos\left(A-B\right)\right], \tag{4.22}$$

$$v_{out}(t)=\frac{A_c^2}{2}\left[\begin{array}{l}\cos\left\{\left(2\omega_c-A_bT_p\right)t+A_bt^2+\left(\frac{A_b}{2}T_p{}^2-\omega_cT_p\right)\right\}\\+\cos\left\{A_bT_p.t+\left(\omega_cT_p-\frac{A_b}{2}T_p{}^2\right)\right\}\end{array}\right]. \tag{4.23}$$

The first cosine-term in (4.23) 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,

$$f_b = \frac{1}{2\pi}\frac{d}{dt}\left[A_b T_p t + \left(\omega_c T_p + \frac{A_b}{2}T_p^2\right)\right], \tag{4.24}$$

$$f_b = \frac{A_b}{2\pi}T_p. \tag{4.25}$$

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 in Figure 4.15 includes both the fixed and chirp terms for illustrative purposes, but in general only the low frequency component is output.

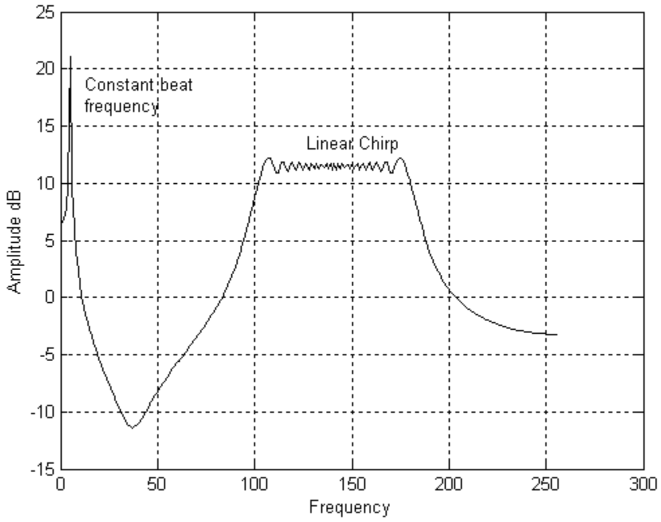


Figure 4.15: Frequency domain representation of the FMCW receiver output including both the high and low frequency components after mixing but before filtering

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 [109], [130].

4.5.2. 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 4.13. 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.

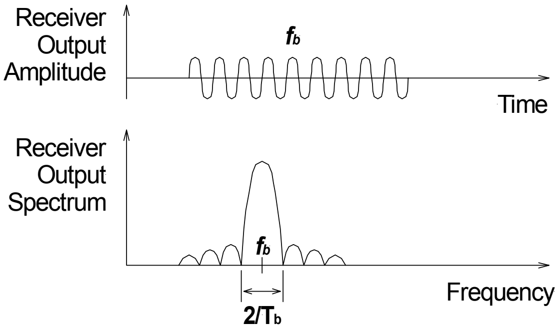


Figure 4.16: Spectrum of the truncated sinusoidal signal output by an FMCW radar

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

$$f_b = \frac{A_b}{2\pi} T_p = \frac{\Delta f}{T_b} T_p . \qquad (4.26)$$

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

$$T_p = \frac{2R}{c}, \qquad (4.27)$$

and substituting into (4.26) gives the classical FMCW formula that relates the beat frequency and the target range

$$f_b = \frac{\Delta f}{T_d} \frac{2R}{c} . \qquad (4.28)$$

For a frequency resolution, δf, (4.28) can be used to show that the range and range resolution, δR, is

$$R = \frac{T_d c}{2\Delta f} f_b , \qquad (4.29)$$

$$\delta R = \frac{T_d c}{2\Delta f} \delta f_b . \qquad (4.30)$$

It was shown earlier that $\delta f_b = 1/T_d$, which when substituted into (4.30) results in a closed relationship between the total transmitted bandwidth and the range resolution

$$\delta R = \frac{c}{2\Delta f}\,. \tag{4.31}$$

This is intuitively quite satisfying at it represents the FMCW equivalent of the classical pulsed radar range-resolution equation where $\tau = 1/\Delta f$.

The analysis presented here assumes that the chirp is completely linear with time. However, in most practical applications this is not true and the analysis in Appendix D shows 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

$$\delta R = R.Lin\,, \tag{4.32}$$

where the linearity, Lin, is defined as the change in chirp slope, $S = df/dt$, normalised by the minimum slope.

$$Lin = \frac{S_{\max} - S_{\min}}{S_{\min}}\,. \tag{4.33}$$

This sensitivity to slope linearity is one of the fundamental problems that limits the resolution of real FMCW radar systems. Techniques to circumvent this limitation are addressed in Appendix E.

4.5.3. The Ambiguity Function

The Ambiguity Function for linear FM pulse compression, Stretch[5] [44] pp.156-157 and for FMCW, is reproduced in Figure 4.17. 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 (4.20) with its Doppler shifted counterpart,

$$v_{fm}(t - T_p) = A_c \cos\left[\omega_c\left(t - T_p\right) + \frac{A_b}{2}\left(t - T_p\right)^2 - \frac{2v_r}{c}\omega_c\left(t - T_p\right)\right]. \tag{4.34}$$

Processing as before to determine the new beat frequency f_b, it is just the old beat frequency offset by the Doppler shift

$$f_b = \frac{2v_r}{c} f_c - \frac{A_b}{2\pi} T_p = f_d - \frac{A_b}{2\pi} T_p\,. \tag{4.35}$$

---

[5] 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.

In this case, the ambiguity function can be expressed as

$$\left|\chi_o(t_d,f_d)\right|^2 = \left[\frac{\sin\left[\pi\left(f_d - \frac{A_b}{2\pi}t_d\right)\left(T_b - |t_d|\right)\right]}{\pi\left(f_d - \frac{A_b}{2\pi}t_d\right)\left(T_b - |t_d|\right)}\left(1 - \frac{|t_d|}{T_b}\right)\right]^2 \quad \text{for} -T_b \le t_d \le T_b \quad (4.36)$$

$$= 0 \qquad \text{for } |t_d| > T_b$$

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.

A cut along the Doppler axis is similar to that of the single pulse discussed in Section 4.3.2 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 Figure 4.17 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.

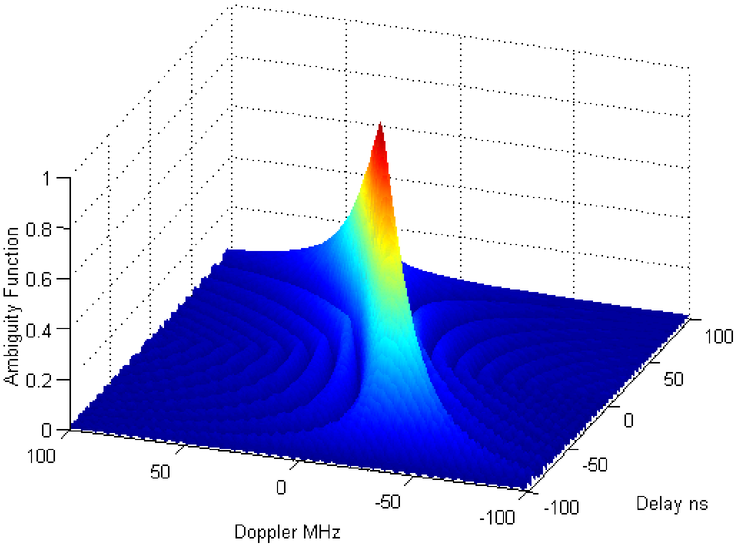


Figure 4.17: Linear FM up chirp ambiguity diagram for a 100ns duration signal showing the interaction between delay and Doppler

4.5.4. Phase Noise and Leakage Effects

For a monostatic FMCW system which transmits and receives simultaneously, it is not possible to isolate the receiver from the transmitter completely, and some power leakage occurs via the paths shown in Figure 4.18. This has the effect of injecting phase noise from the transmitter into the receiver.

This leakage power has three major effects that are discussed briefly below, with more information available in Appendix F. Leakage may result in desensitisation of the receiver [154]; it will lift the noise floor, particularly at short range; and it will generate phase-noise skirts around each target return and this reduces image contrast.

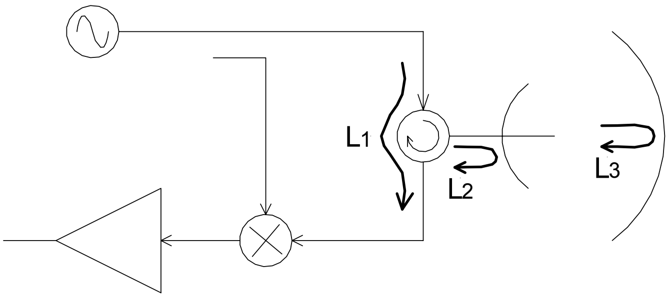


Figure 4.18: Transmitter signal leakage paths in an FMCW front-end

For a classical FMCW design based on an InP Gunn VCO, and a moderately low leakage level, L_2+L_3 = -20dB, then the noise levels into a 1kHz bandwidth (derived in Appendix F) are as shown in Figure 4.19.

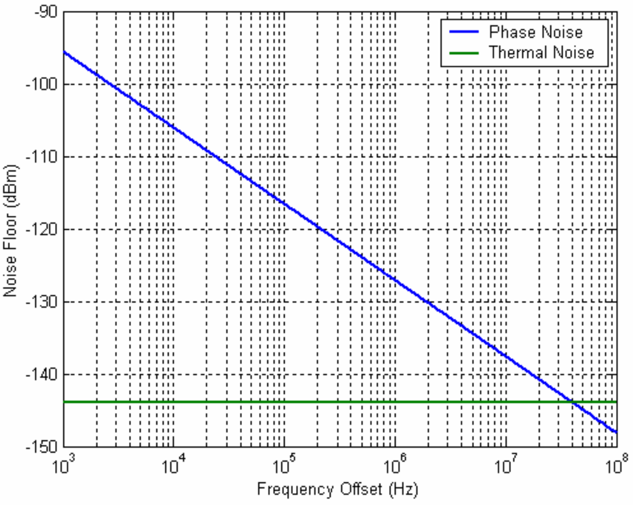


Figure 4.19: Comparison between phase-noise and thermal noise at the mixer output of a typical Gunn-oscillator based millimetre-wave FMCW radar

It can be seen that the performance of the radar is severely compromised at short range where the phase-noise level exceeds that of the thermal noise by many dB. Approaches to improve performance include using better matched and lower leakage components, generating a chirp with lower phase-noise [31] and introducing additional circuitry to actively cancel the leakage. Recent work [163] has demonstrated a moderate-bandwidth cancellation technique that reduces the leakage by >10dB over a 150MHz bandwidth. The addition of an EH tuner between the circulator and the antenna allows the frequency of the null to be adjusted to suit the characteristics of the transmitter [136].

As regards the issue of image contrast reduction, Figure 4.20 shows that the phase noise levels surrounding a target are proportional to the amplitude of the return and that they taper off slowly with frequency. In this example, the contrast between the target amplitude and the noise level in the adjacent bins is only 14dB.

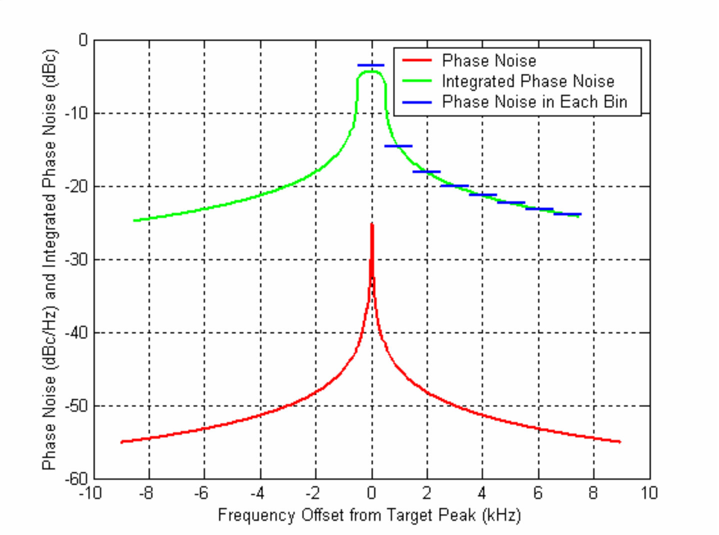


Figure 4.20: Phase-noise and integrated phase-noise in each 1kHz range-bin around a point target at a range of 400m

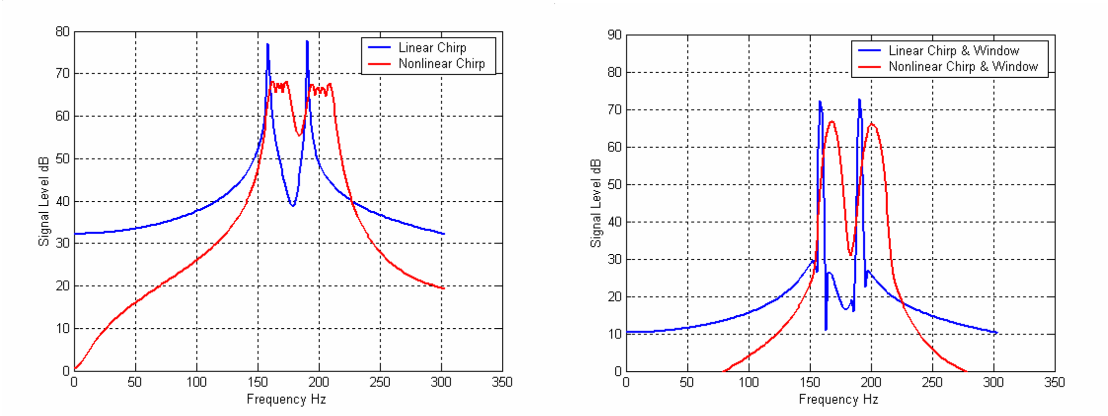
For large amplitude returns, these phase skirts rise above the noise floor to produce bright radial lines on images (see Appendix F for an in-depth explanation of this phenomenon).

4.5.5. Radar Performance Evaluation

A computer model has been written that combines all of the effects discussed in the chapter so far, to examine the returns from point targets.

In the following figures, the effect on the output spectrum of increasing the nonlinearity of the chirp signal and adding phase-noise is shown in comparison to the perfectly linear clean reference.

Figure 4.21 shows the theoretical output signal for two closely-spaced point targets if the oscillator chirp is completely linear, and if a nonlinearity of 10% is introduced for the rectangular (uniform weighting) and Hamming window cases. The two important differences between the linear and nonlinear returns are that the nonlinear return peaks are broader and their level is lower. The fact that they are also displaced in frequency is an artefact of the modelling process. The addition of a positive quadratic error to the linear chirp increases the chirp slope which in turn leads to an increasing beat frequency over a single sweep period.



(a) (b)

Figure 4.21: Effect of quadratic chirp-nonlinearity on the spectra of two closely-spaced point targets for (a) a rectangular and (b) a Hamming window

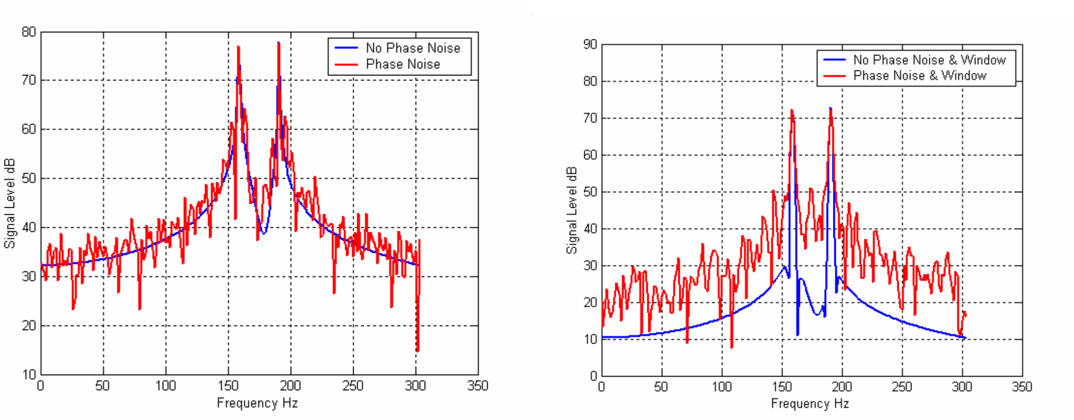
The well-behaved broadening is the residual chirp in the beat signal due to the quadratic chirp-nonlinearity modelled in this simulation. In most systems, where the nonlinear chirp is not necessarily quadratic in nature, the spread will not be uniform, and often results in the formation of a number of distinct peaks. An interesting side effect of the nonlinearity visible in Figure 4.21 is that the amplitudes of the far-out sidelobes are reduced when compared to the linear case.

The decrease in the signal level is the inevitable result of this spectral spreading because, with the same total power in each echo, the power density will be lower if it is distributed over a wider bandwidth. In this case the FFT bin width is 1.5Hz and the bandwidth of the nonlinear signal is 16Hz, so the power is spread uniformly over about 10 bins in the case of the rectangular window, with a resulting decrease in power density of $10\log_{10}(10) = 10$dB. The magnitude of this decrease is confirmed in Figure 4.21.

Interestingly, the application of the Hamming window to the received spectrum reduces the amplitude difference to about 5dB as well as decreasing the difference in resolution between the nonlinear and the linear signals. The primary reason for this is that the Hamming window reduces the contributions of the frequencies from extremes of the chirping target signal.

In this simulation model, the phase-noise is generated by passing normally-distributed white noise through a perfect integrator and adding it to the carrier phase to produce an f^2 power spectrum around the carrier. This spectrum is modified by the homodyne mixing process as discussed earlier in the chapter.

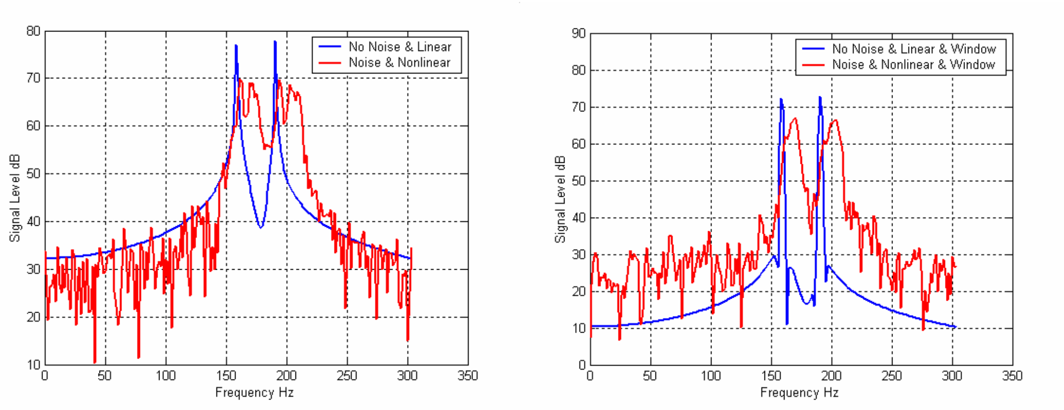
It can be seen from Figure 4.22 that the phase-noise does not widen the peak appreciably but it raises the level of the noise skirts that straddle the peak. This is particularly noticeable in the Hamming windowed signal in which the close-in sidelobe levels are increased by more than 20dB by the phase-noise.



(a) (b)

Figure 4.22: Effect of phase-noise on the spectra of two closely-spaced targets for (a) rectangular and (b) a Hamming window

Finally, in Figure 4.23 the combined effect of the introduction of both a nonlinear chirp and phase-noise to produce realistic radar outputs is shown.



(a) (b)

Figure 4.23: Effect of nonlinear chirp and phase-noise on the spectra of two close spaced targets for (a) rectangular and (b) a Hamming window

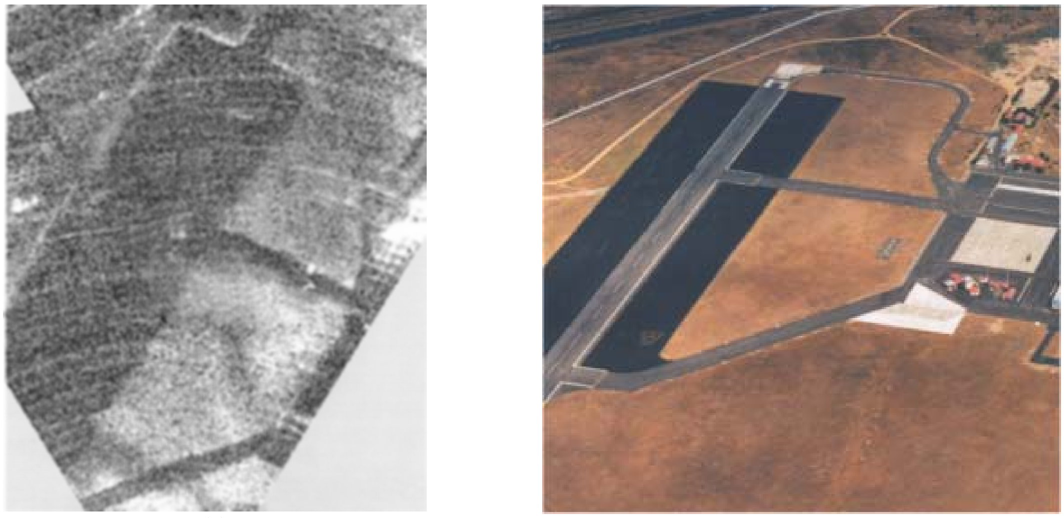
This model does not incorporate the effects of AM noise leakage (discussed in Appendix F) which also exceeds the thermal-noise level by a significant margin.

4.5.6. Potential for Autonomous Applications

A number of FMCW radar systems have been built and tested for both ground-based (see Chapter 3), and airborne applications. In general the quality of the images is a little disappointing in the latter, as can be seen for the image made from a hovering helicopter shown in Figure 4.24. This is probably due to the fact that the system performance is being pushed to the limit in terms of the long operational-range and low transmitter-power in conjunction with the notoriously poor noise-figure intrinsic to the FMCW process.

The radar used for these flight trials was a closed-loop linearised FMCW unit with a chirp bandwidth of 150MHz, a sweep time of 1ms and a transmit power of 10mW at 94GHz. The antenna used was a 300mm diameter Cassegrain unit that had been modified to produce a fan-beam.

A number of other issues contribute to the relatively poor image quality. These include problems with maintaining a sufficiently stable airborne platform for image integration if no INS is available. Uncompensated base motion results in image blurring. In addition artefacts of the chirp-linearisation process generate the bright arcs that bisect the image.



(a) (b)

Figure 4.24: Airborne images of a section of runway and taxiway showing (a) the composite millimetre-wave radar image and (b) an aerial photograph

Measurements made on a number of images not reproduced here are summarised in Table 4.2. They confirm that the runway reflectivity is so low that the returns are below the noise floor for ranges beyond 250m, and that the maximum possible operational range is less than 800m.

Table 4.2: Reflectivity contrast analysis for Integrated FMCW images

	Image A	Image B	Image B	Image B	Image B
Range	450m	480m	800m	250m	300-700m
Runway	-78.8	-79.2	-78.7	-76.7	-78
Verge	-75.7	-75.1	-77.5	-69.2	-74.2
Contrast	3.1dB	4.1dB	1.2dB	7.5dB	3.8dB

This section has shown that a conventional FMCW radar system is limited by phase-noise and AM noise leakage from the transmitter. Therefore increasing transmitted power does not improve the operational range.

The two techniques than can be used to counteract this problem are operation from a pair of antennas to minimise leakage or the implementation of an active reflected power canceller. The former is not really feasible for an airborne application because the antenna isolation will be compromised by a radome, and the latter requires millimetre-wave components that are not readily available.

A simple non-adaptive reflected power cancellation technique can be applied if the antenna VSWR remains constant. However as soon as a scanned antenna is placed behind a fixed radome, as is required for airborne imaging, the VSWR changes with scan angle and this technique is no longer effective.

In summary is not feasible, without additional hardware, to use millimetre-wave FMCW radar to produce images from an airborne platform at a range of more than about 500m.

4.6. Interrupted FMCW

One solution to the problems described above is to implement an interrupted version of FMCW, also known as FMICW. This technique has been proposed and used for point target seekers, but is seldom used in imaging applications because of complications introduced by discontinuities in the received signal. The operational principles are identical to those described for an FMCW system with the exception that the transmit duty-cycle is limited to about 50% at the longest operational range as shown in Figure 4.25.

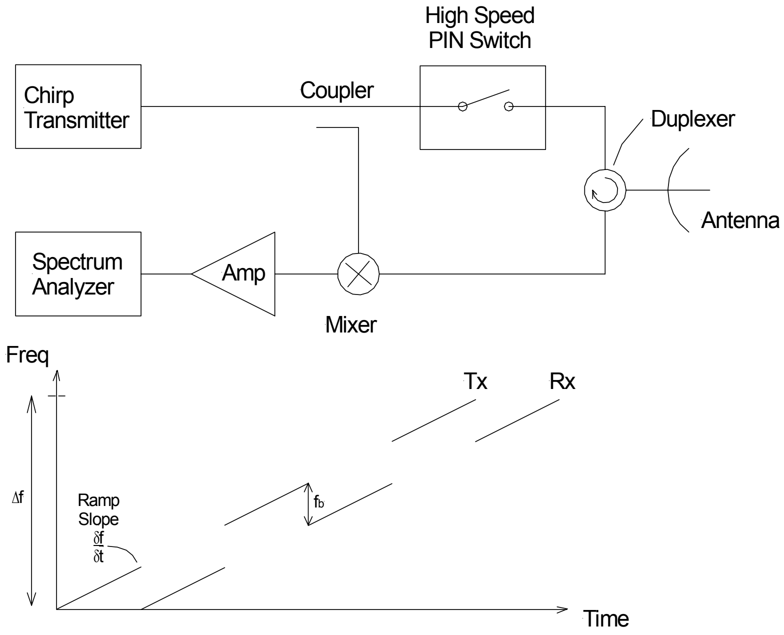


Figure 4.25: Schematic diagram illustrating the FMICW concept

Because the transmitter is turned off during the receive period there is no leakage and the receiver noise-figure is fixed so the SNR can be increased by increasing the transmitter power. However it is shown [156], [110], [168] that the effect of eclipsing results in an ambiguity function that is dependent on the range to the target as well as the relative transmit and receive periods.

For a deterministic interrupt period it has been shown [111] that the received spectrum contains a number of spurious signals that can easily obscure the real target echo. One solution is to tailor the period to obtain a spurious-free region covering the ranges of interest. As alternatives, pseudo-random interrupt periods [156] or weighted sequences [168] can be used to spread the spurious returns can be used. Both of these techniques increase the noise floor significantly.

When compared to the 10mW FMCW system discussed in the previous section, the increase in the transmitter power to 250mW, the decrease in the noise figure and the slight decrease in required resolution just balances the increased range requirement in clear air as shown in Table 4.3.

Table 4.3: Received power balance comparison between an FMCW radar at 500m and that of an FMICW radar at 3000m indicating equivalent performance requirements

Contributor	Short Range	Long Range	Signal
Tx Power	10dBm	24dBm	+14dB
Noise Figure	22dB	10dB	+12dB
RCS	-8dBm2	-3dBm2	+5dB
Range	500m	3000m	-31dB
Total			0dB

It can be seen that this technique is viable under good conditions, but if the atmospheric attenuation is degraded due to rain, then the operation out to 3km becomes marginal unless the angular resolution requirement is relaxed or if the transmit power can be increased. Relaxing the range-resolution requirement is not helpful as it requires an increase in the gate bandwidth, with a consequent increase in the noise level.

An implementation which is described in detail in Chapter 5, processes each of the received periods independently in a manner which is more akin to Stretch than to conventional FMICW. This technique overcomes the problem of spurious returns and the increased noise floor, but because the observed chirp bandwidth is decreased, the range resolution is poorer than that which can be obtained using the other techniques.

As with the FMCW technique, the matched-filter process involves mixing the received signal down to baseband where it is digitised and the range profile extracted using a standard FFT.

4.7. Conclusions

This chapter starts with a brief introduction to using the ambiguity function as a measure of radar performance from a moving platform. It shows that it is possible to produce large-bandwidth waveforms which increase range and velocity resolutions simultaneously without limit independently of transmitted the pulse width.

An analysis of a typical pulsed time-of-flight radar shows that it cannot operate out to a range of 3km and maintain the required resolution without increasing the transmitter power to 5kW. As this is impractical due to the high cost and large size of high-power transmitters, an alternative technique must be used.

Pulse compression can be employed to divorce signal bandwidth (range resolution) effects from the transmitted pulse length by modulating the signal within the pulse in amplitude, phase or frequency. This results in a transmitted pulse with a duration which is much longer than the effective compressed pulse width. Because range is proportional to the product of the transmitter power and the uncompressed pulse width, while the resolution is proportional to the compressed pulse width, there are advantages to developing a system with the longest possible pulse width.

In the extreme, a system that transmits continuously such as FMCW would appear to be the ideal candidate. However a detailed analysis conducted of this technique shows that transmit-power restrictions imposed by having to transmit and receive simultaneously limit its useful range to less than 500m for most imaging applications.

A number of other pulse-compression techniques including step frequency and phase coding are discussed and considered as candidates for the long-range application. It is concluded that any of the other CW techniques would suffer from the same shortcomings that limit FMCW performance.

The longest pulse width that can be accommodated without an overlap occurring between the transmitted and received signal is one that is equal to the round-trip time to the target. It is shown that a number of the techniques can be modified to accommodate this requirement. For range and resolution performance based criteria, these techniques are similar, but in terms of the signal processing overheads, the FMICW/Stretch technique offers the best solution. It is therefore selected as the technique of choice for the long-range imaging system that is developed in the subsequent chapters of this thesis.

Chapter 5.

Interrupted FMCW

5.1. Introduction

As discussed in Chapter 4, there are serious limitations to the operation of FMCW radars at long range unless special precautions are taken to minimise the amount of transmitter power reflected into the receiver. Unfortunately, these precautions are not feasible either in the millimetre-wave band where active leakage cancellation techniques are prohibitively expensive, or for airborne applications where the dual-antenna approach is difficult to implement successfully.

The increases in power required for long-range operation can be accommodated without significantly reducing radar performance by temporally separating the transmit and receive cycles. This separation can be accomplished using the classical short pulsed configuration, with its own associated disadvantages, or by using one of the wide bandwidth techniques such as SAW pulse-compression or FMICW.

A derivative form of the latter technique combined with classical Stretch processing was chosen as the most efficient and easiest to implement of the wide bandwidth techniques.

With a 50% duty factor, the average transmitted power for FMICW is reduced by 3dB compared to FMCW but the improved isolation improves the system noise by more than 3dB for short-range seeker applications [67] p683, [171]. Unfortunately the process of interrupting the received waveform introduces a large number of spurious spectral-components which make this process unsuitable for anything but single target applications.

In the case where multiple targets must be tracked or in an imaging application, these spurious components must be removed. Pseudorandom interrupt sequences to spread the peaks across the full spectrum can be used [156], but this results in an unacceptably large increase in the noise floor. Using an extension of the classical ambiguity function to cater for the interrupts, [110] shows that the function is dependent on the relationship between the interrupt period and the round-trip time to the target. This information is used to tailor the relationship between transmit and receive periods to produce a spurious-signal free region around the target range of interest.

Doppler processing, waveform design and optimal windowing techniques are derived [110], [190], [168] as part of the development of long-range, low-frequency oceanographic studies in which a limited range span is examined.

This chapter which forms one of the core contributions of this thesis starts by developing a “matched” filter to optimise the performance of an FMICW/Stretch imaging system given certain constraints to the linearity, chirp and processing bandwidth. This leads naturally to a dual gating scheme in which a constrained-bandwidth stepped down-conversion is followed by digitisation and the Fast Fourier Transform (FFT) to produce range gates with bandwidths matched to the range resolution.

The second half of this chapter documents practical aspects of the development of a millimetre-wave (94GHz) FMICW/Stretch system for a UAV application. It starts by dividing the hardware into the two primary radar-functions which are linear-chirp generation and signal amplification and pulsing. Both open-loop and closed-loop chirp-linearisation techniques are discussed as a justification for the use of the classical delay-line discriminator and phase-locked loop technique. The section on amplification discusses the theory and implementation of IMPATT diode based pulsed injection locked oscillators.

Section 5.4 covers all of the remaining hardware systems relevant to the radar implementation and performance including effects of switching on the performance of the IF amplifier, thermal control, linearity and noise figures.

The front-end performance evaluation is described in Section 5.5. It is determined by placing a corner reflector at a range of 2600m and measuring the return echo amplitude and shape using a spectrum analyser. The analysis of these measurements shows that signal-level calibration closes to better than 1dB but that the range resolution is worse than expected by a factor of four.

The remainder of the chapter describes the signal processing techniques used to implement the range gating function. This is an implementation of the dual process for which the theoretical foundation is derived at the start of this chapter. The process starts with a stepped down-converter which breaks the returned echo into twelve narrow band signals (coarse gates) which are processed sequentially after sampling and digitisation. The fine gating process uses a complex FFT to produce 32 overlapping range bins across each coarse gate to generate a total of 384 bins spanning the 1.5km to 3km range. The theoretical range-resolution using this process is about 5m, but it was found that the actual resolution achieved was closer to 20m for reasons explained in Appendix G.

Notwithstanding the poorer than expected performance, the images produced by the system, when the beam was scanned over 30°, were exceptionally clear. The system proved to be sufficiently sensitive to see low-reflectivity ground returns at low grazing-angles, and yet it did not suffer from saturation effects with high RCS targets that are common to FMCW systems.

5.2. FMICW Matched Filter

5.2.1. Derivation of the Matched-filter Equations

It is shown in Chapter 4 that the interruption process involves breaking the transmitted signal into an integral number of receive periods during the chirp period. To determine the effectiveness of this waveform, it is necessary to evaluate its response function by a process which is summarised in Figure 5.1.

To start with, the response of an FMCW waveform can be described by the product, in the time-domain, of an infinite chirp and a gating function with duration T_b. The transmit and receive FMICW waveforms are then obtained by multiplying the FMCW waveforms with the following transmit and receive interrupt sequences

$$\begin{array}{ll} 0 \le A_T(t) \le 1 & -\infty < t < \infty \\ 0 \le A_R(t) \le 1 & -\infty < t < \infty \\ \text{and} & \\ A_R(t) = 0 \quad \text{if} \quad A_T(t) \ne 0 & \end{array} \tag{5.1}$$

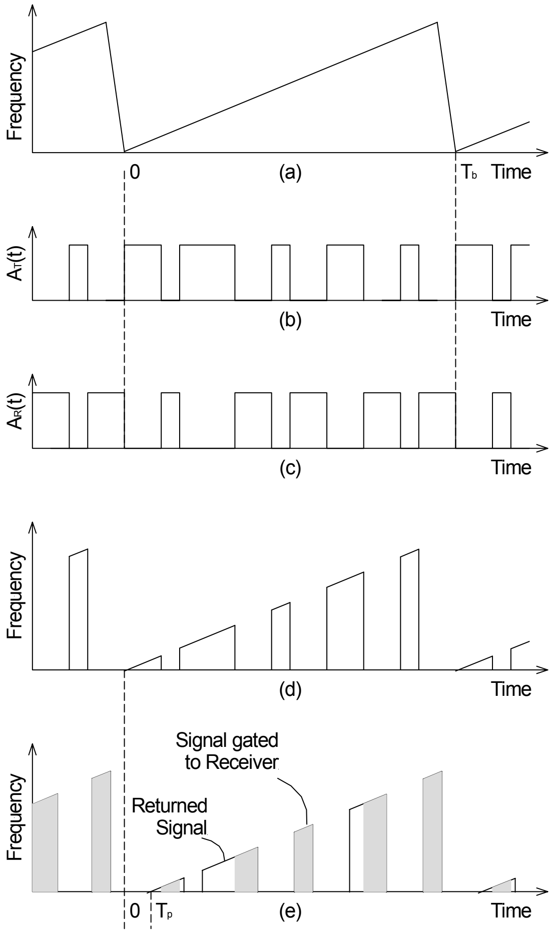


Figure 5.1: Process to determine the effect of target range on an FMICW radar signal where (a) is the FMCW signal (b) the transmitter gating sequence (c) the receiver gating sequence (d) the transmitted FMICW signal and (e) the received FMICW signal for a round-trip time of T_p sec.

The transmitted waveform is the gated version of the standard FMCW signal $v(t)$,

$$f_T(t) = A_T(t)v(t)\,. \tag{5.2}$$

The radar returns are delayed by a round-trip time T_P and further modified by the receiver gating function $A_R(t)$,

$$f_R(t) = A_R(t)A_T(t - T_P)v(t - T_P)\,. \tag{5.3}$$

This can be rewritten displaced in time by T_p seconds as follows

$$f_R(t + T_P) = A_R(t + T_P)A_T(t)v(t)\,. \tag{5.4}$$

The cross response of the receiver and transmitter gating functions can then be determined as

$$\chi_0(t_d, f_d, T_P) = \int_{-\infty}^{\infty} A_T(t + t_d/2)A_R(t + T_P - t_d/2)A_T(t - t_d/2)\exp(-j2\pi f_d t)dt\,. \tag{5.5}$$

Equation (5.5) can now be convolved in the frequency domain with the response function of the standard FMCW waveform to give the cross ambiguity function for the FMICW waveform.

Under normal circumstances, the rate of change of frequency (chirp slope) for an FMICW waveform results in a sufficiently high beat frequency that the velocity-induced Doppler shift of most targets can be ignored. This allows the zero-Doppler ambiguity function to be used to describe the performance of the waveform adequately.

From (4.26) it can be seen that the Doppler frequency corresponding to any delay t_d is simply $-A_b t_d/2\pi$ which allows the zero-Doppler response function to be calculated

$$\chi_0(t_d, f_d, T_P) = \int_{-\infty}^{\infty} A_T(t + t_d/2)A_R(t + T_P - t_d/2)A_T(t - t_d/2)\exp(jA_b t_d t)dt\,. \tag{5.6}$$

To simplify equation (5.6) it is rewritten replacing t with $t+t_d/2$ [110], [111],

$$\chi_0(t_d, f_d, T_P) = \int_{-\infty}^{\infty} A_T(t + t_d)A_R(t + T_P)A_T(t)\exp[jA_b t_d(t + t_d/2)]dt\,. \tag{5.7}$$

5.2.2. Evaluation of Options

It is convenient to evaluate the various FMICW options described by the response function (5.7) in terms of the power spectrum of the received signal. In the following examples, a chirp with $\Delta f = 360$MHz with a duration $T_b = 240\mu$s, ($A_b = 9.92\times10^{12}$) is broken up into six equal sections consisting of a 20µs transmit and a 20µs receive period. Two closely spaced point targets, separated by 10m and with amplitudes differing by 10dB are simulated at a range of 3km to produce the following results.

Figure 5.2 shows the continuous spectrum (a) where $A_R(t) = A_T(t) = 1$, followed by the interrupted spectrum (b).

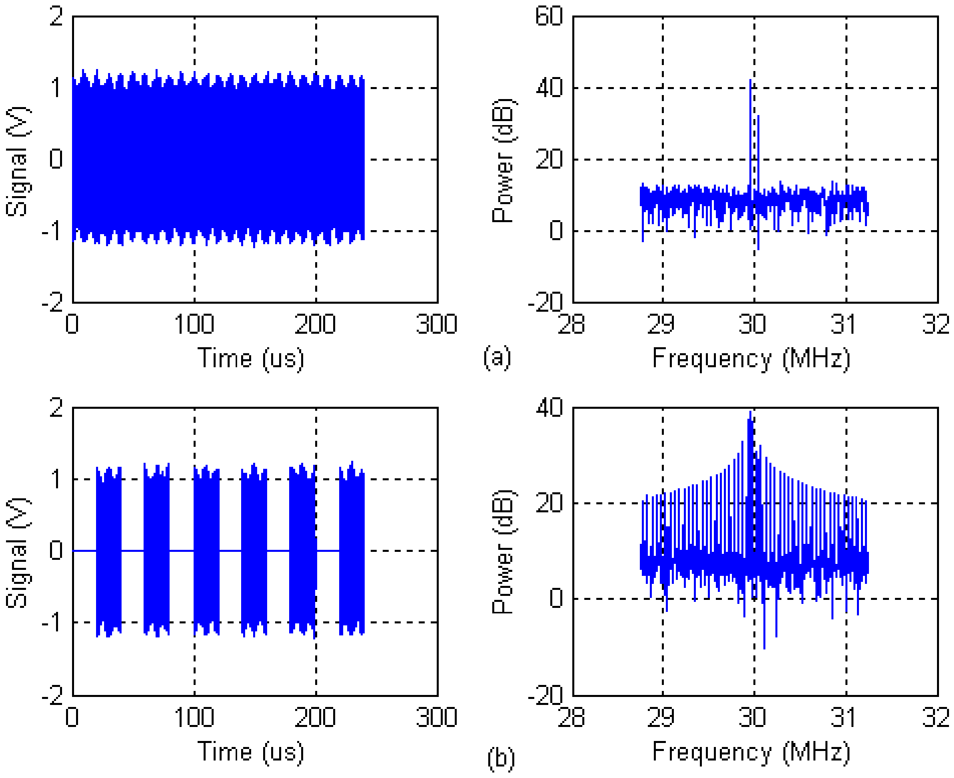


Figure 5.2: Comparison between the received signals and spectra for two closely spaced targets of different amplitudes for (a) an FMCW radar and (b) an FMICW radar with a deterministic interrupt sequence

In the FMICW case the 50% interrupt period has been matched to round-trip time to maximise the received SNR, and it is clear from the figure that the target amplitudes have been reduced by 3dB from the FMCW case as predicted. However even with a difference of only 10dB in the amplitudes of the two targets, it is difficult to identify the return from the smaller one as its amplitude is smaller than that of the surrounding spurious peaks.

In the FMICW case, the received signal can be considered to be the product of a square wave with a period of τ_{sq} = 40μs and the continuous beat frequency signal. In the frequency domain, this is equivalent to taking the convolution of the Fourier transform of each. As the Fourier transform of the square wave includes components at a frequency $f_{sq} = 1/\ \tau_{sq}$ = 25kHz and harmonics thereof, the output spectrum will include the target spectra repeated at this interval and weighted by the magnitude of each harmonic.

Figure 5.3 shows the results for a uniformly-distributed pseudorandom time-weighting of the interrupt sequence. It is assumed that the target range is sufficiently short that the roundtrip time can be ignored. In this case the sidelobe energy is spread randomly across the spectrum with an associated reduction in the individual amplitudes of the spurs when compared to the previous example. However it can be seen that the amplitude of the range sidelobes surrounding the larger target still swamps the return from the smaller one.

A practical implementation of this technique that included round-trip times to different targets would result in an effective receive duty cycle of less than 50% with a resulting decrease in the SNR.

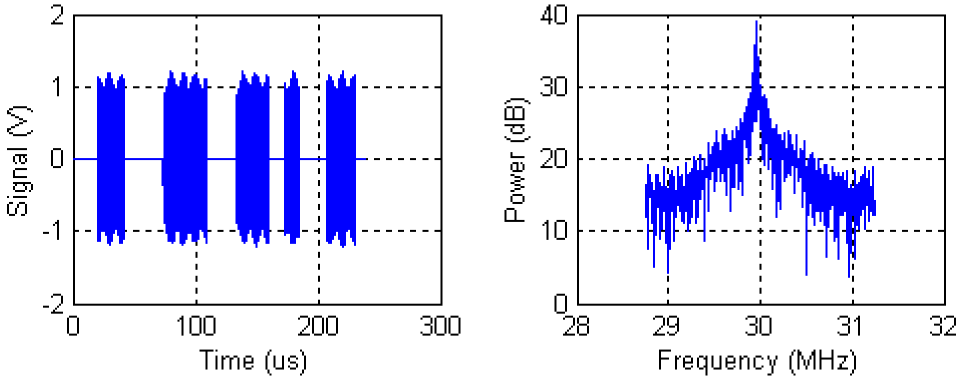


Figure 5.3: The received signals and spectra for two closely spaced targets of different amplitudes for an FMICW radar with a pseudorandom interrupt sequence

An alternative is to process each burst independently. To produce the same number of points as shown in the previous examples, the remainder of the time series is packed with zeros as shown in Figure 5.4. In this case the second target is still difficult to identify because of the range sidelobes and because the reduced chirp bandwidth of 30MHz can only just resolve the two targets. It should also be noted that the signal level for both targets is reduced by 10.8dB compared to the continuous spectrum.

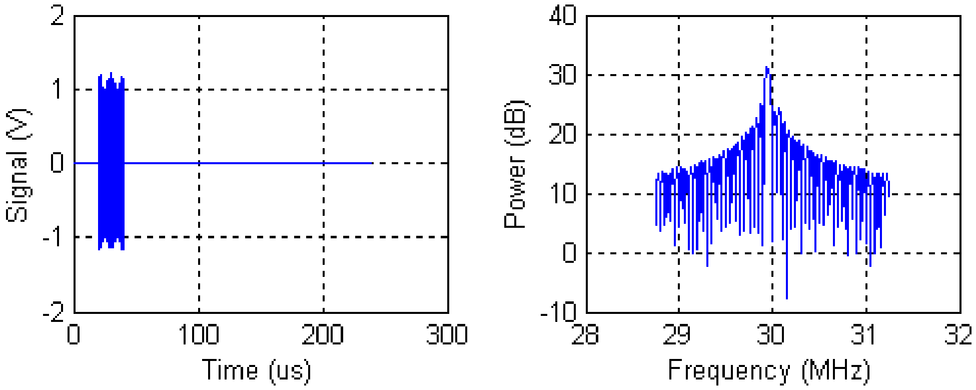
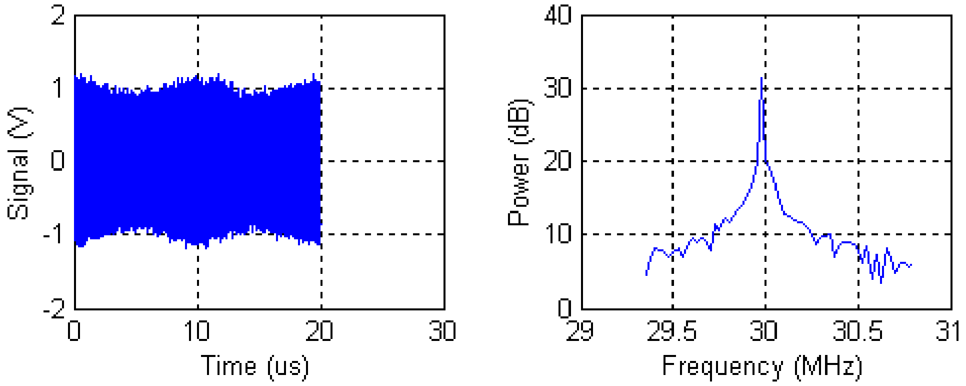
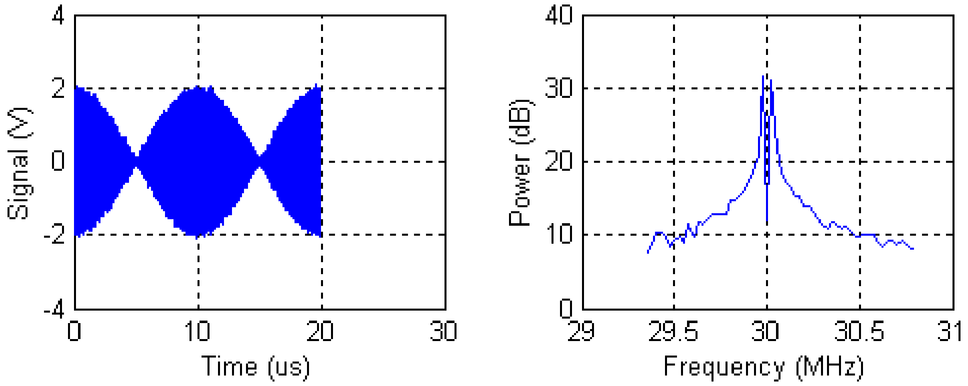


Figure 5.4: The received signals and spectra for two closely spaced targets of different amplitudes for an FMICW radar processing a single return packed with zeros

In the final option shown in Figure 5.5, the received signal is not packed with zeros, and only the 20μs during which the signal is present is processed. This produces a similar spectrum to that of the previous option with the same peak power but made up from fewer points. In this case the smaller target return cannot be identified at all in (a) because of the reduced range-resolution and high sidelobe-levels. The signal is clearly visible if the level is increased as shown in (b). With the appropriate window to reduce the sidelobe levels, the smaller target would be visible as the two peaks are sufficiently far apart.



(a)



(b)

Figure 5.5: The received signals and spectra for two closely spaced targets for an FMICW radar processing a single return only for the case in which (a) the target amplitudes differ by 10dB and (b) the target amplitudes are identical

In summary, an ideal FMCW system with a chirp bandwidth of 360MHz has a range resolution of about 0.4m. The process of interrupting this signal either with a deterministic or a pseudorandom duty cycle does not degrade this resolution significantly, though it does introduce spurious signals which make imaging impractical. If instead of processing the complete signal, each received block is processed individually, the effective chirp bandwidth is reduced to 30MHz and the range resolution is degraded to 5m. In this case the spectrum is clean and so is more suitable for building images, albeit with a reduced resolution.

5.2.3. Processing Efficiency and Integration

If the number of floating point operations required to generate the FFT is $O(n.log_2n)$, where n is the number of points in the FFT, then it is more efficient to generate a number of shorter FFTs and integrate them non-coherently than to generate a single long FFT. In this example the processing speed for a single burst is 15.79 times as fast as that for the whole sweep, so to process six bursts is still 2.63 times as fast.

The non-coherent integration improvement factor for six bursts is about 6.2dB compared to about 7.8dB for coherent integration of the complete interrupted signal, resulting in a theoretical loss in SNR of 1.2dB as a result of the improvement in processing efficiency.

A technique to further improve the range resolution of block-processed frequency-stepped chirp by deramping (Stretch) and concatenation is discussed in a recent paper [170]. This method goes beyond what has been developed for this thesis, but may be an option in the future.

In the light of this processing efficiency, and the relative ease of dealing with shorter FFTs, it was decided to implement the individual burst processing method followed by non-coherent integration of the resulting spectra as described in Section 5.8.

5.3. Implementation of an FMICW Front-End

Figure 5.6 shows the block diagram of the second iteration of FMICW radar developed for this application. It is the culmination of more than three years of development and testing.

It can be seen that the RF front-end consists of two main components. The module on the left (a) uses the delay-line discriminator method to produce a linear chirp, while the module on the right (b) gates and amplifies the signal to produce the interrupted transmissions.

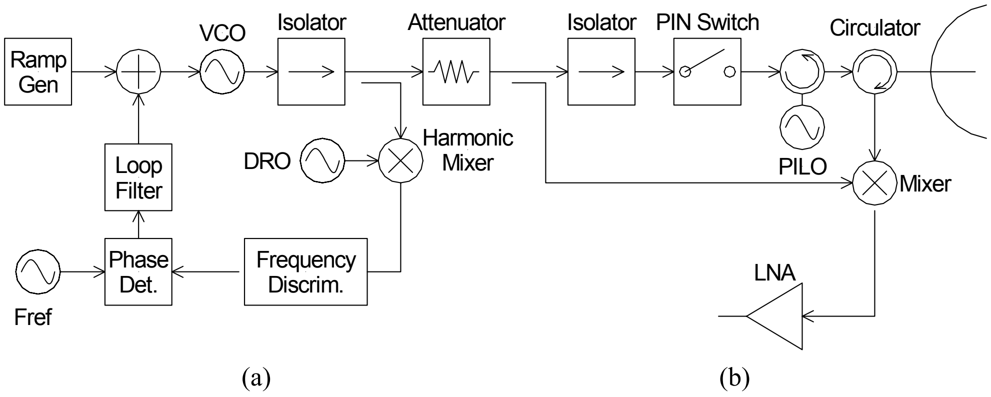


Figure 5.6: Schematic block diagram of an FMICW radar system showing that it consists of two major components (a) a linear ramp generator and (b) transmit signal gating and amplification

5.3.1. Open-loop Linearisation

It was shown in Chapter 4 that a non-linearised FMCW or FMICW radar would produce poor images because of the poor range-resolution that is achieved at long range. To circumvent this problem, a number of practical methods of linearising the chirp signal have been considered.

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 is controlled, or a number of curves is implemented and referenced appropriately. The latter is easily achieved, as shown in Figure 5.7 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.

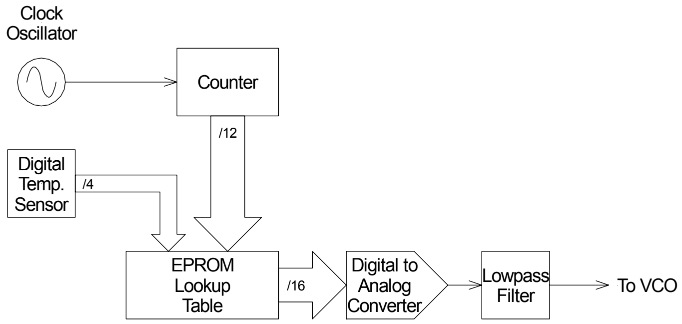


Figure 5.7: Schematic diagram of a linear chirp generator based on a lookup table

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 which can reduce the nonlinearity of a well-behaved VCO by a factor of 10 is shown in Figure 5.8. This alternative was implemented successfully for this application. It 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.

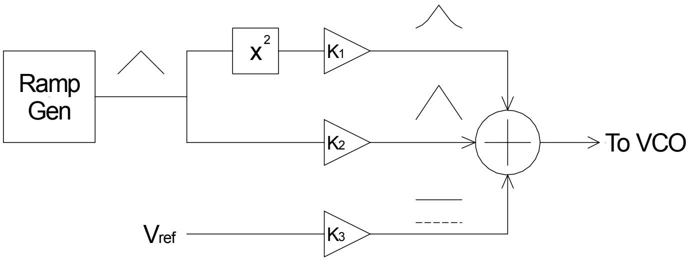


Figure 5.8: Quadratic frequency chirp correction circuit using an analog multiplier chip

5.3.2. Determining the Effectiveness of Linearisation Techniques

The obvious method to determine the effectiveness of a linearization 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 an 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 most applications SAW or bulk acoustic wave (BAW) devices are used to fulfil this function as shown in Figure 5.9.

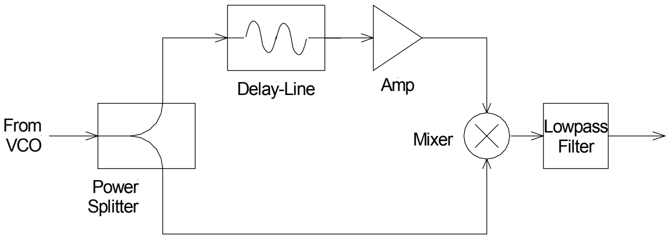


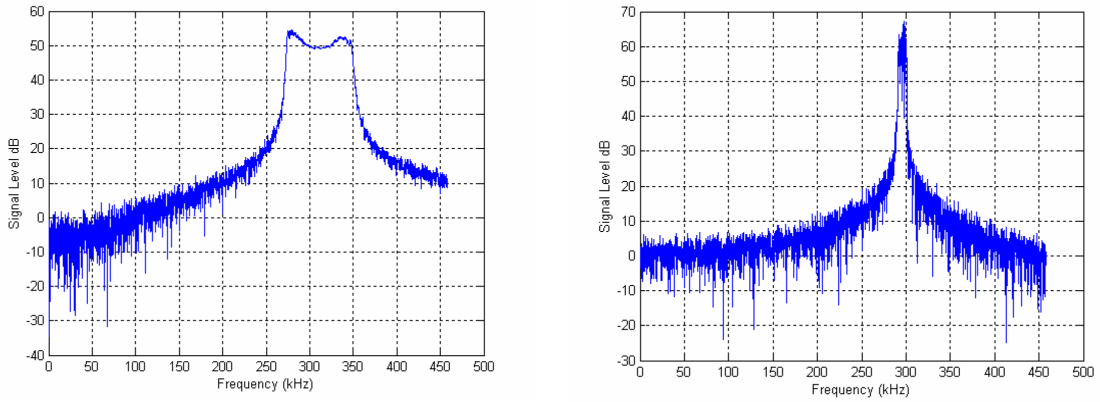
Figure 5.9: Schematic diagram of a delay-line discriminator

Disadvantages of commercially available 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 (typically 700MHz) to take advantage of commercially available delay-lines.

If, as is often the case in high-resolution applications, a chirp bandwidth in excess of 300MHz must be monitored, then the VCO output is down-converted to a higher IF (for example 1.4GHz) and then divided by two using a high-speed digital divider (prescaler) to produce a signal at the appropriate frequency with a reduced bandwidth.

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 Figure 5.10, the discriminator outputs are shown for the Hughes VCO (discussed in Appendix E) both completely unlinearised (a) and after open-loop linearization (b).

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 5.10: Discriminator output spectra for (a) an unlinearised Hughes VCO and (b) after open-loop correction

Unfortunately, many of the millimetre-wave voltage controlled oscillators that are available do not have well-behaved characteristics. A VCO produced by Millitech that has been incorporated into another radar shows characteristics (in Figure 5.11) which are not easily linearised by a simple quadratic correction.

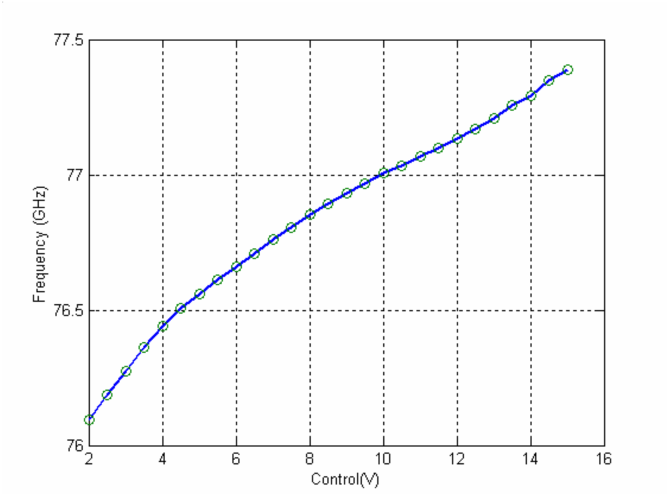


Figure 5.11: Example of the non-linear characteristics exhibited by a typical Gunn VCO

For applications where the required chirp-bandwidth is small, it is still possible to select a region and apply the classical quadratic-correction effectively. However if a bandwidth of more than 200MHz is required, then only the digital look-up table can be used. Even using this technique, the smallest discriminator bandwidth that could be achieved in experiments was 40kHz which corresponds to a linearity of 0.11. This is poor compared to that which was obtained using the better-behaved Hughes VCO with quadratic correction.

5.3.3. 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 configuration [89] p9, if it is remembered that the loop must maintain a constant rate-of-change of frequency and not a constant frequency as is more usual.

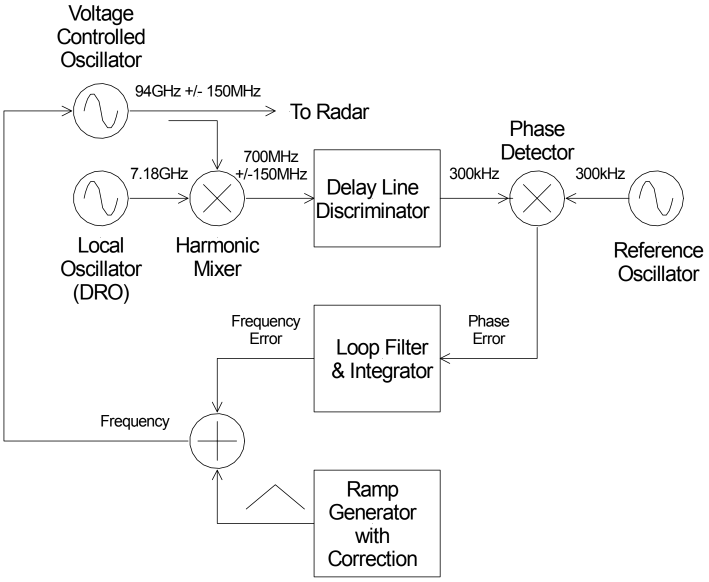


Figure 5.12: Schematic diagram showing the process of chirp linearisation based on a combination of open-loop correction and closed-loop delay-line discriminator output feedback

The delay-line discriminator is effectively a differentiator in the frequency domain and produces a constant output frequency if the chirp-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 Figure 5.12.

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 generated by this process from the receiver spectrum completely.

5.3.4. Amplifying and Pulsing

Injection locking is often used to improve phase and frequency stability of a source. In this case it is used to lock the phase and frequency of a high power IMPATT source to the low power linear ramp generated by a Gunn Oscillator. More information on this topic can be found in Appendix G.

The locking bandwidth, $\delta\omega_L$, can be derived for a small injection signal as

$$\delta\omega_L = \frac{2\omega_0}{Q_e}\sqrt{\frac{P_i}{P_o}}\frac{1}{\cos\theta}, \tag{5.8}$$

where ω_o – Free running frequency,
Q_e – Cavity Q ($\omega_0 L/R_L$),
θ – Phase angle of Z*(A),
P_i – Injection signal power,
P_o – Output power.

If Q_e is large then $\delta\omega_L$ is very small, whereas if Q_e is low, oscillator behaviour, stability and noise performance are degraded. Generally a compromise is required.

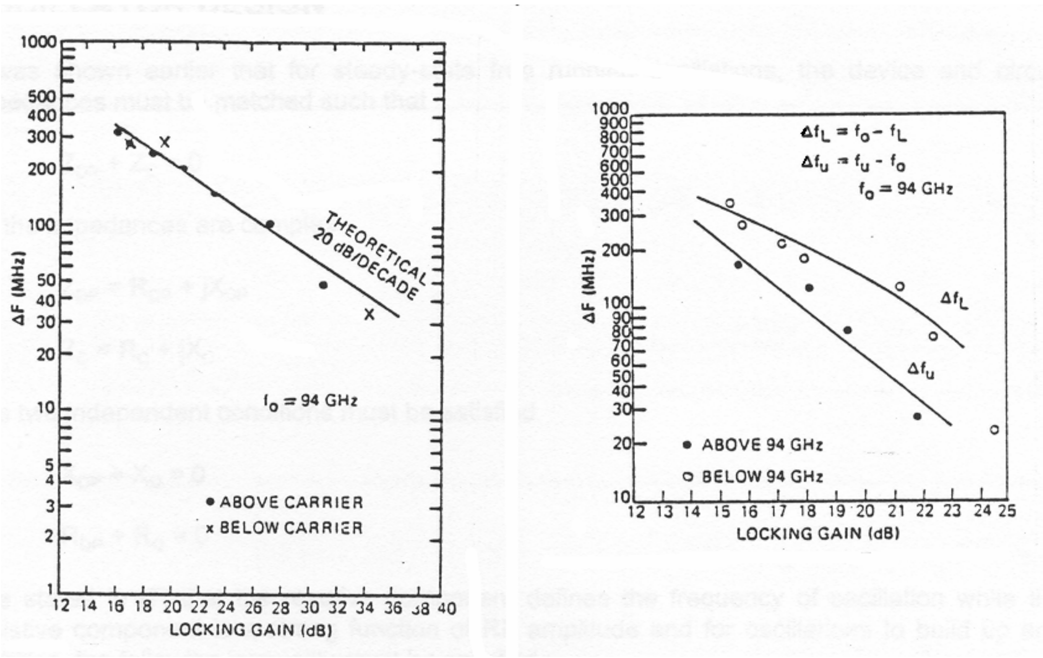
The locking bandwidth is kept to below 1GHz for ILOs to maximise output power. It is can be shown [122] that ILOs follow the classical gain-bandwidth relationship,

$$\sqrt{G} \times \delta\omega_L = \text{constant}. \tag{5.9}$$

Figure 5.13 shows this relationship directly.

Pulsing (or interruption) is accomplished by removing the signal source from the IMPATT oscillator using an electronic (PIN) switch and by turning off the drive current that supplies the oscillator with power. This type of oscillator is known as a pulsed injection locked oscillator (PILO).

The total attenuation of the PIN-switch used to isolate the VCO signal is 20dB and the leakage through the PILO is measured to be -10dB which reduces the magnitude of the signal leakage by about 30dB from that in a conventional FMCW radar. The effect of this reduction on the overall noise figure is discussed in Section 5.4.5.



(a) CW ILO (b) Long Pulse ILO

Figure 5.13: Locking gain-bandwidth characteristics of millimetre-wave IMPATT injection-locked oscillators[122]

To maximise the return SNR and the range resolution, the burst period is made equal to the round-trip time to the longest operational range plus a small margin for switching. In this case the range is 3km, and 1µs is allowed for switching, making the total transmit period 21µs. The receive time is set to the same duration so that the full echo can be received. The chirp period is selected to be equal to an even integer multiple of the transmit period, which is in this case equal to four, making the total chirp duration 84µs as shown in Figure 5.14.

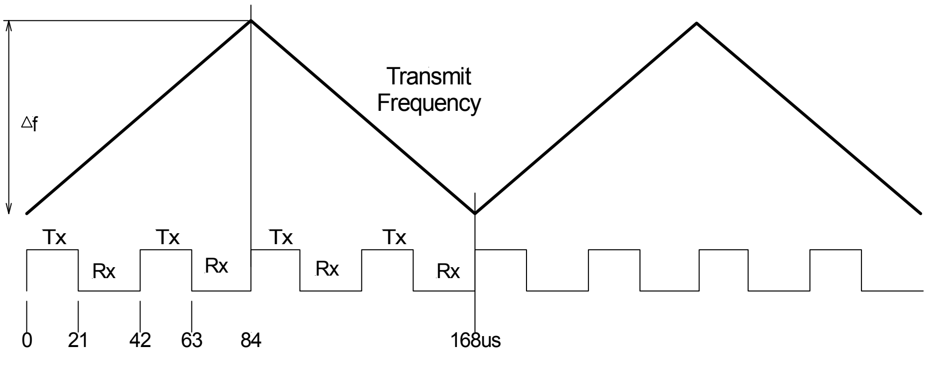


Figure 5.14: FMICW timing optimised for a 3km operational range with two interrupt cycles per ramp period

The primary advantage of increasing the total chirp bandwidth and accommodating more bursts than those shown here is because the wider bandwidth decorrelates the ground clutter returns better [17] p86. For example, for a bin size, L = 5m, the clutter correlation frequency will be

$$f_c = \frac{c}{2L} = 30\,\text{MHz}. \tag{5.10}$$

The effective number of independent samples, n_e, that can be accommodated due to this frequency diversity in a total chirp bandwidth, Δf, is

$$n_e = 1 + \frac{\Delta f}{f_c}. \tag{5.11}$$

As n_e increases, the variance of the clutter distribution decreases and the resultant image appears to be more homogenous and less speckled.

In this application it was not possible to linearise a larger swept band and so only two cycles per sweep were implemented. The resultant images showed a large amount of speckle that could only be reduced to acceptable levels by using spatial integration.

5.3.5. Effective Range Resolution

As the range to the target decreases, the round-trip time is decreased. Transmitter eclipsing of the received signal at ranges below 3km, shown in Figure 5.15, decreases the available swept bandwidth which leads to a degradation of the range resolution. However, because of the simultaneous improvement in the cross-range resolution (due to the shorter range and fixed antenna-beamwidth), the effective pixel area, and hence resolving capability, remains constant [143].

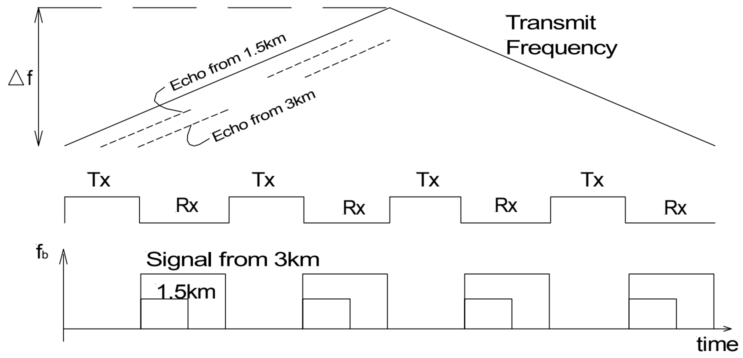


Figure 5.15: Diagram showing the received waveforms at 1.5 and 3km and the interrupt timing for the implemented FMICW radar

5.4. Radar Hardware Considerations

The front-end was manufactured using "brass block" techniques. This involves assembling by bolting together, or connecting using short sections of waveguide, all of the millimetre-wave components. This cannot occur before a complete RF level and frequency analysis has occurred to ensure that each of the modules is fed with the appropriate level signals. Figure 5.16 shows the initial loss budget for the RF bock, these include the individual component losses (measured at 94GHz) and the calculated signal levels at each flange.

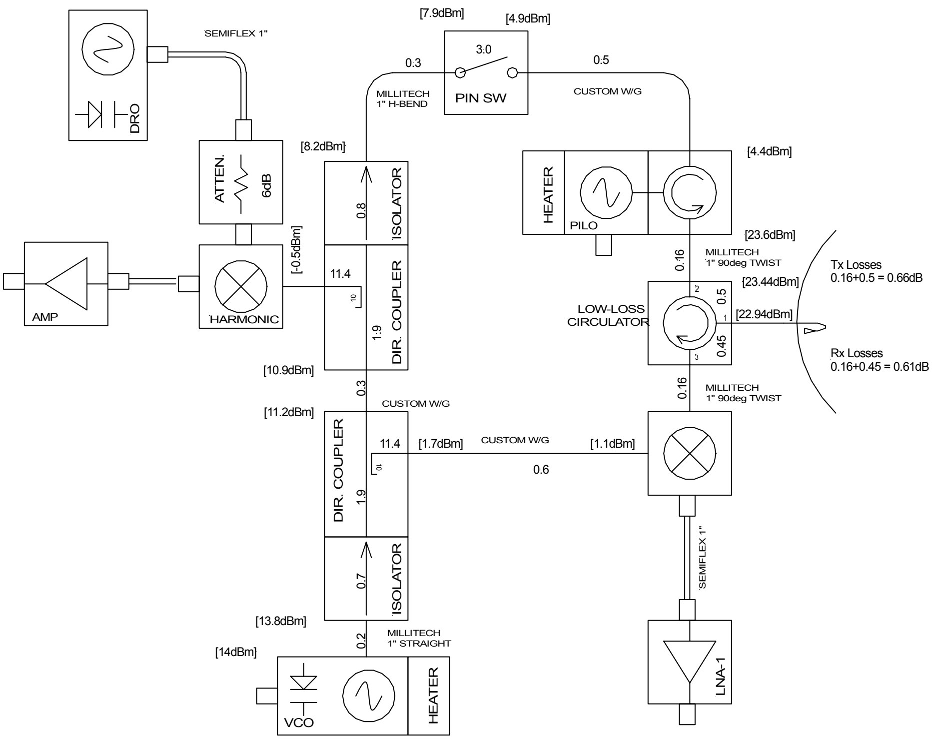


Figure 5.16: Block diagram showing the RF levels and loss budget for the 94GHz FMICW radar

It was found that the PILO performance was very sensitive to drive level (see Section 5.5.6), and a level-set attenuator was inserted in the transmitter path between the two directional couplers to allow for some adjustment.

The layout design was then undertaken with the primary objectives of producing a compact unit and of minimising the lengths of critical waveguide runs from the PILO to the antenna, and the antenna to the mixer in an attempt both to maximise the radiated power and to minimise the receiver noise figure. The final layout is shown in the photograph of the completed front-end in Figure 5.17. A low insertion-loss circulator with good isolation was also manufactured to help with this process.

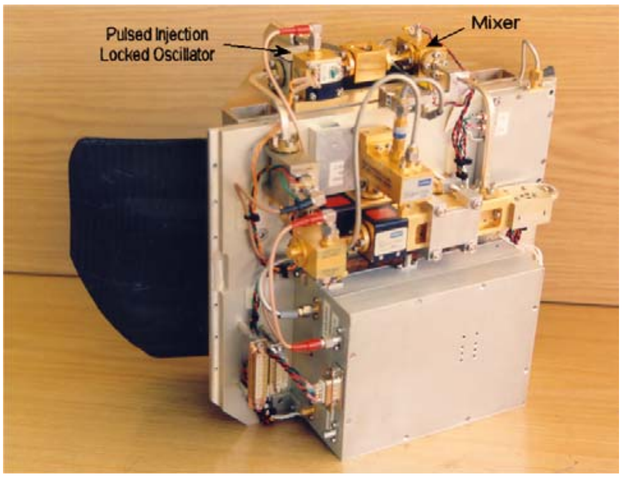


Figure 5.17: Photograph of the FMICW front-end showing the brass-block construction and indicating the positions of the critical components that are mounted close to the antenna

5.4.1. Mitigation of Switching Effects

One of the problems with an FMICW system is that both the PIN-switch noise and transmitter leakage inject large signals into the input of the LNA causing saturation and ringing. Recovery from these effects takes some time which results in poor receiver sensitivity for the period just after the transition from transmit to receive.

One method of guarding against this problem is to clamp the input and gate the LNA using a TR switch shown in Figure 5.18 so that the switching transients do not propagate as far as the digitiser circuitry.

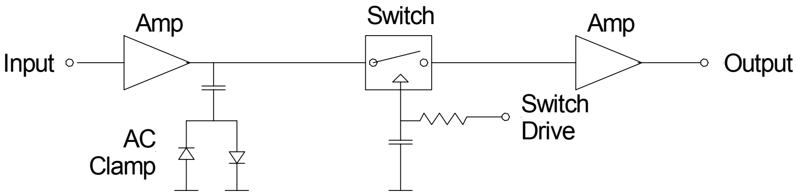


Figure 5.18: Schematic diagram of the switched LNA showing the positions of the clamp and FET switch designed to minimise leakage-transient effects

This circuitry was unable to completely eliminate the switching transients although their magnitude was reduced to a reasonable level as shown in Figure 5.19. This ensured that saturation did not occur which made recovery almost instantaneous.

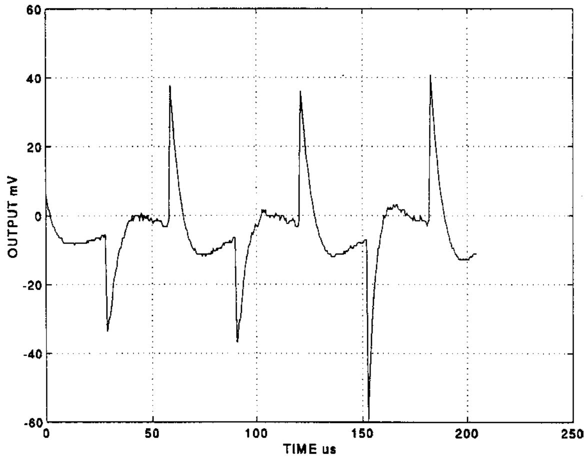


Figure 5.19: LNA output signal showing the residual switching transients that introduce wideband noise into the received signal spectra

5.4.2. Practical Aspects of Linearisation

The technique used to linearise the sweep has been documented in detail in Section 5.5.3 and Appendix E. The discriminator output spectrum is shown in Figure 5.20 in which a comparison is made between the open-loop and closed-loop linearised output spectra. It can be seen from the magnitude and width of the peak that this function was reasonably successful.

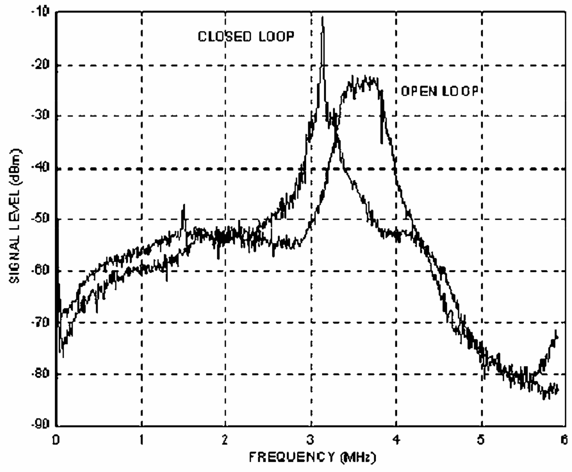
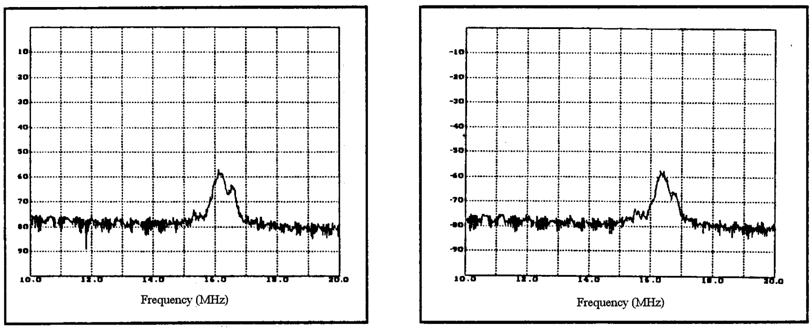


Figure 5.20: Delay-line discriminator output spectrum showing the effect of closing the linearization loop. Chirp linearity is proportional to the width of the spectral peak.

In this development, a number of different phase comparators were evaluated to see which one performed the best. They included the following:

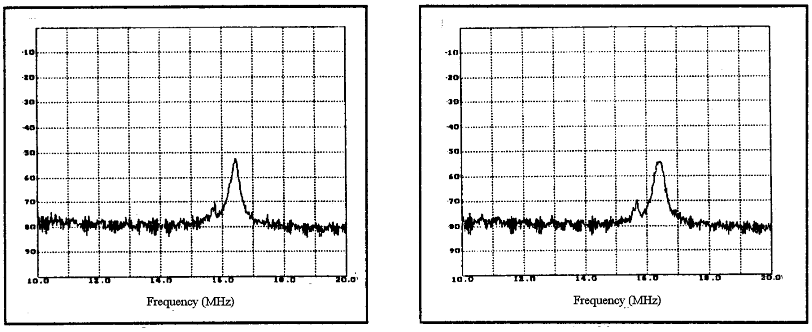
- Exclusive OR
- Counter/DAC [7 bit]
- Limited Counter/DAC [2 bit]

Figure 5.21 shows the results obtained using a building at 2.6km whose primary reflectors are two point-targets separated in range by about 50m.



(a) Unlocked

(b) Exclusive Or



(c) Counter DAC

(d) Limited Counter DAC

Figure 5.21: Phase detector performance comparison showing the processed return from two point-targets separated by 50m at a range of 2.6km. It can be seen that only the counter DAC techniques (c) and (d) adequately reproduce the smaller target at the shorter range

The counter-DAC method performed best in general, but if phase-slip occurred it took some time to recover as it had to integrate down. The compromise option which is a 2 bit counter-DAC was selected as being more robust and superior to the exclusive-OR method which produced surprisingly poor results in this application.

The measured 3dB bandwidth of the linearization signal centred at 3.15MHz is 10.7kHz. This is a linearity of 0.34% which is worse than the 0.1% goal. However, it is not possible to measure this parameter accurately because the sweep duration is too short to obtain the required frequency resolution.

5.4.3. Automatic Frequency Control

The technique of implementing a long time-constant frequency-control loop to keep the UHF output centred on 700MHz was implemented for this front-end. A frequency discriminator centred at 700MHz was constructed using a short length of coax cable as a delay line. The output was then filtered, amplified and added to the VCO ramp input as a correction. The effectiveness of the AFC loop is shown by Figure 5.22 which shows the magnitude of the frequency drift with and without control.

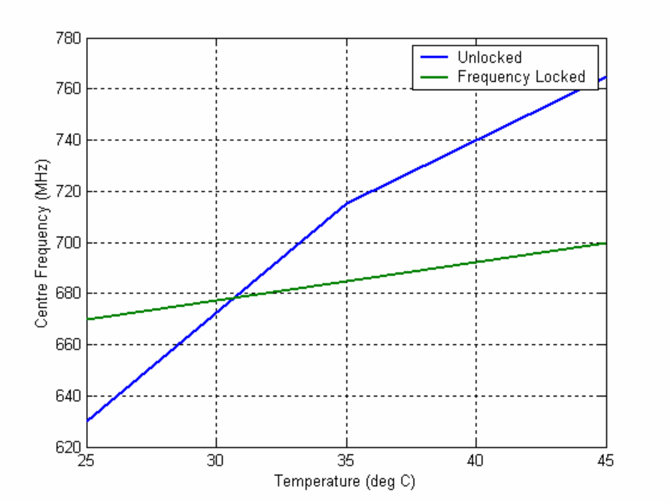


Figure 5.22: Downconverted VCO centre frequency showing the effectiveness of the frequency control loop in counteracting temperature induced frequency shift

5.4.4. Transmit Power and Spectrum

Injection locking an IMPATT oscillator involves introducing a small signal of the required frequency into the tuned cavity containing the diode to force the oscillator phase to match that of the injected signal, as discussed earlier in this chapter and in Appendix G. The amount of gain that can be achieved depends on the tuning range that is required, and as the output power remains fairly constant, the injected power level determines this.

Because this PILO is really an oscillator and not a true amplifier, it is gated by switching the bias current. An adjustable precision current source coupled to a high speed FET switch was developed to perform this function [34].

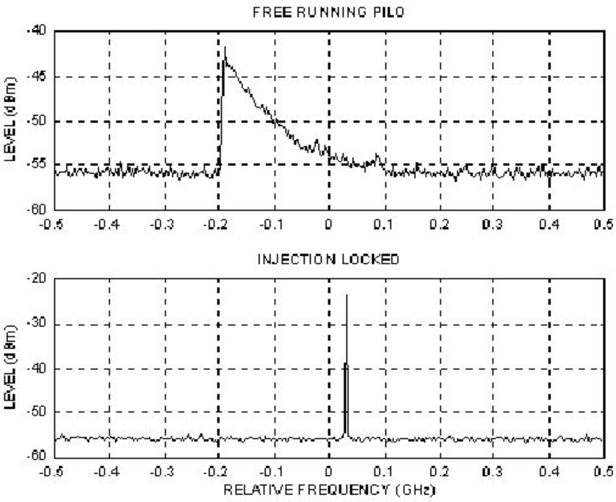


Figure 5.23: Relative spectra centred at 94GHz showing (a) the temperature induced frequency chirp in an unlocked PILO and (b) the effect of injection locking this spectrum to a constant frequency source.

The upper graph in Figure 5.23 shows the spectrum of an unlocked PILO. It can be seen that as the diode heats with every pulse, the oscillation frequency changes by about 300MHz. The non-linear power density is a reflection of the non-linear nature of the chirp. The lower graph shows the spectrum of the PILO locked to a fixed frequency.

The transmitted power is measured at the antenna flange using the test set-up shown in Figure 5.24.

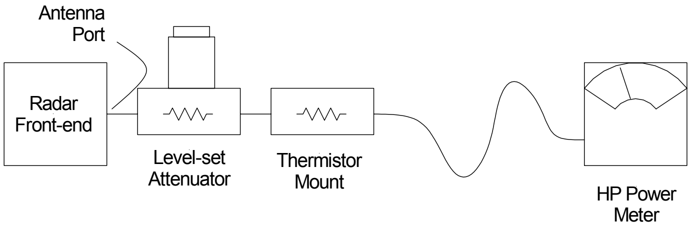


Figure 5.24: Schematic diagram showing the set-up used to measure the transmit power

The actual transmitted power can be determined to an accuracy of better than 0.5dB if all of the path losses are considered

$$P_{tx} = P_m + L_{met} + L_{attn} + L_{dut} = 23.3\text{dBm}\,, \tag{5.22}$$

where P_{tx} – Transmitted peak power (dBm),
P_m – Measured power (0.8dBm),
L_{met} – Power meter calibration (1.9dB),
L_{attn} – Attenuation through system (17.2dB),
L_{dut} – Duty cycle, $10\log_{10}(28.1/62) = 3.44$dB.

It can be assumed that the power meter calibration is exact.

This measured power is slightly lower than the maximum possible as the PILO is set up to run conservatively with an output power of 24dBm. There are also waveguide and circulator losses of 0.16 + 0.5 = 0.66dB between the transmitter and the antenna.

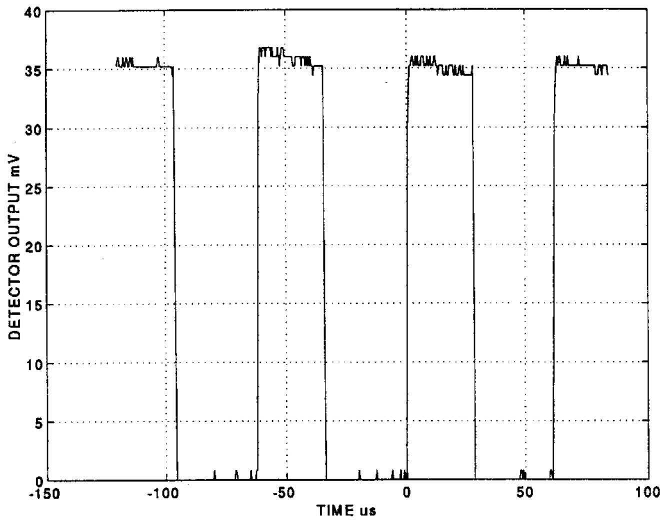


Figure 5.25: Detected output showing the four transmitted pulses with almost constant amplitude across each pulse

A detector replaced the power meter to determine whether the transmit power remained constant during the on-time. This measurement is shown in Figure 5.25 from which it can be seen that a very slight droop (<0.5mW) occurs across the pulse.

There are four individual Tx pulses per up and down ramp. Because the duty cycle is just less than 50%, these appear as individual spectra as shown in Figure 5.26. The loss due to harmonic down-conversion is not accounted for, and so the actual power measured by the spectrum analyser is not indicative of the actual transmitted power.

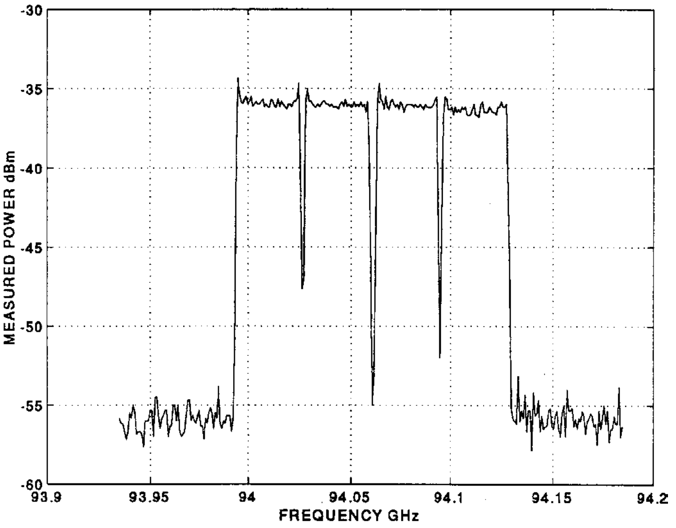


Figure 5.26: Transmitter spectrum showing the four distinct pulses visible because the duty cycle is slightly less than 50%

5.4.5. Noise Power Measurements

The radar noise performance is determined by the Y-factor method [32] using a wideband millimetre-wave noise source with a known excess noise ratio (ENR). It is found to be about 12dB.

The theoretical noise figure for the system is determined with the help of Figure 5.27.

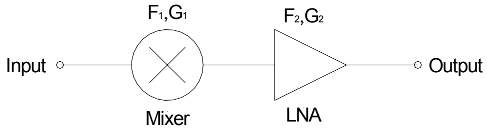


Figure 5.27: Receiver chain schematic used to determine the noise figure

It can be shown that the receiver noise-figure is determined using the formula for a cascaded network,

$$F_0 = F_1 + \frac{F_2 - 1}{G_1} = 10.15dB\,, \tag{5.23}$$

where F_1 – Mixer noise-figure (8dB or 6.31),
F_2 – LNA noise-figure (2.5dB or 1.778),
G_1 – Mixer gain (-7.15dB or 0.192).

To the thermal noise should be added the phase-noise due to leakage of the oscillator back into the receiver. The SSB phase-noise power density due to leakage is derived in Chapter 4. To obtain the noise power output at each gate, the power density must be integrated over the gate bandwidth of $\beta \approx$ 50kHz in this case. The oscillator power remains unchanged at 15dBm, and the leakage level determined when the PIN-switch is open and the PILO is off will be $L_{leak} = L_{pin} + L_{pilo} + L_{circ} = 20 + 8 + 20 = 48$dB.

The noise power level will change by 15 – 48 + $10\log_{10}(50)$ = -16dB to account for the new leakage-level and the wider bandwidth compared to the FMCW case illustrated in Figure 4.19. By extrapolation to a beat frequency between 15 and 30MHz, and taking into consideration the increase in the thermal-noise level due to the wider gate bandwidth, the phase-noise level will be below -180dBm. This is some 60dB below the thermal-noise level and so can be ignored.

The AM noise, as determined in Appendix F, is also reduced by the additional attenuation around the leakage path to a level below that of the thermal noise, so it can also be ignored.

This leaves only the thermal noise contribution and the receiver noise-figure to determine the minimum detectable signal level for the system.

The difference of 2dB between the measured and calculated noise-figures is probably due to a problem with the measurement technique or instrument calibration. The former is most likely as it is notoriously difficult to measure pulsed signal levels accurately using a spectrum analyser.

5.5. System Measurements

With the integrated front-end shown in Figure 5.17 mounted on a tripod, measurements were made at a range of about 2500m using calibrated corner reflectors with RCS values of 100 and 1000 m^2 at 94GHz [33]. The reflectors were mounted in turn on a pulley mechanism that would allow them to be raised up a 6m high pole as shown in Figure 5.28. The radar was aimed at the target and measurements were made of the received signal amplitude using a Hewlett Packard 8562A spectrum analyser.

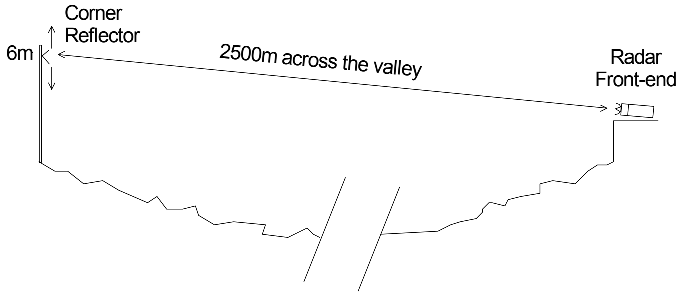


Figure 5.28: Measurement set-up to investigate radar performance. The corner-reflector can be moved to test for multipath induced effects

5.5.1. Antenna Gain

For an imaging radar it is important to maintain a constant return echo amplitude irrespective of range. It is shown in many radar texts that this can be achieved, for a constant RCS target, by shaping the gain as a function of the square of the cosecant of the elevation angle. If the radar is used to produce images of the ground, then the RCS changes with range because the beam diverges and the gate size increases. In this case the gain function should theoretically be proportional to cosec^{2}cos$^{1/2}$, but in practice this additional correction is term seldom implemented.

For a cosec^2 function starting at the horizon, the reduction in peak antenna-gain is only 3dB. However, if the function starts at a steeper grazing angle, to accommodate the required operational range, then the loss in gain is significantly larger.

Antenna polar plots made at an outdoor range [76] are reproduced in Figure 5.29 and Figure 5.30. These show that the antenna elevation pattern conforms reasonably well to the cosec^2 function required though it does roll off at larger depression angles. The azimuth pattern confirms that the antenna beamwidth is close to 0.5° and the sidelobe levels are lower than -18dB.

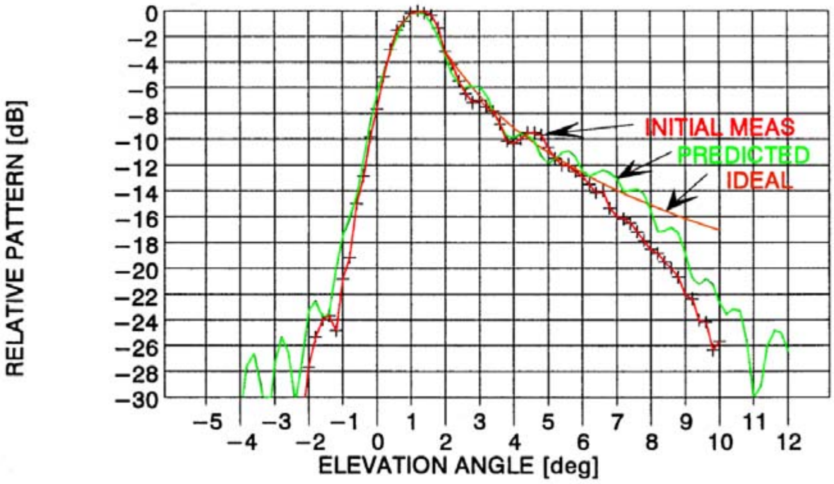


Figure 5.29: Normalised elevation polar plot showing the predicted pattern, the measured pattern before environmental tests and the measured pattern after the environmental tests

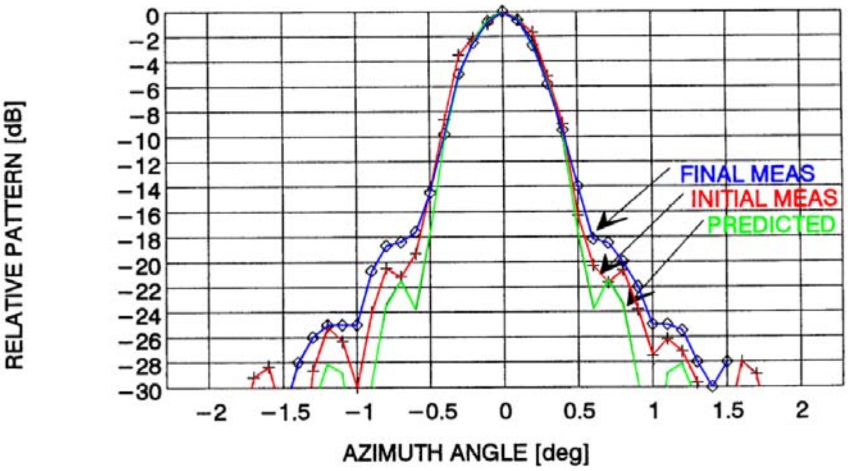


Figure 5.30: Normalised azimuth polar plot showing the predicted pattern, the measured pattern before environmental tests and the measured pattern after the environmental tests

The peak gain measured for the antenna (using a calibrated standard-gain horn as a reference) is 44.2dBi taken as an average over a number of measurements.

5.5.2. Multipath Effects

Using the configuration shown in Figure 5.28, the reflector was moved up and down the pole to determine whether there was a multipath effect. If the effects were present they would manifest themselves as strong variations in the returned echo level similar to those generated in the “flat earth” model shown in Figure 5.31.

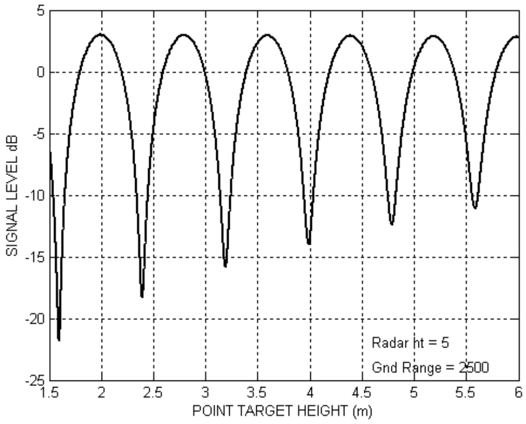


Figure 5.31: Multipath-induced amplitude lobing at a range of 2500m for a flat-earth model

In this case, a combination of surface roughness and the valley shape appears to eliminate any multipath effects as shown in Figure 5.32 where no lobing is present.

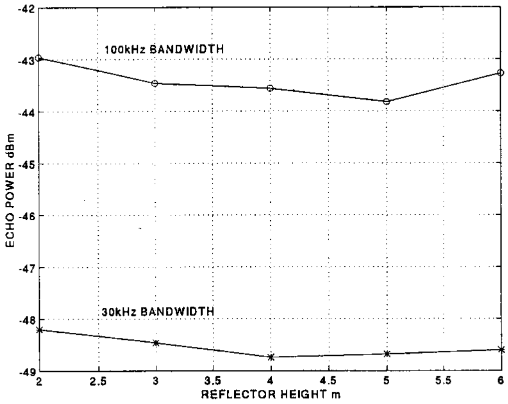


Figure 5.32: Measured echo-signal power as a function of reflector height above the ground shows a small variation in signal level with reflector height, but no lobing

The residual droop of less than 1dB towards the centre of the traverse may have been due to a number of phenomena, from variation in transmit power to a change in the corner reflector RCS with alignment. However, the most important point is that the assumption can be made that these results are largely free of multipath cancellation (or enhancement) and as such can be used to close the radar calibration.

5.5.3. Bandwidth Effects

If the radar output is measured with a spectrum analyser it should appear as repeated spectra due to the sampled nature of the transmit signal. The bandwidth of each component should equal the reciprocal of the 20µs "pulse" width to be about 50kHz wide. However various other phenomena within the radar have the effect of spreading the effective received signal more. These include the following:

- Residual chirp non linearity
- Limited sweep bandwidth
- Phase-noise
- Injection locking effects

These effects manifest themselves as changes in the measured signal level as a function of the spectrum analyser resolution-bandwidth setting as shown in Figure 5.33.

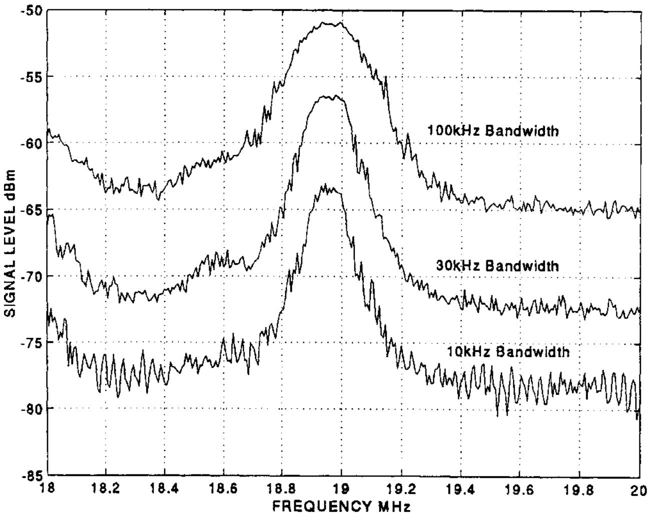


Figure 5.33: Measured echo-signal power spectrum of the FMICW radar for a 100m^2 corner reflector as a function of the spectrum analyser bandwidth

This measurement shows that the received signal is spread fairly uniformly over a 120kHz bandwidth by a combination of the effects listed above. As yet, the model that has been developed to describe the radar does not fully explain this phenomenon, where the 10dB bandwidth of 130kHz is predicted at 2000m.

With regard to closing the calibration, the radar range equation assumes matched filtering and hence the signal level should be calculated over a 120kHz bandwidth. Any frequency mismatch manifests its self as a loss in SNR when compared to the matched filter performance.

5.5.4. Noise Effects

Measurements made with a 100kHz bandwidth for 1000m^2 and 100m^2 reflectors should be different by 10dB. However, it was found that the difference was only 8dB. This is because the peak signal-to-noise value is measured as shown by the diagram in Figure 5.34.

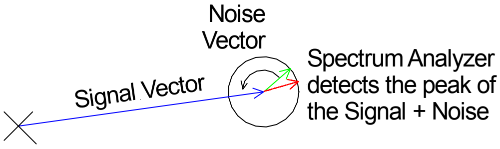


Figure 5.34: Vector addition of signal and noise

The peak value of the noise floor is measured to be -64dBm and the 100m^2 reflector power ($S + N$) is measured to be -50dBm. The true signal level is

$$S = 20\log_{10}\left(10^{-50/20} - 10^{-64/20}\right) = -52\text{dBm}. \tag{5.24}$$

In the same way, the true value of the return power level from the 1000m^2 reflector is calculated to be -42.7dBm compared to the -42.1dB measured value.

5.5.5. Closing the Calibration

Closing the calibration involves verifying that the measured and modelled signal return levels are the same, and by doing so, confirming that the radar system is operating to specification.

To simplify the calculation, the following parameters have been assumed:

- The radar bandwidth is matched to 100kHz.
- The large reflector RCS is exactly 1000m^2 (30dBm^2)
- The peak signal obtained is -42.1dBm.

The system parameters are measured as follows:

P_t	23.3dBm	Tx power measured at the antenna
G_{ant}	44.17dBi	Measured wrt standard gain horn
σ	30dBm^2	Calibrated reflector
R	2527m	Sweep Linear 18.95MHz
G_{amp}	45.33dB	Measured using spectrum analyser
λ	3.19mm	94GHz
L_{rx}	0.61dB	Measured 0.16 + 0.45dB
L_{mix}	7.15dB	Specified by Alpha at 20MHz
α	0.2dB/km	Clear air attenuation dB/km

Applying the radar range equation

$$P_r = P_t + 10\log_{10}\left(\frac{\lambda^2}{(4\pi)^3}\right) + 2G_{ant} + \sigma + G_{amp} - 40\log_{10} R - 2\alpha R_{km} - L_{rx} - L_{mix} \quad (5.25)$$

= -40.81dBm.

The calibration closes to -42.1 - (-40.81) = -1.29dB.

5.5.6. Range Resolution

The range resolution is defined as the separation required between two targets of equal amplitude so that they are just resolved. For an FMCW radar this is generally taken to be the 3dB bandwidth of the received signal. For this radar the theoretical 3dB bandwidth is 77kHz.

Tests conducted using the radar in FMCW mode, with the PILO bridged out, resulted in a 70kHz bandwidth for a CW signal. From the measured linearity (0.34%) a bandwidth of 56kHz was expected. Gating the receiver widened the received echo width to 95kHz, and re-attaching the PILO increased the signal level by 21.3dB but widened it to 205kHz in this case.

From these tests, it appears that the PILO is the main cause of the degraded range-resolution. This was verified when the PILO was tested in isolation where it was found that if the locking signal was too high (>5dBm), the transmitted signal spectrum widened and then divided into two individual spectra. Even careful adjustment of the injection signal level could not reproduce as narrow an output spectrum as that on the input. Measurements made showing this phenomenon can be found in Appendix G.

The PILO built by us does not exhibit this phenomenon either in isolation, or as part of the radar. Unfortunately as it does not produce the same amount of power as the Millitech unit, it was decided to install it only if the range resolution proved to be a problem.

The measured range-resolution for a point reflector using different spectrum analyser bandwidths is shown in Figure 5.35 and Figure 5.37.

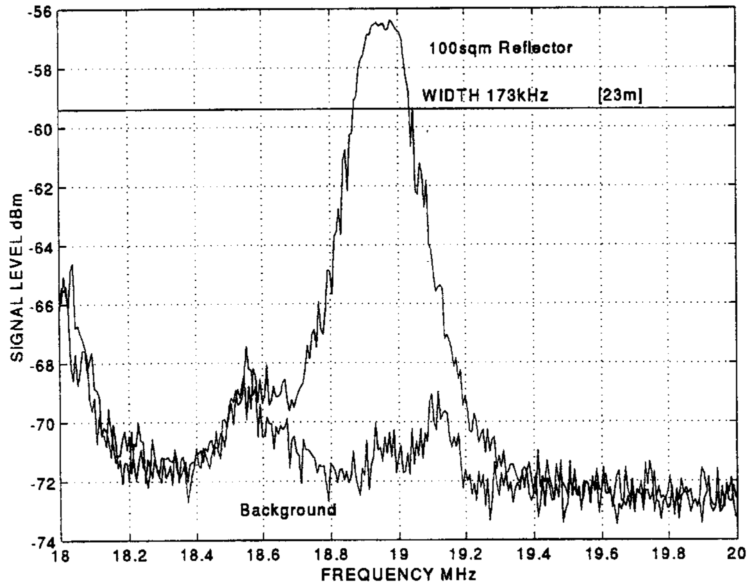


Figure 5.35: Echo return from a 100m^2 corner-reflector and background (30kHz spectrum analyser bandwidth) showing spectral widening caused by imperfections in the PILO, nonlinearities in the transmitted chirp and interrupt effects

The signal-processing bandwidth (gate size) is 50kHz when widened by the Hamming window as discussed in Section 5.8.3, so it should in theory be possible to resolve targets something over 5m apart if they have comparable RCS values. However, measurements made using a spectrum analyser are affected by imperfections in the PILO, chirp nonlinearities and by the spectral widening due to the interrupt process.

It was shown in Figure 5.2 that the spectrum for an FMICW radar with a 20μs + 20μs interrupt cycle produces a spectrum with peaks separated by 25kHz. Measuring the received signal using a spectrum analyser bandwidth of 30kHz does not have sufficient resolution to resolve the individual components of the spectrum. Figure 5.36 shows the simulated result obtained by convolving the spectrum obtained in Figure 5.2 with the spectrum-analyser transfer function.

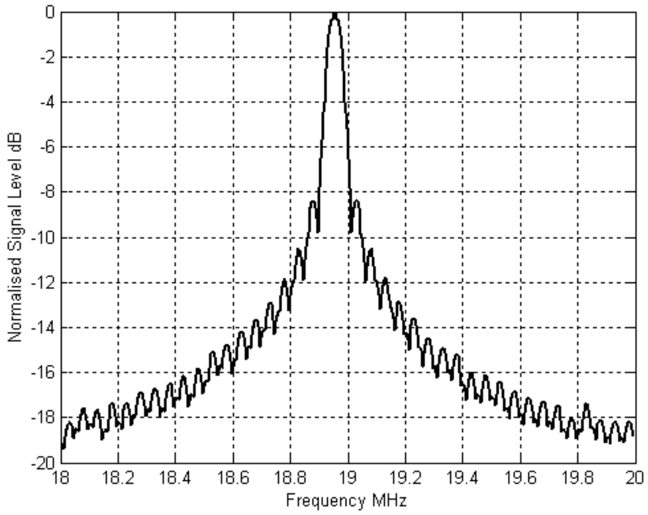


Figure 5.36: Simulated spectrum for an FMICW radar with a 20μs + 20μs interrupt cycle and a point target for a 30kHz spectrum analyser bandwidth

These results show that a combination of the interrupt process and the spectrum analyser bandwidth widens the spectrum to more than 60kHz and that the remaining spread is caused by a combination of the nonlinearities in the chirp signal and the PILO effect.

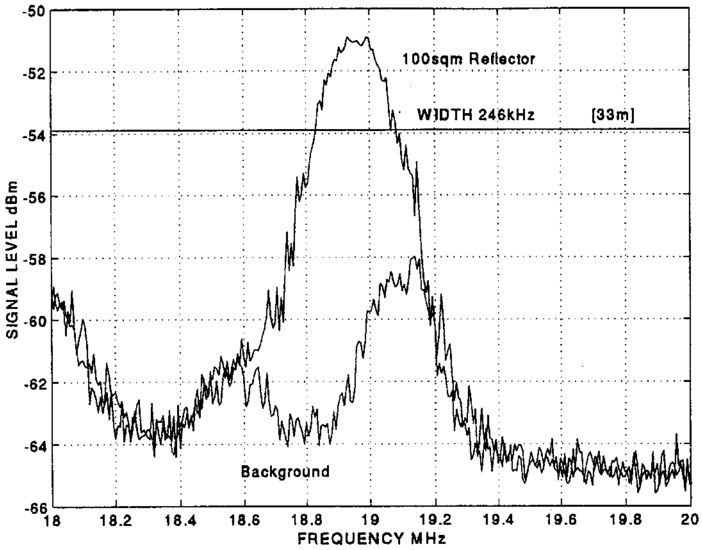


Figure 5.37: Echo return from a $100m^2$ target and background (100kHz spectrum analyser bandwidth) showing spectral widening caused by imperfections in the PILO, nonlinearities in the transmitted chirp and interrupt effects

5.6. Performance Prediction

The view of the target area across the valley was masked by the top of a hill at just over 2600m. Reflectors were placed at a range of about 2526m for these tests. As the site was a little higher than the radar position, the radar was rotated up by 1° to maximise the return echo amplitude. At this point the ground slope was 3° making the grazing angle close to 2°. Moving from the crest of the hill towards the radar, the slope falls away with a resultant increase in ground slope as shown in Figure 5.38.

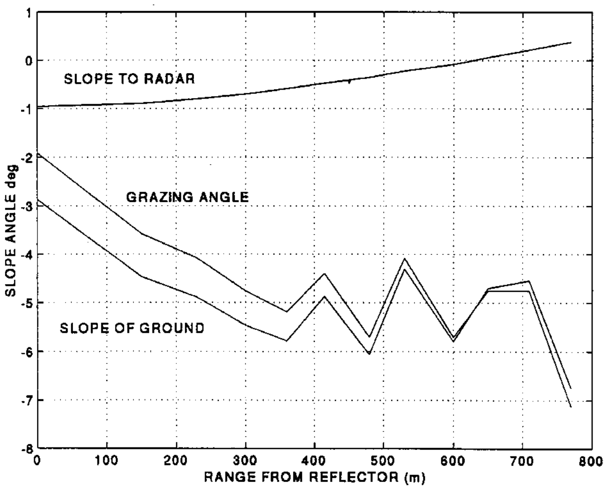


Figure 5.38: Characteristics of the measurement site showing the elevation angle to the radar, the slope of the ground determined from a contour map, and the grazing angle at target

Figure 5.39 and Figure 5.40 show the measured range profile for two different spectrum analyser bandwidths. It can be seen from the range to the targets and the frequency of the peak that the radar timing was configured for operation over a 2km to 4km span for these tests.

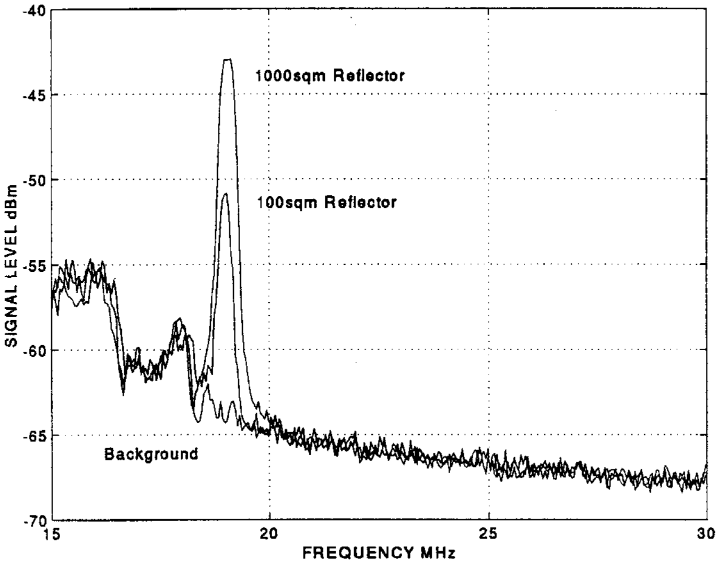


Figure 5.39: Measured range profile for the FMICW radar using a spectrum analyser with a 100kHz bandwidth

These range profile spectra show ground clutter returns at frequencies from 15MHz to just less than 20MHz (2000m – 2660m) with the target return just below the brow of the hill at 2526m. The figures include spectra made with and without the reflectors so that the magnitude of the target return can be measured and the SNR and signal to clutter ratios can be determined.

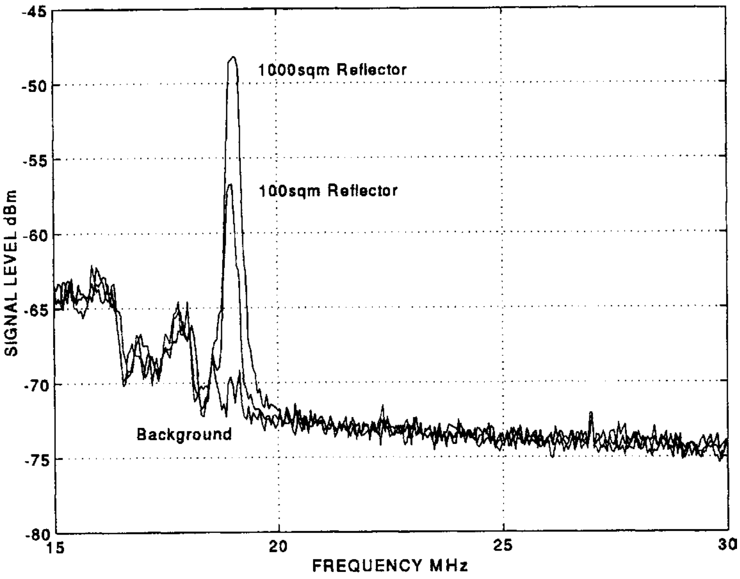


Figure 5.40: Measured range profile for the FMICW radar using a spectrum analyser with a 30kHz bandwidth

As the spectrum analyser displays the peak signal and noise levels for a single sweep, the levels presented include the artefacts of the interrupt process as well as the broadband noise introduced by the switching transients shown in Figure 5.19. This will result in poorer SNRs than will be obtained when the return echoes are processed independently. In addition, the non-coherent integration of a number of cycles will improve the SNR still further.

5.7. Antenna Scanning and Beam Stabilisation

A set of gimbals was developed to perform the antenna scanning and beam stabilisation functions needed for airborne imaging. These gimbals were developed in two phases. In the first phase the elevation and azimuth axes were built, and in the second phase which followed some two years later, the roll axis was built to produce the mechanism shown in Figure 5.41.

Mounted on the roll axis are three servo amplifiers to power the three motors which drive the various axes using wire-drive pulleys as shown. Precision potentiometers with analog state-variable filters to smooth position and estimate angular rate are used as feedback elements. A digital controller was implemented to control the antenna position, and to compensate for base motion using signals from an accurate IMU.

To test the stabilisation, the radar and IMU were suspended from a single rope that was balanced to allow the complete structure to pitch and yaw freely. A laser pointer was mounted to the antenna and the position of the spot projected onto a distant wall was monitored as the unit was rocked. A rough analysis of the spot position showed that the beam position accuracy was better than 0.2°. Though this method could not reproduce the specific base-motion of a helicopter it did give an indication of the performance which was later verified by the quality of the images obtained during flight trials.

Low-power FMICW radars are extremely sensitive to extraneous sources of noise, particularly that from switch-mode power supplies and pulse-width modulated servo amplifiers. A great deal of careful design was required to ensure that these potential noise-sources were adequately filtered and shielded to minimise both radiated and conducted interference. Though this aspect of the radar design is beyond the scope of this thesis, its importance in ensuring good radar performance cannot be stressed enough.

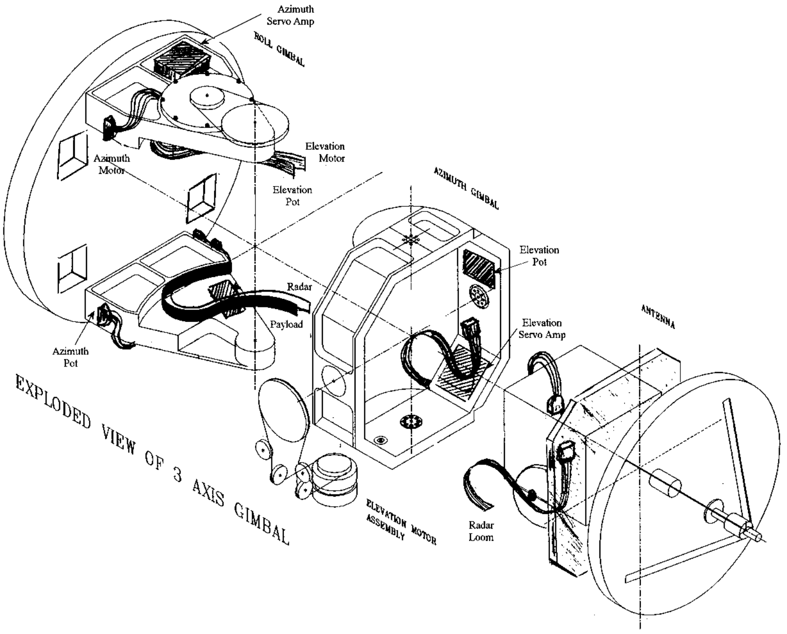


Figure 5.41: An isometric drawing of the radar hardware showing the roll, yaw and pitch axes with a front-end mounted to the inner gimbal plate

5.8. Signal Processing

The signal processing involves range gating to produce a range-echo profile over a range from 1.5 to 3km. This function is repeated as the antenna is scanned in azimuth so that a ground reflectivity image can be constructed. This image can then be processed in a number of ways to perform navigation and guidance functions.

5.8.1. Coarse Range Gating

Range gating is achieved by analysing the spectrum of the beat signal. For a coherent system it is common practise to downconvert to baseband using two quadrature oscillators to produce in-phase and quadrature components prior to filtering and digitisation. However, gain and phase imbalances between the two channels can result in poor performance.

An alternative technique [159] is shown in Figure 5.42. In this case the signal is mixed down to a new centre frequency, B. The negative frequencies, $2f_o-B$, are removed by a bandpass filter so that the signal spans a frequency from $B/2$ to $3B/2$. These data are then sampled and digitised at a sampling rate of $4B$.

This process could be performed directly on the signal spanning 15 to 30MHz which corresponds to the range from 1.5 to 3km, but this would require that the received echo be sampled and digitised at 120MHz which is not feasible for the required 12-bit resolution. An alternative is to divide the spectrum into 12 coarse gates of 1.25MHz each using a stepped down-converter.

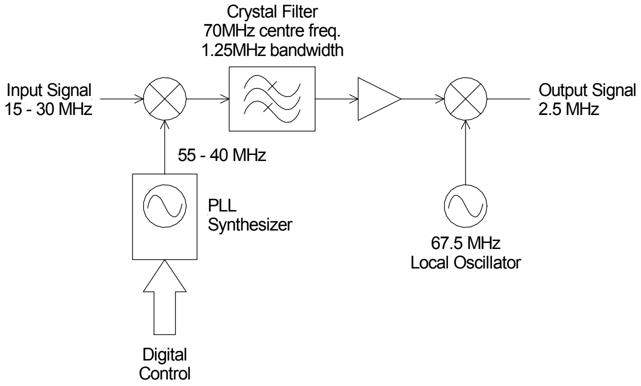


Figure 5.42: Schematic block diagram of the coarse analog-range-gating process implemented using a stepped down-converter and a narrow-band crystal filter

In this implementation, the signal is first up-converted to a centre frequency of 70MHz where it is filtered using a crystal bandpass filter with a bandwidth of 1.25MHz, extremely steep roll-off and low sidelobes. It is then down-converted to a centre frequency of 2.5MHz where it is sampled and digitised at a more practical rate of 10MHz.

5.8.2. Generating In-Phase and Quadrature Signals

For the sampled-data signal, the objective of the digital filter $H(z)$ shown in Figure 5.43 is to eliminate the negative-frequency sideband while leaving the positive sideband unaltered. This is achieved by implementing a pair of all-pass filters $H_1(z)$ and $H_2(z)$

$$\left|H\left[\exp(j\theta)\right]\right| \approx \text{constant} \qquad \pi/4 < \theta < 3\pi/4 \tag{5.26}$$

$$\left|H\left[\exp(j\theta)\right]\right| \approx 0 \qquad -3\pi/4 < \theta < -\pi/4 . \tag{5.27}$$

The two filters are designed so that the phase difference in the pass band deviates from 90° by a small error of no more than ε for a gain of 2 and that the stop band gain is no greater than $\sin(\varepsilon)$ which is 60dB down. This requires a phase error $\varepsilon < 0.1145°$.

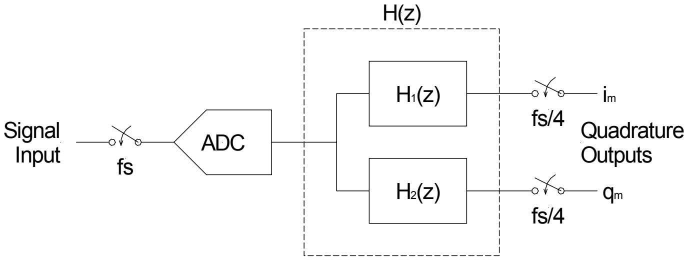


Figure 5.43: Block diagram showing the components required to generate the in-phase and quadrature signals from the radar IF output

A procedure for designing all-pass phase shifting networks based on Jacobian elliptic-functions [92] is applied to produce the following 5th order filter construction

$$H_1(z) = z^{-1}\frac{z^{-2}-a^2}{1-a^2z^{-2}}, \tag{5.28}$$

$$H_2(z) = -\frac{z^{-2}-b^2}{1-b^2z^{-2}}, \tag{5.29}$$

$$H(z) = H_1(z) + jH_2(z), \tag{5.30}$$

where $a^2 = 0.5846832$ and $b^2 = 0.1380250$.

For a sample rate of 10MHz, this filter generates the magnitude and phase responses shown in Figure 5.44.

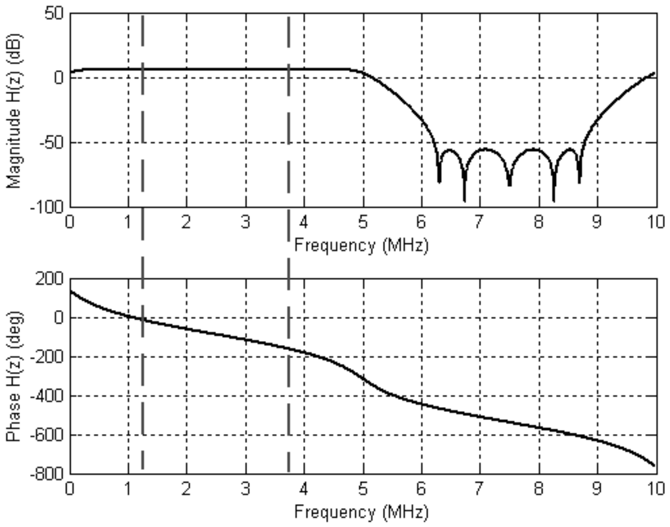


Figure 5.44: Magnitude and phase response of H(z) where (a) is the magnitude response and (b) is the phase response with step jumps of π removed to make it continuous

In the region of interest between $B/2$ (1.25MHz) and $3B/2$ (3.75MHz) marked by the dotted lines, it can be seen that the amplitude response is completely flat and that the phase is linear. When the filter output is sampled at 2.5MHz, it has the effect of aliasing this section of the filter down to the baseband where it can be processed to extract the echo signal spectra.

The efficient implementation of this filter relies on the fact that output decimation occurs (only one in four samples is used), and that $H_1(z)$ uses only odd samples while $H_2(z)$ uses only even samples. This results in the realisation shown in Figure 5.45.

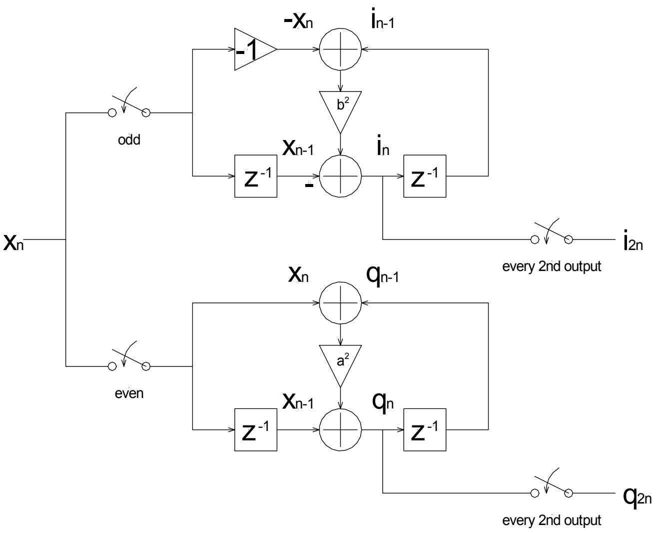


Figure 5.45: Realisation of the I-Q generation transfer function, H(z) configured to exploit post sampling decimation

5.8.3. Fine Range Gating

The in-phase and quadrature signals are weighted by the Hamming window to reduce range sidelobes and processed through a complex 64 point FFT to produce 64 bins spanning 2.5MHz that corresponds to the input frequency between 1.25 and 3.75MHz.

With 64 bins spanning 2.5MHz, the individual bin size is 39kHz which, with Hamming weighting, increases to 50kHz resulting in a some bin overlap and leakage. In theory the down conversion and I-Q split prior to the spectral analysis step does not introduce any processing mismatch. However, the Hamming weighting does introduce a slight loss in SNR compared to a matched filter as discussed in Appendix F. This 1.34dB loss is a small price to pay for the reduction in the level of the range sidelobes.

To improve the signal to noise ratio, the range spectra for a complete up and down chirp (4 interrupt cycles) are shifted to compensate for any radar motion-induced Doppler and integrated into a single group of 64 bins. As the down-conversion process uses a free-running oscillator, coherence is lost and only post detection (non-coherent) integration of the spectral envelope is possible. This results in an integration improvement of about 5.4dB compared to a maximum of 6dB.

New spectra are generated sequentially after the down-converter frequency has been stepped until the whole spectrum from 15 to 30MHz has been produced. Because the down-converter is stepped by 1.25MHz, the output spectra overlap by 50% meaning that only the 32 central bins are unique. However, as the analog signal is filtered by a filter with a bandwidth of 1.25MHz prior to digitisation, the signal in the bins at the two extremities is attenuated. A continuous seamless spectrum is built up by averaging a few of the overlapping bins from adjacent blocks.

As shown in Figure 5.15, it is only at a range of 3km that the return signal fills the full 20μs receive period. At all other ranges the shorter round-trip time results in the transmission eclipsing part of the return. Because the range to each block of bins is known, the approximate proportion of the 20μs receive period that contains an echo signal is known and the remainder is packed with zeros to maximise the SNR. This shortened observation time degrades the theoretical range resolution, until at a range of 1500m it has doubled from 50kHz to 100kHz as shown in Figure 5.46.

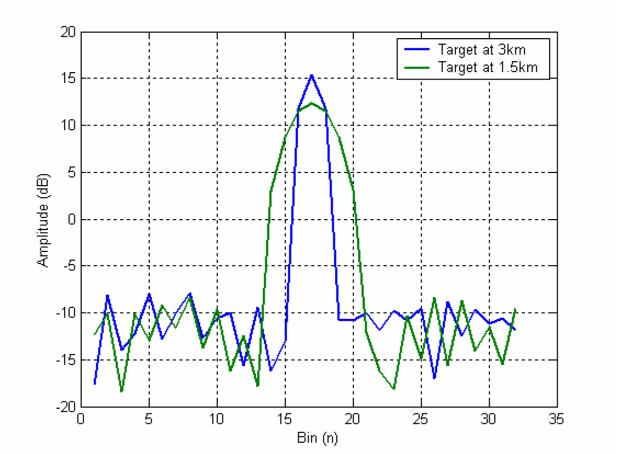


Figure 5.46: Simulation of the received signal spectra showing the difference in the range resolution at 3km and 1.5km caused by eclipsing

5.8.4. Image Generation

Because the passband of the bandpass filter is not completely flat, as can be seen from noise floor measurements shown in Figure 5.47, a gain correction is applied to the contents of each of the bins to compensate for this variation. A gain-table is automatically generated by pointing the radar at the sky and measuring and storing the average noise-floor in each of the 384 bins which is used for this purpose.

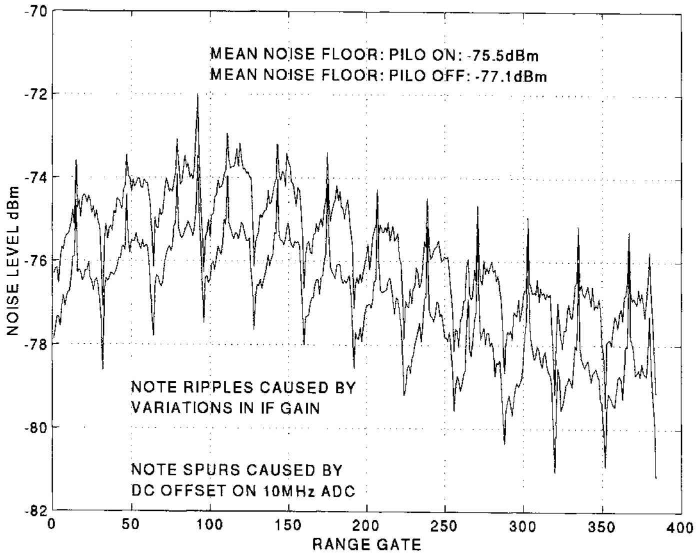


Figure 5.47: Measured noise floor for the FMICW radar aimed at the sky, showing gain ripples and DC offset effects due to imperfections in the down-converter and ADC modules

The stabilised antenna is scanned in azimuth over an angle of 30° in about 7s during which time equally spaced range profiles are made every 0.23° and assembled to produce a complete polar image as shown in Figure 5.48.

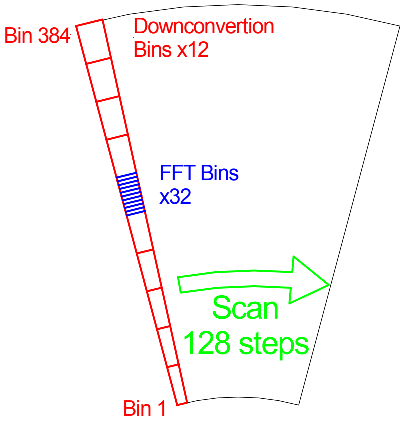
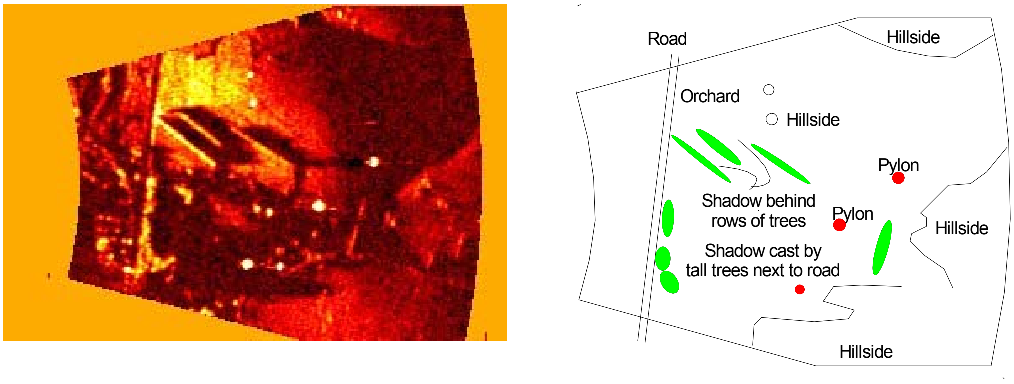


Figure 5.48: Gating process for image generation showing the construction of 384 range gates made up of 12 coarse × 32 fine gates mechanically scanned over 30° in 128 steps

From the measured antenna beamwidth shown in Figure 5.30, the two-way 3dB beamwidth is determined to be about 0.29° which is only slightly larger than the 0.23° sampling interval. Ideally the return should be sampled at an interval corresponding to half the beamwidth to avoid spatial-aliasing, but this was not feasible in this case because of the limited available processing capability of the radar and the limited time available to perform a scan.

Some enhancements have been made to improve the image quality. These include re-sampling the polar image into Cartesian space (which is discussed further in Chapter 6), and the integration of a number of images to produce the clear image shown in Figure 5.49.



(a) (b)

Figure 5.49: FMICW radar reflectivity image of a rural environment and a sketch describing the features visible in the image. Tiny marks towards the bottom of the image show the theoretical resolution of the radar at 1.5 and 3km

5.8.5. Measured Resolution Effects

The radar image shown in Figure 5.49a includes markings towards the bottom left and the bottom right of the image which represent the theoretical resolution of the radar in both range and angle. Non saturated point targets visible in the image show that the range resolution is much worse than this. Figure 5.50 below shows a log-compressed mesh plot of a point target[6] made from polar data. From this plot it can be determined that the 3dB width of the target is about 5 pixels in range and between 2 and 3 pixels in azimuth. This reduction in range resolution is as predicted in earlier in this chapter, while the angular resolution is accurately determined by the antenna azimuth-beamwidth and the sample interval.

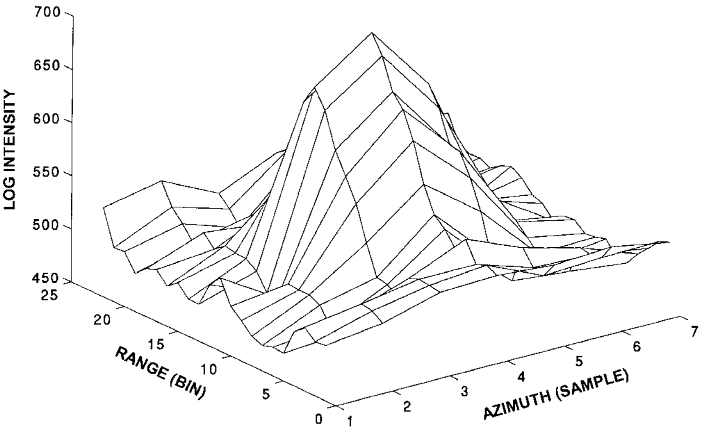


Figure 5.50: Measured radar image of a point target at a range of 2.75km shows that the azimuth resolution is about 2 samples wide and that the range resolution is about 5 bins deep

The range resolution could be improved by first looking at component and implementation problems, then by increasing the chirp bandwidth, improving the linearisation and then replacing the PILO.

5.8.6. Resolution Enhancement Methods

Over the past decade, techniques have been developed to obtain resolutions better than those dictated by the transmitted bandwidth. The literature contains numerous examples of this including the method of moments [18], the method of extent estimation [23] and maximum likelihood length estimation [150]. A novel technique called Linear FM envelope processing [63] relies on some motion between the two targets to aid with their resolution. Additional techniques using modern spectral estimation methods to increase resolution have also been used, and these methods are commonly referred to as super-resolution methods [69]. If the feature of interest is known to be a point target, then it is possible to use a split gate or interpolation to determine its centroid to within a fraction of one bin

Angular resolution, on the other hand, is a more intransigent problem as it is dependent primarily on the antenna aperture, a parameter which is difficult to alter. For a 94GHz radar it is difficult and expensive to obtain an aperture diameter much larger than 400mm, and hence a beamwidth of about 0.55° is the limit. This equates to a cross-range resolution of 30m at a range of 3km which is larger by a factor of four than the required resolution specified in Chapter 1.

---

[6] The "point target" was the steel structure of a windmill from which the blades had been removed.

A two-pronged approach can be followed to solve this problem. In the first instance, it has been shown [143] that it is the pixel area and not its shape that defines the effective resolution required for feature identification (out to a 5:1 aspect ratio). This allows a trade off to be made between the radar's range and cross-range resolution. Secondly, a well documented technique called monopulse imaging uses the difference in the amplitude of a "point" target return in two overlapping beams to determine its true position.

Though it may not be practical to build a monopulse antenna and receiver, both the sequential lobing and the conical scan techniques also produce measures of angle error and should therefore be able to pinpoint the position of a single scatterer. A further assumption is that there is a single dominant scatterer within the resolution cell, otherwise this technique measures the position of the centroid of the ensemble of scatterers. A reduction in the size of the range gate increases the probability that only a single scatterer is present, so it is advantageous to improve the range resolution. This process has been extended to include multiple receivers as well as transmitter position multiplexing [108] in an application which may be practical for a ground-based system.

The required 7×7m resolution for a square pixel can therefore be replaced by a 3.5×15m pixel without significant loss of performance. To achieve a range resolution of 3.5m requires a transmitted bandwidth of about 50MHz which is easy to achieve. To obtain the desired angular "resolution" requires that the 50% overlap in the azimuth scan be processed to determine the scatterer position to within a half beamwidth. This should be feasible using the sequential lobing generated angle error extraction algorithm.

Even though an elongated pixel is not the ideal shape from which to produce images for navigation purposes, under some circumstances it is possible to use the finer resolution in the range dimension to advantage. The following examples illustrate the process for a radar beam traversing a rock-face.

Consider a radar return that contains echoes in eight consecutive range bins shown in Figure 5.51.

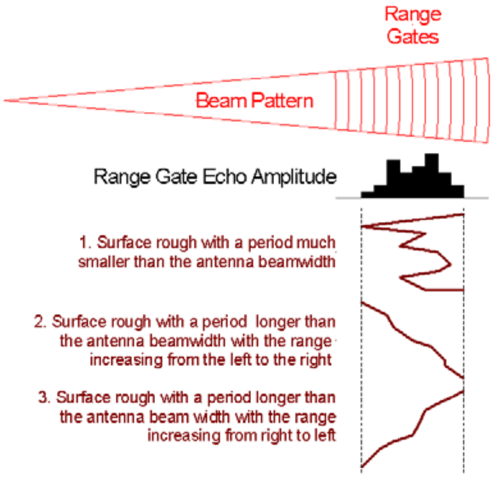


Figure 5.51: Hypothetical surfaces that can be reconstructed from a single range-gate echo profile

From a single observation it is not possible to make any predictions about the likely structure of the target surface unless some *a priori* knowledge is applied. If hypothesis one (listed in the Figure) is true, then no inferences can be made with respect to the shape of the target except that it contains facets that reflect the radar signal at all of the ranges indicated. If one of the other hypotheses is true, then from a single observation, they are equally likely. However, as soon as a second observation is made by rotating the antenna even a fraction of a beamwidth, by noting which bin grows and which shrinks due to changes in the antenna gain in that direction, or in which bin a new target appears, one hypothesis begins to collapse and the other is reinforced. To automate this process in 2D is explained with the help of Figure 5.52.

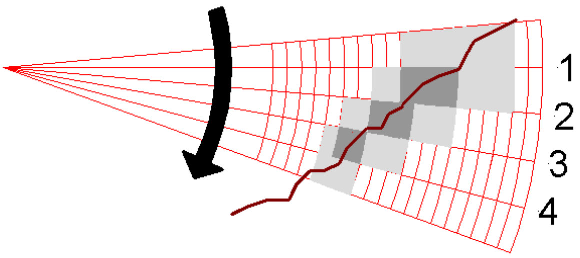
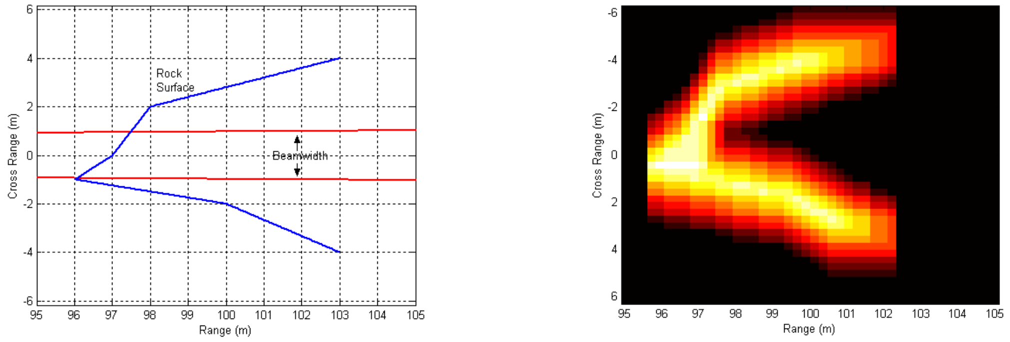


Figure 5.52: Doubling the effective angular resolution by stepping the beam by one half beamwidth and integrating the contents of the overlapping beams

After the first sample (1), each of the 16 bins (eight range bins over two half beamwidths) is considered to be occupied. The beam is then stepped by one half beamwidth to position (2) and a new reading made. In this case, 12 bins are occupied of which the four shaded a darker grey overlap with those from the previous measurement. This proceeds for step (3) and (4) to produce an overall measurement with an effective resolution, in this case, of about half the beamwidth.

A simulation of a 2D radar scan across an irregular target at a range of about 100m using a radar with a beamwidth of 1.12° and a range resolution of 0.25m is used to produce integrated images for an angular step size of 0.1 beamwidths.



(a) (b)

Figure 5.53: Simulation result showing (a) the radar beamwidth superimposed on the rock profile to be scanned and (b) the resultant integrated output for a step size of 0.1 beamwidths.

It can be seen that the simulation output produces a reasonably accurate 2D representation of the rock face with a better resolution than that provided by the antenna.

5.9. Conclusions

The lower noise floor and fewer leakage problems encountered during the use of FMICW as compared to FMCW makes it the modulation technique of choice for long-range operation.

In this chapter the generalised equations that describe the response function for an FMICW waveform were derived. These show that the ambiguity function for this waveform is a function of the round-trip time to the target as well as Doppler and delay. Simulation results based on the process compare the performance of a classical FMCW radar with both deterministic and pseudorandom interrupt sequences. It is shown that neither technique is suitable for use in an imaging radar where the range of interest spans 50% of the range from the radar to R_{max}. The proposed solution to this shortcoming is to process each of the transmit-receive interrupt cycles independently and then to integrate the results non-coherently. This solution is a hybrid of the FMICW/Stretch technique.

The rest of the chapter documents the development of a radar for an airborne navigation application based on this technique, with its associated gimbals and signal processing hardware and software. The protracted development of this radar which is based on a number of less successful prototypes (not documented here) highlights some of the complexities associated with developing radars in the millimetre-wave band. These are generally issues relating to unexpected component interactions.

It is shown using measurements that, for the most part, the radar in its final form functions as expected and in so doing meets the sensitivity requirements for terrestrial imaging at long range. It does suffer from unexpectedly poor range-resolution (about a factor of four worse than predicted) which is primarily due to a non-ideal amplifier component.

The poorer resolution notwithstanding, images produced by the radar more than meet expectations in terms of their dynamic range and the ease with which ground-based features can be identified.

Finally, the limited bandwidth and rigid signal processing structure implemented makes it difficult to use this radar for broader-based research applications. Of particular interest are techniques that can be used to obtain both long-range and high-resolution images from a real aperture. This has led to the requirement for more flexible hardware and software which is discussed further in Chapter 7.

Chapter 6.

Radar-based Navigation and Guidance

6.1. Introduction

This chapter is concerned with the generation of gate-width-limited radar images from airborne and ground-based platforms and their use for navigation of an autonomous vehicle.

A typical airborne autonomous-navigation application would have a UAV traversing unknown territory between waypoints guided only by GPS/INS. In the event that GPS is not available, any drifts and biases in the INS will result in an increasing uncertainty in the position of the platform.

A radar system mounted in the platform can be used to perform a number of different functions that may be used to minimise this error:

- Produce a Doppler map of the ground return to determine the sideslip angle and the rate of lateral drift.
- Track high reflectivity features on the ground in range and angles to perform drift estimation.
- Match known features from the surveyed waypoint area with those measured by the radar to determine the accumulated radial and lateral navigation errors to that time.
- Perform a correlation of an aerial photograph or synthetic aperture radar image of the waypoint area with the radar image to perform the offset estimation.

This chapter deals with long-range waypoint-based navigation methods in isolation. It assumes waypoints are placed with sufficient regularity and cover sufficient area to ensure that the radar image will always fall within the surveyed area.

This process of navigation starts with the production of a reference image or feature database that may be used for correlation purposes. It is then concerned with obtaining an appropriate high-resolution radar image produced using a scanned system such as discussed in Chapter 5.

The gate-width-limited image-building process consists of the production of Cartesian images pinned to a fixed point on the earth as determined by a GPS/INS based estimate of the radar position at that time. This image can either be used in that form as one of the inputs to a 2-D correlator, or salient features can be extracted and matched to features within the waypoint area. Either way, a highly accurate estimate of the platform position relative to the waypoint can be obtained.

AGV navigation is more complex and can be broken up into four distinct processes. The highest level process involves using GPS and INS to plan the route. The following two processes involve 2-D and 3-D time-of-flight sensors for long and medium range path planning using reflectivity and surface slope information. Obstacle detection and collision avoidance is undertaken by the lowest level process.

The chapter concludes by examining the feasibility of using a number of modified gradient-descent techniques on a long-range reflectivity map to guide an AGV over various terrain types. The assumption made is that the terrain in which the vehicle finds itself at the start of the traverse is safe, and that if possible it should remain within a similar clutter environment. Under some conditions the vehicle will get stuck at local minima unless the gradient is made very steep. It was found that the most successful technique involved setting waypoints within the image to ensure a specific trajectory that minimised the required gradient.

6.2. Millimetre-wave Radar Image Generation and Interpretation

The FMICW radar discussed in Chapter 5 was integrated onto the 3-axis gimbals and inserted within a pod attached to the side of a Puma helicopter as shown in Figure 6.1 and Figure 6.2. An aircraft grade INS/GPS unit provided aircraft attitude data to the gimbals for stabilisation against base motion.

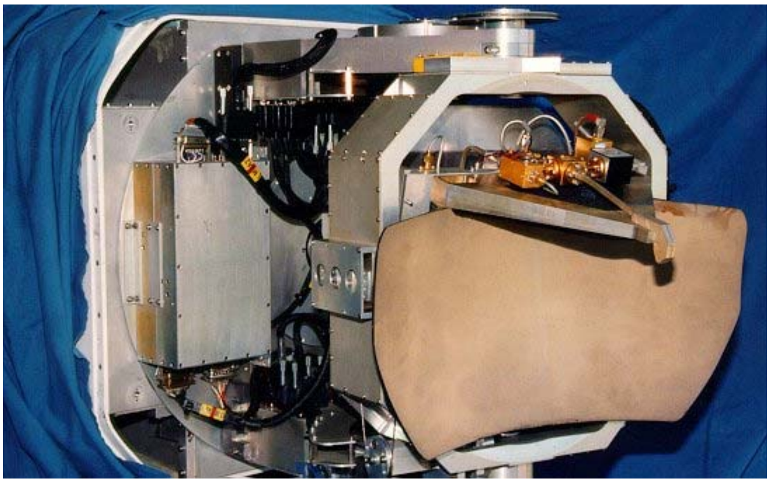


Figure 6.1: Photograph of the FMICW radar front-end attached to 3-axis gimbals

Figure 6.2: Photograph of the radar installed in a pod cantilevered from the side of the Puma helicopter, and the development team, the pilots and engineer

Permission to publish much of this work was obtained from the original client, and more details of the development and results can be found in a number of papers [41], [37], [39], [40].

6.2.1. Image Construction

The range-gating process is repeated a total of 128 times (every 0.23°) as the stabilised antenna is scanned over a 30° angle to build up a polar image of 384×128 pixels. During the process each scan is time-stamped and tagged with the aircraft position and the measured antenna azimuth-angle so that it can be converted from polar to Cartesian space and "laid" down onto a rectangular grid pinned to the earth.

As the Cartesian pixel size is chosen to be smaller than the radar resolution, each polar pixel overlaps on a number of Cartesian pixels which are filled using a weighting proportional to the overlap as described in Section 2.3.1.

This process allows multiple scans to be integrated into overlapping spaces as shown in Figure 6.3. This integration process, if conducted over a period when the drift of the INS/GPS is smaller than one pixel, results in the generation of a clear image with good resolution and good contrast.

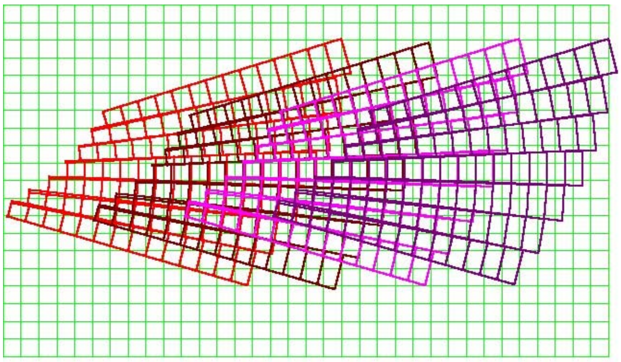


Figure 6.3: Graphical illustration of the spatial integration process from a moving platform

For the initial airborne tests a single 30° scan took about 7s to complete, which for a forward velocity of 30m/s resulted in a displacement of about 200m on average before the same slice was re-illuminated. This period is determined by the range gating and processing overhead which could be reduced by using more processors or faster hardware. It was found experimentally that 3 to 4 scans produce the best results.

Reflections from the individual scatterers in a range cell can combine constructively or destructively to produce an image that is mottled and difficult to interpret by eye. A number of techniques can be implemented to reduce this effect.

It is shown in [184] pp70-78 that the critical frequency change, Δf_{crit}, that is required to decorrelate the returns from two targets D metres apart is given by

$$\Delta f_{crit} = \frac{c}{2D}. \tag{6.1}$$

It is interesting that this corresponds to the chirp bandwidth required to obtain a range resolution of D metres. In an FMCW radar, if the transmitted bandwidth, Δf, exceeds this minimum value, then coherent frequency averaging of the ensemble return from all of the reflectors within the range bin occurs. This is equivalent to averaging N independent measurements where N can be approximated by

$$N \approx \frac{\Delta f}{\Delta f_{crit}}. \tag{6.2}$$

This averaging results in a reduction of the variance of the measured data (less speckle) and a return that is more consistent with the mean reflectivity of the ground at that range [67] p.681. A similar discussion may be found in Section 5.3.4.

This averaging can also be achieved by integrating the image from slightly different perspectives, by a process known as spatial averaging [184] pp67-69 (see Section 3.2.4). This effect is clearly shown below where the image (a) on the left is unintegrated and the image (b) on the right is produced by the integration of a sequence of four images.

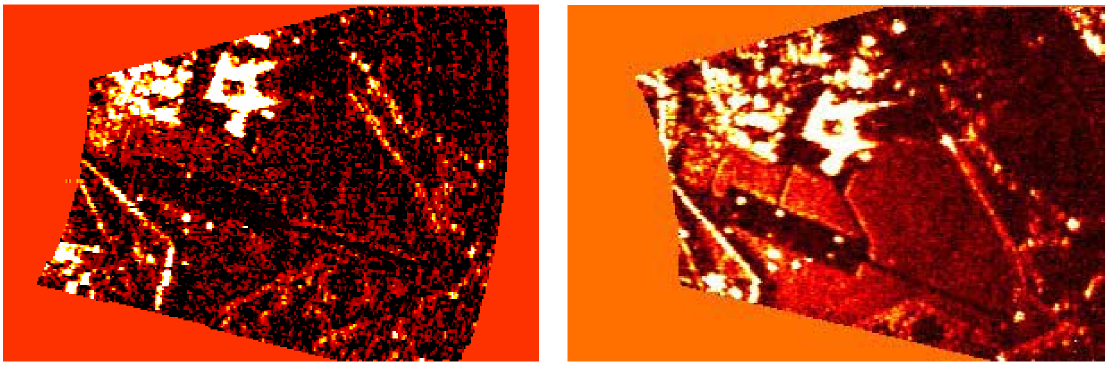

(a) (b)

Figure 6.4: The effect of integration on image quality showing (a) a single unintegrated image and (b) integration of four images from a moving platform

The main features that are visible in these images are a section of runway about 130m wide and 500m long followed by a narrow section about 25m wide and 600m long. Two taxiways lead towards a cluster of hangars that is surrounded by a concrete apron. A wall and clusters of buildings are visible at the narrow end of the image.

The limited dynamic range of the printing process requires that the image be clipped with resultant loss of resolution for the bright objects (particularly the corrugated iron clad hangars) in a process which is discussed in Section 6.3.2.

6.2.2. Image Interpretation from an Aerial Perspective

(a) (b)

Figure 6.5: Comparison between (a) an aerial photograph and (b) a composite radar image of an airfield. This confirms that the runway and taxi ways are clearly visible due to their contrast with the surrounding verges and that buildings and walls appear as bright, high reflectivity returns

The quality of the images is sufficiently good that they can be interpreted by eye. This is shown quite clearly when a comparison is made between the composite radar image and an aerial photograph of the same airfield shown in Figure 6.5. An aerial photograph of the test site taken from directly above is not available, and so the perspective of the two images is different. In this case the radar image appears to be viewed from directly above, as discussed in Section 1.2, and the visual image is viewed from an angle of about 35° above the horizontal.

6.2.3. Image Interpretation from a Pilot's Perspective

If the radar images are to be compared to video images taken from a co-located camera, or from the pilot's perspective, then the display is projected from the range-azimuth (r, θ) plane onto the elevation-azimuth (ϕ, θ) image plane [14] using the transform shown in Figure 6.6. This allows the guidance system to interpret the radar image in the same way it would a video image.

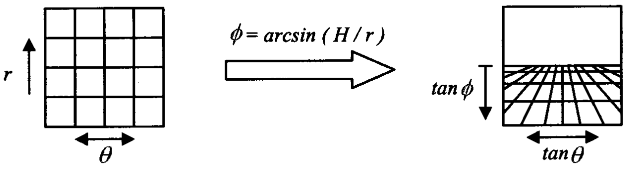
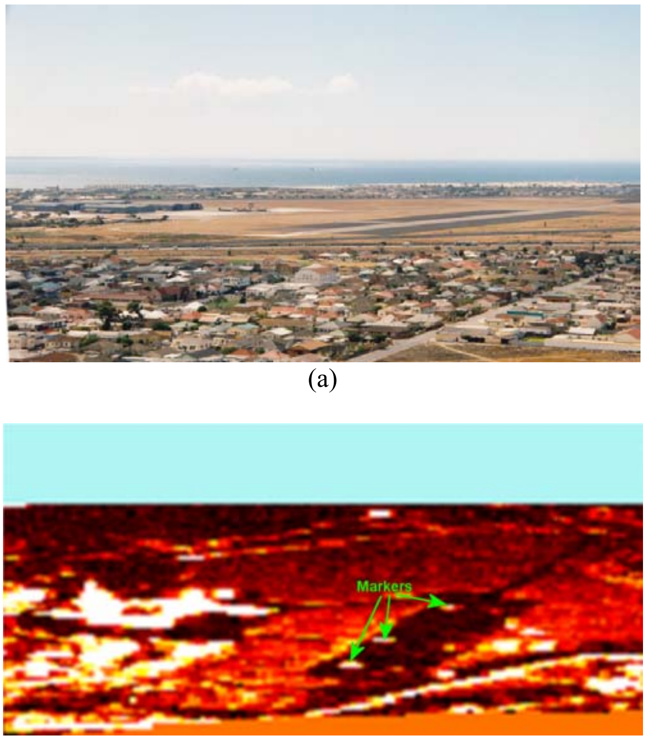


Figure 6.6: Projection from the Range-Azimuth to the Elevation-Azimuth Plane

One advantage of this presentation is that the magnitude of potential obstacles within the image can be exaggerated. This is well illustrated in Figure 6.7 where the runway markers, which are invisible in the aerial photograph, appear quite clearly lining both sides of the runway. These images should be compared to those produced by a commercial synthetic-vision system shown in Figure 2.10.



(a)

(b)

Figure 6.7: Comparison between (a) a low grazing-angle aerial photo of an airfield and (b) a radar image transformed to correspond to a similar perspective

6.3. Feature Based Navigation

In the past, navigation using radar has mostly used specifically placed point-reflectors with known characteristics [57], [125], but some work has been undertaken to extend these techniques to use natural features [58], [112], [124] as discussed in Chapter 2.

The process of localisation is performed by extracting features from the radar image and associating them to reference features stored in a database. The differences between the positions of the reference features and their measured locations are used to determine the vehicle position. In this way the process is very much like a multiple-target tracking problem, but with static targets and a moving observer [15].

One of the major weaknesses of this technique concerns the data association process. Correct error estimates rely on finding the correct correspondence between observed and stored features. If there are sufficient features involved in the process, then erroneous associations of a small fraction of the total will not have a major effect, but as the percentage increases, the error will grow proportionally until it exceeds the uncertainty in the platform position and the vehicle becomes lost.

In the airborne application discussed here, the process of estimating the navigation error is based on natural features and is described in the following section.

6.3.1. Reference-feature-base Construction

Firstly, the area of interest is analysed using aerial photographs and any other available data to produce a series of features that will probably be visible on a radar image as illustrated in Figure 6.8. This process is fraught with difficulty, not only because some large objects, like the hangars, turn out to have relatively small radar cross-sections[7], but also because some incredibly small objects, like corner reflectors, have astonishingly large radar cross-sections.

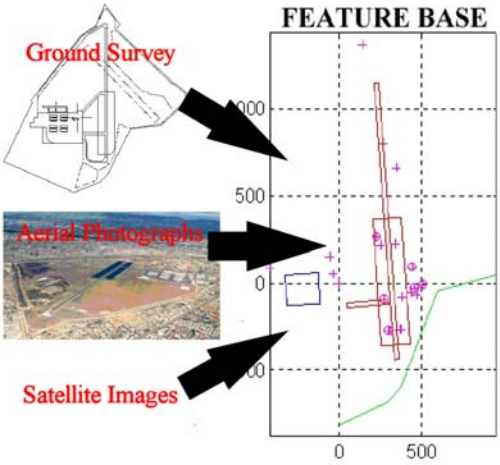


Figure 6.8: Compilation of a feature base for point, line and edge features

Useful features for this process include point reflectors with dimensions smaller than the resolution of the radar, “blobs” with dimensions larger than the resolution and usually made up of an ensemble of closely packed point reflectors, lines which generally consist of rows of point reflectors closely spaced and finally edges across which a change of contrast occurs.

The individual features are reduced to simple polar co-ordinate-based structures defined relative to a fixed point on the image and stored in a database for the area.

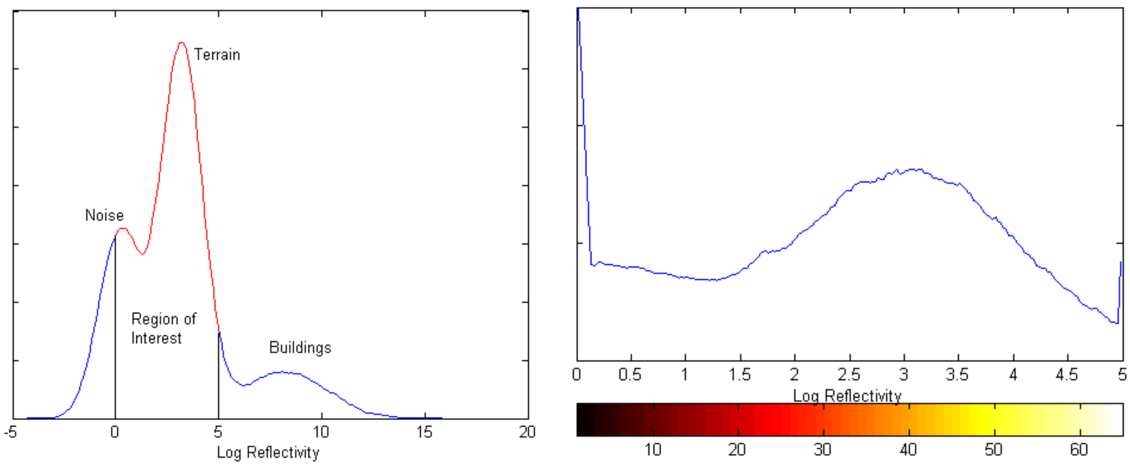
All of the features are defined by their centroids in Cartesian space. Blob features include an additional parameter proportional to their area, while line and edge features include both a direction and a length.

6.3.2. Radar Image Processing

A sequence of four radar images is obtained as the radar flies towards the waypoint. These are normalised by the pixel integration count to produce a single integrated image which is then scaled and clipped to enhance the contrast for low SNR regions shown in Figure 6.9. The un-scanned image background over the rectangular Cartesian space is then filled to the average of the scanned area before further processing.

---

[7] Measurements made by the author have shown this to be the case, particularly for hangars clad with corrugated asbestos sheeting.



(a)

(b)

Figure 6.9: Image enhancement process showing (a) the original image reflectivity histogram and (b) the expanded and clipped histogram and colour map

The literature on low-level image processing and computer vision is full of operators that are optimal for detecting sharp intensity changes in digital images. These techniques are critically dependent on the characteristics of the image. If there is good contrast and little texture, then their performance is good, but if there is little contrast or the contrast is embedded in a textured region, their performance is generally poor. Unfortunately, at short range, millimetre-wave images are generally textured, and at long range the contrast is poor as can be seen from the images shown later in this chapter. Most research on edge detection for radar images is based on high-resolution SAR data. In most SAR data, speckle makes pixel-based filtering ineffective for edge detection, and the larger neighbourhood surrounding each pixel is used to decide whether an edge passes through that pixel [25]. If the images have been averaged, then linear regression and pattern matching techniques can be used to identify partially obscured structures [97].

Another common application for edge detection is for the extraction of road boundaries from automotive-radar images. Because considerable *a priori* knowledge is available concerning the edges of roads and runways, it is possible to apply a few simple rules to aid with the extraction process. The boundaries of the homogenous runway surface are separated from the rougher surroundings by straight, parallel lines, the positions of which can be determined using a deformable template [123]. This can be extended to include curved roads [151] using tightly coupled parabolic curves to define the boundaries. The process is further refined [107] using a simulated annealing process to optimise the fit and escape from local minima. Curved templates are used to extract straight lines because the images are processed from polar data laid-down on a rectangular grid.

For the integrated images described in this chapter, the form of the processing is common to all feature types. First the image is passed through a filter matched to enhance a specific feature type. It is then re-normalised to allow for histogram-based binarisation. Finally the binary image is processed to produce “fact” features which conform to the same format as the reference features. As each feature (including so-called point features) covers a number of pixels, the actual position is determined from the centroid of the normalised return using a sub-resolution technique employed by star trackers [131]. This allows the centroid to be determined to an average accuracy of about 0.02 pixels if the signal to noise ratio is sufficiently good. More details cannot be included here because this algorithm is proprietary information.

As shown in Figure 6.10, the fact feature base contains a list of all of the features that have been extracted from the radar image. It can be seen that the algorithms developed are very effective in extracting point and blob features, but less effective in extracting lines unless the image contrast is very good. Due to this shortcoming, no attempt has been made to actually extract the edges of the runway itself.

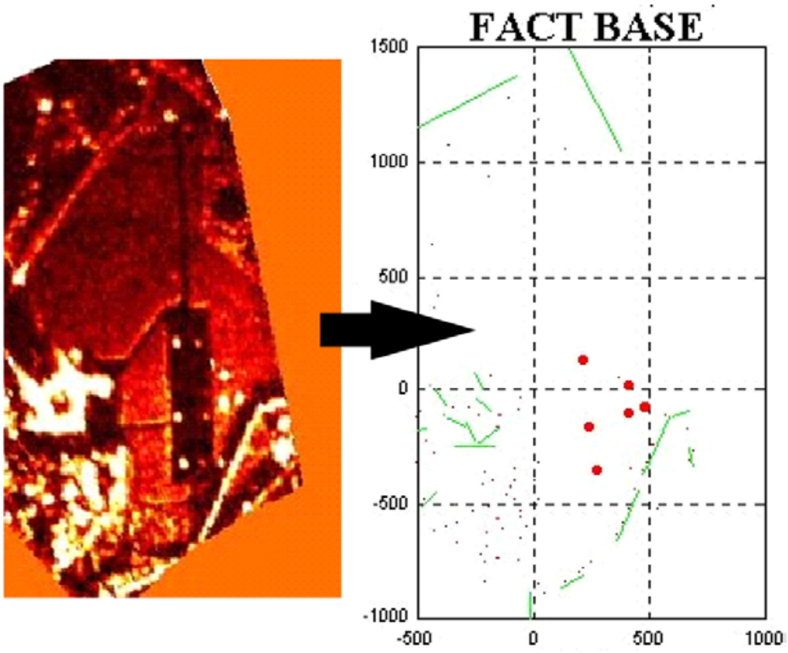


Figure 6.10: Compilation of a fact base from the radar image

6.3.3. Navigation Error Estimation

The process used to determine the navigation-error estimate is as follows [131], [178]:

- Each reference feature of a specific type is compared to similar fact features within a given radius and with similar characteristics (length, orientation, area, etc.) to determine whether they match. A list of possible matches for each feature is stored.

- The distance between all pairs of reference features (shown in green in Figure 6.11) and the distance between pairs of all their possible matches (shown in blue) is then compared and rated according to some tolerance criteria.

- The rating of each fact feature accumulates asymptotically towards unity during the whole search process until finally the fact feature with the highest score is considered to correspond to its associated reference feature.

- The distances between all the references and their associated fact features are then combined optimally to produce an estimate of the navigation offset with respect to the GPS/INS based position estimate.

The software can be run off-line using the stored radar image so that the process can be “tweaked” and anomalous results checked.

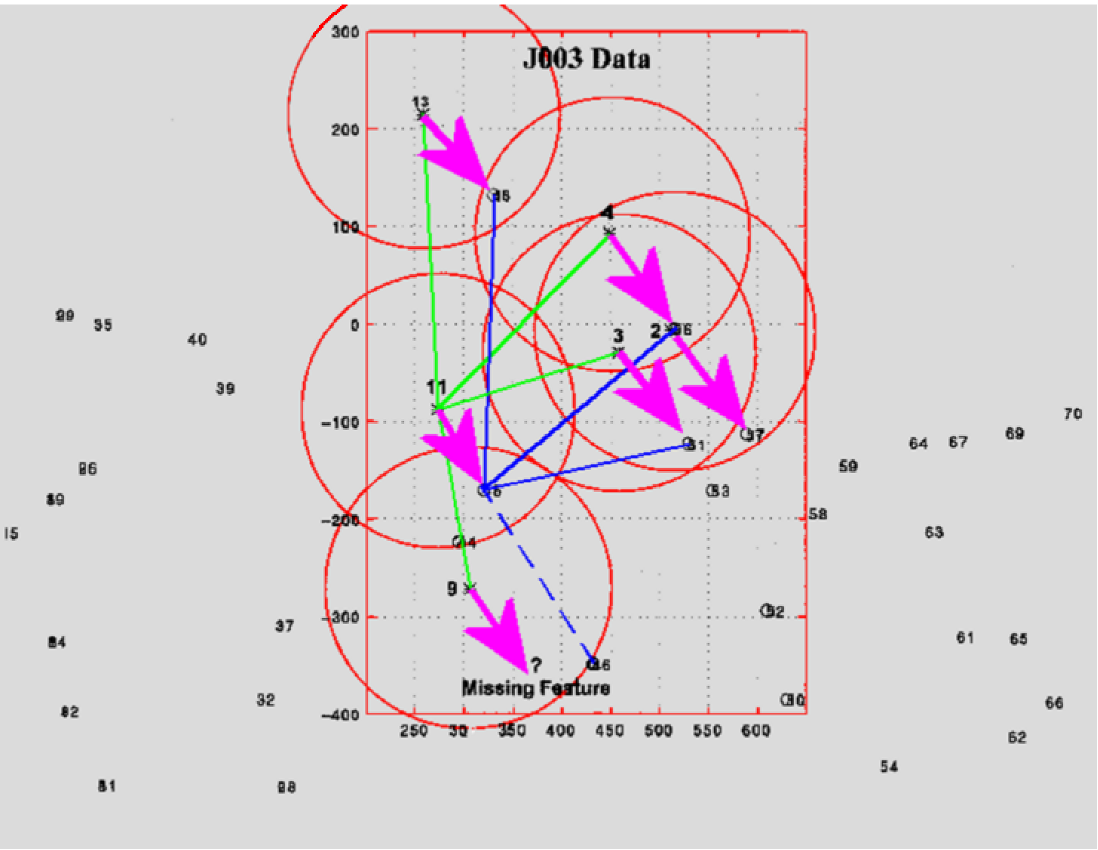


Figure 6.11: Comparison of the distances between feature pairs (green) and fact pairs (blue) which leads to the fact-feature association shown by the purple arrows

Figure 6.12 shows the results that were obtained for six of the images obtained during one flight test. In each of the images a complete list of all point and blob fact-features is displayed at the correct coordinate point. A frame near the centre of the image highlights the area within which the selected reference features, (1,2,3,9,11,13), are situated.

In each case, the algorithm described above is run to associate a specific fact feature with a reference feature. This association is highlighted by plotting a line between the pair. If the positions of the fact features were registered perfectly, then all of the lines would be of equal length and would run in the same direction. However, it can be seen that there is some variation in both the line length and direction.

Small variations in the positions of some of the fact features are probably due to the fact that two of the blob features (a guard tower and a cluster of antennas) are actually each made up of an ensemble of closely spaced scatterers. It is speculated that when these features are observed from even slightly different perspectives, their apparent centroids shift significantly to cause the differences in length and direction observed.

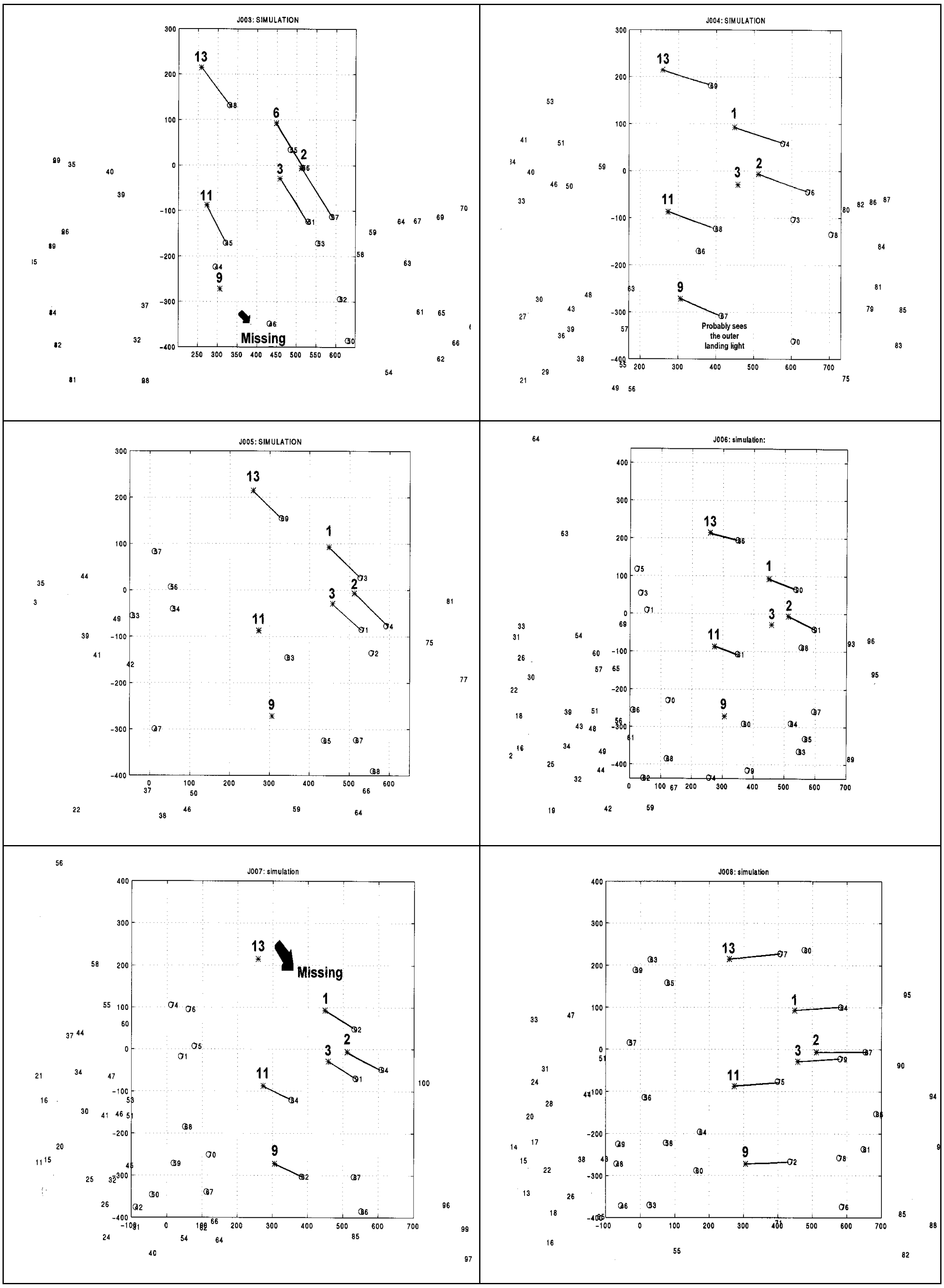
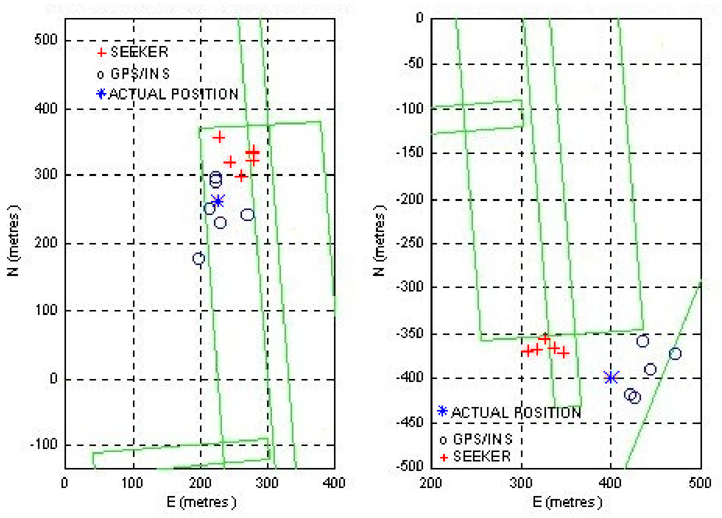


Figure 6.12: Off-line processed results from six images obtained during a flight test

6.3.4. Results

The system performance was measured by flying in the direction of the runway, but offset from its axis and performing the feature extraction and navigation processes before landing at a known reference. The position estimate generated by the GPS/INS system and that generated by the radar were logged. This process was repeated as often as possible during each flight test to obtain a statistically significant sample of the differences between the actual position and the two position estimates, as shown in Figure 6.13.



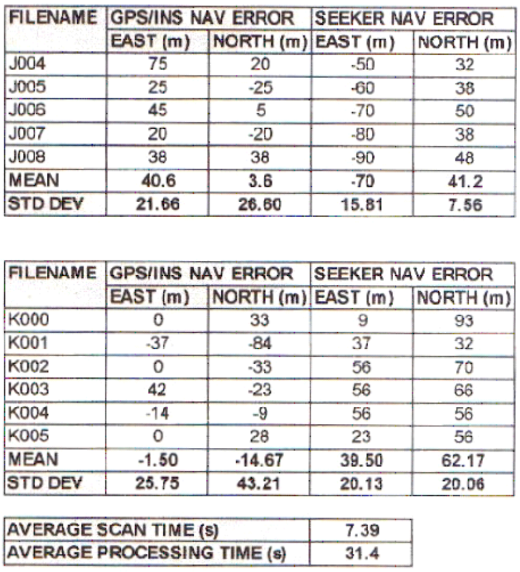
(a)

(b)

Figure 6.13: Measured results of the navigation performance of the radar for two separate flight tests comprising (a) six sorties and (b) five sorties

These results [35] tabulated in Table 6.1, assume that there is no drift in the navigation system between making an image and landing, which is not true as these measurements were conducted with GPS Selective Availability on, and no differential GPS.

Table 6.1: Radar-image based navigation-correction results



FILENAME	GPS/INS NAV ERROR		SEEKER NAV ERROR	
	EAST (m)	NORTH (m)	EAST (m)	NORTH (m)
J004	75	20	-50	32
J005	25	-25	-60	38
J006	45	5	-70	50
J007	20	-20	-80	38
J008	38	38	-90	48
MEAN	**40.6**	**3.6**	**-70**	**41.2**
STD DEV	**21.66**	**26.60**	**15.81**	**7.56**

FILENAME	GPS/INS NAV ERROR		SEEKER NAV ERROR	
	EAST (m)	NORTH (m)	EAST (m)	NORTH (m)
K000	0	33	9	93
K001	-37	-84	37	32
K002	0	-33	56	70
K003	42	-23	56	66
K004	-14	-9	56	56
K005	0	28	23	56
MEAN	**-1.50**	**-14.67**	**39.50**	**62.17**
STD DEV	**25.75**	**43.21**	**20.13**	**20.06**

AVERAGE SCAN TIME (s)	7.39
AVERAGE PROCESSING TIME (s)	31.4

Because the radar system and INS had to be installed into the helicopter prior to each set of flights [36], some alignment differences occur between installations which result in angular offsets between the two measurements. In addition to this there is a slight scale error in the radar range measurements which results in a range offset between the two measurements. However if the mean displacement error is ignored and the standard deviation of the errors in each case is compared, it is found that the radar navigation results are superior to those provided by the combined GPS/INS navigation pack.

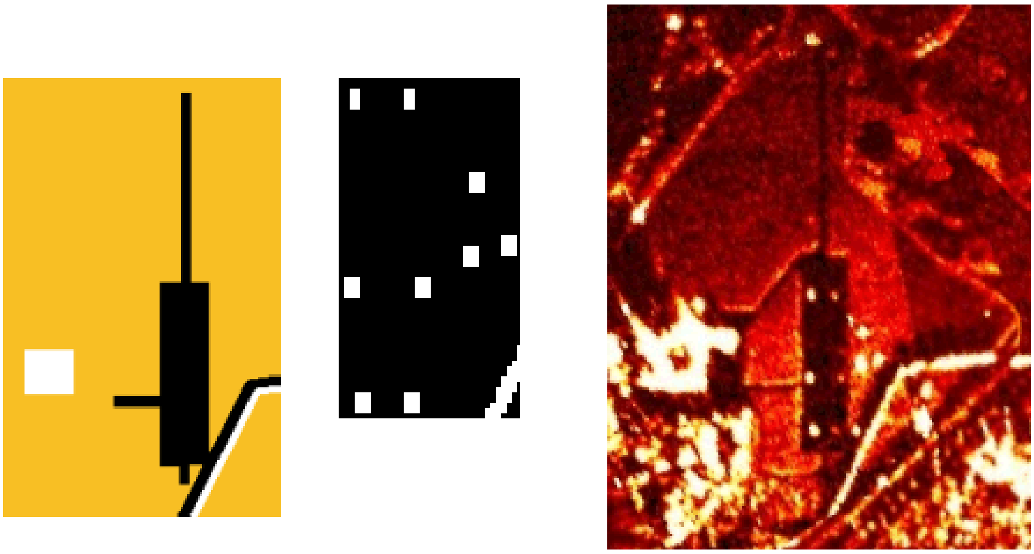
One of the main advantages of this feature-based navigation technique is that it requires very little memory to store points. However because of the uncertainties involved in the image acquisition process, in which even very reliable features sometimes don't appear, as seen in Figure 6.11, it is not as robust a technique as one that includes more information. This limitation is probably outweighed by the fact that the process is insensitive to both image rotation and scale.

6.4. Correlation Based Navigation

2-D cross-correlation [26] between a template containing a number of known target features and a radar image can be used to determine their relative offsets. These offsets (x_{off}, y_{off}) can then be used to determine the position of the platform relative to the ground with sub-pixel accuracy.

6.4.1. The Correlation Process

Using the reference-feature-base described in the previous section as a template, an image of the known features of the airfield was constructed. The first image constructed consisted of the main distributed features like the runway and the wall with the shadow cast shown in Figure 6.14a. A 2-D correlation between this reference image and the radar image was performed to produce a correlation surface [113]. With this reference, the correlator output was a plateau and there was not a well defined peak.



(a) (b) (c)

Figure 6.14: Inputs to a correlation tracker showing (a) a distributed reference image, (b) a point reference image and (c) the raw radar image

A second reference was made that included only point targets similar to (b) but without the wall, and this produced good correlations but with two areas of ambiguity from the two built-up regions across the bottom of the radar image. Finally, a section of wall was added to eliminate this ambiguity, and good sharp correlations were obtained.

It is not known exactly why the distributed image correlator does not function well, because if a section of the radar image is used as a reference, then a well defined, albeit small, correlation peak occurs in the correct position. Even if this radar image is subsequently modified by squaring up the runway edges and selecting the same colours as used in the distributed reference, a reasonable correlation peak is achieved. It would appear that very subtle differences between the radar-based reference image and that constructed from surveyed data points is sufficient to smear the peak into the surrounding surface.

6.4.2. Correlation Results

Figure 6.15 shows the correlation output from reference image (b) and the raw radar image (c). It is clear that a single well defined correlation peak appears near the centre of the image. The peak has twice the amplitude of any of the other peaks on the correlation surface.

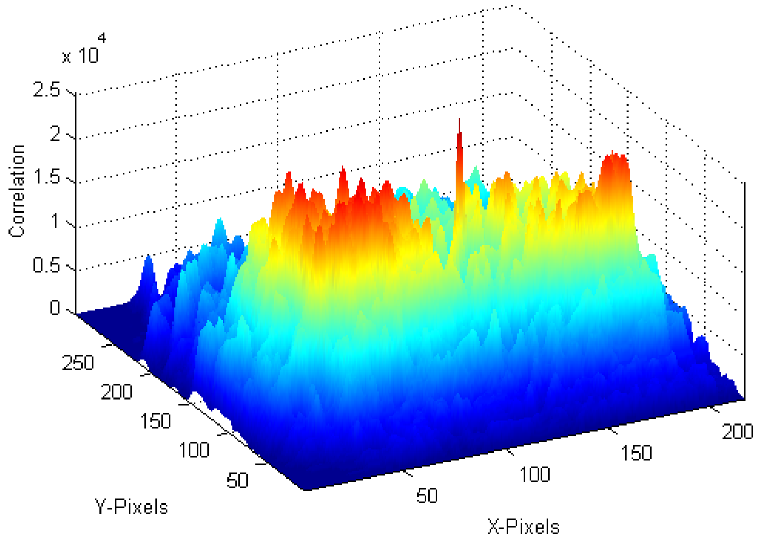
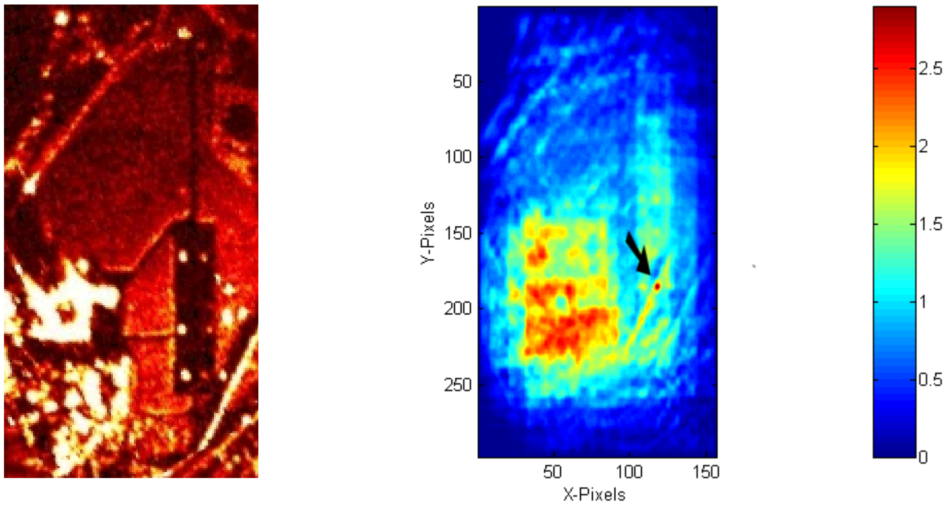


Figure 6.15: Correlation output from a reference image made up of point targets and a section of wall, and a raw radar image shows a single unambiguous peak

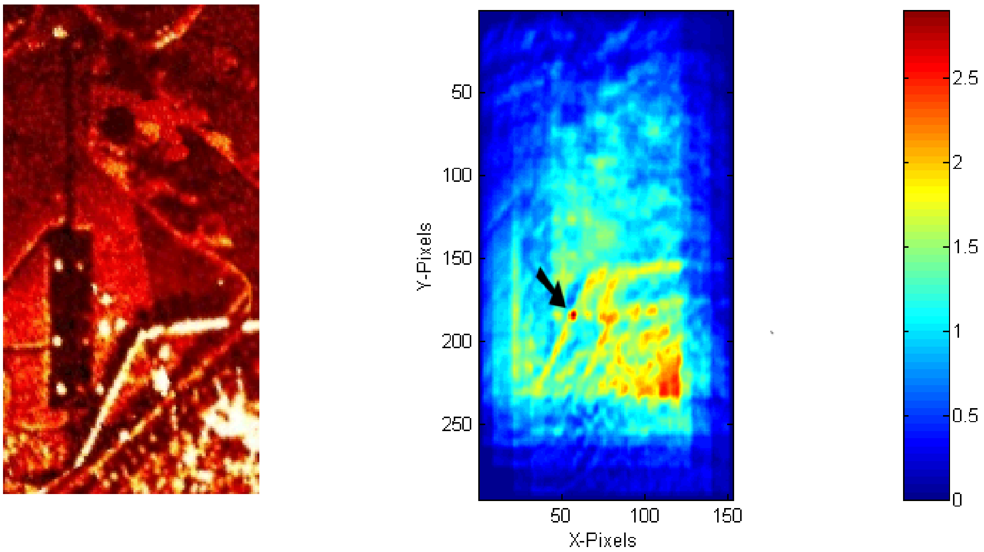
It is incredible that even though this image covers a square kilometre, a specific target can be found within it, to an accuracy of better than one pixel.

To confirm that the correlator peak is in fact generated by the correct features of the image, the radar image was broken into two views, one of the rightmost two thirds of the image, and the other of the leftmost two thirds. The point-and-wall based image was correlated with these to produce the two correlator outputs shown, displaced by the correct number of pixels, in Figure 6.16 and Figure 6.17 respectively.



(a) (b)

Figure 6.16: Right image view and correlation output showing a single clear peak



(a) (b)

Figure 6.17: Left image view and correlation output showing a single clear peak

6.4.3. Positioning Accuracy

Taking orthogonal slices through the coordinates of the peak correlator output shows that the peak may not be centred on a pixel. By applying quadratic interpolation to the peak and the points straddling it as shown in Figure 6.18 and Figure 6.19, it is possible to obtain an estimate of where the maximum should occur and hence to measure of the error down to about 0.01 pixel widths (about 40mm).

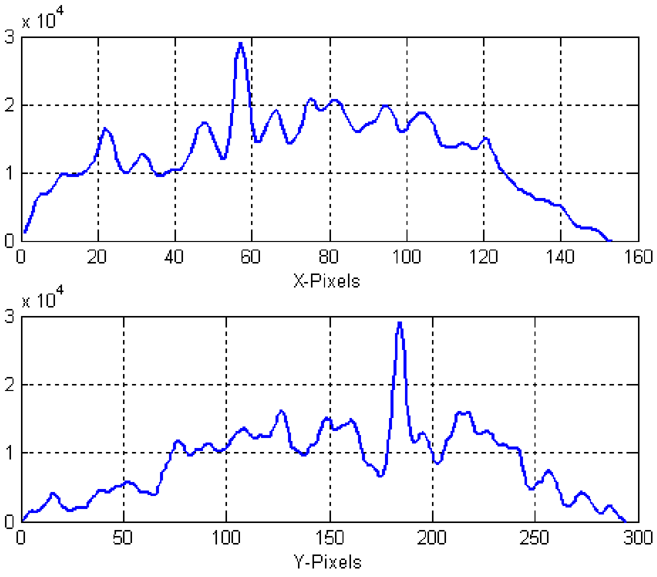


Figure 6.18: Orthogonal slices through the correlator output cutting through the peak

In this example, five points have been used to fit a 2nd order polynomial and the highest point has been found for each of the two peaks.

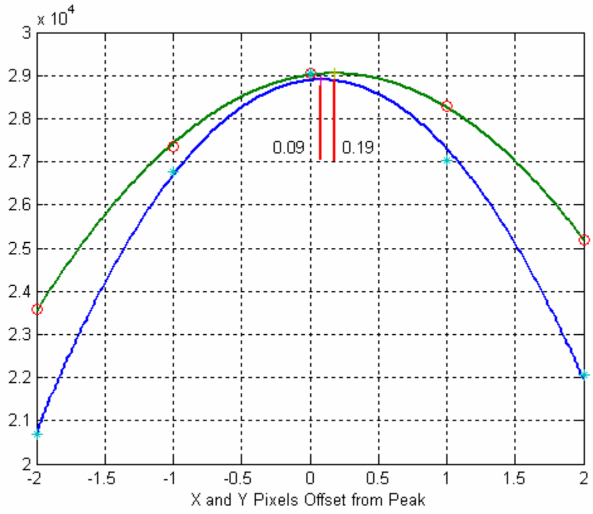


Figure 6.19: 2nd order polynomial fit across the peak in the two orthogonal axes is used to determine the point of correlation to a fraction of a pixel

The coordinates of the peak (x_{off}, y_{off}) are at (57.09,184.19) as shown. To try to verify this alignment accuracy, both the reference and the radar images were up-sampled by a factor of 10, and then the two images matched by eye. The large size of the features as, determined by the angular and range resolution of the radar, made it difficult to judge the alignment even to this level of accuracy. The best fit that could be found had the reference image at a displacement of (571,1842) as shown in Figure 6.20.

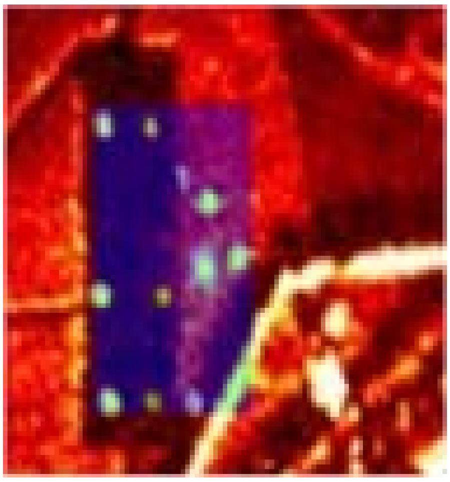


Figure 6.20: Manual placement of the reference image at a displacement of (571,1842) on images up-sampled by a factor of ten confirms a perfect correlation to a fraction of a pixel

Obviously the accuracy of this measurement would depend on the accuracy with which the reference image has been constructed as well as the relative positions of the selected features.

6.4.4. Effects of Reference Image Errors

In the previous example both the scale and the rotation of the radar and reference images are matched, but, in reality this is unlikely to be the case. The following section investigates the effect of some angular and scale misalignment.

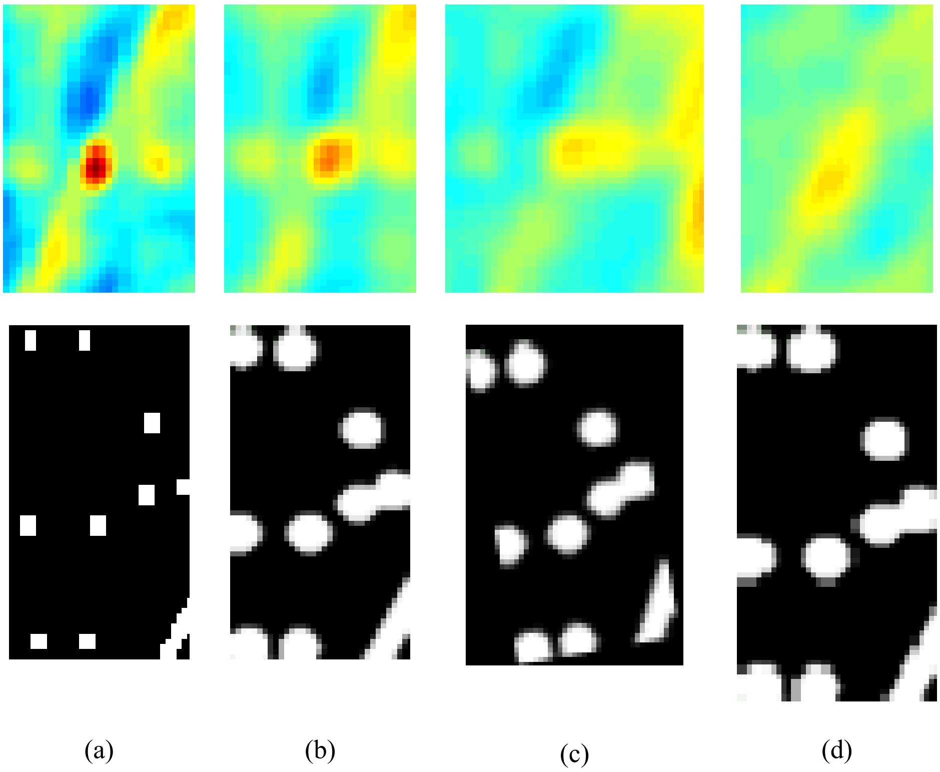


Figure 6.21: Correlation outputs for different reference input image template options for (a) standard unmodified template, (b) dilated, (c) dilated and rotated by 10° and (d) dilated and magnified by 120%

For the template shown in Figure 6.14b, the correlation process appears to be quite robust in rotation with a clear peak still visible if the reference image is rotated by 5° but once it has been rotated by 10° the peak is hardly discernable. The correlation process is also quite robust in scale with a clear peak still visible if the reference image scaled by 110%, but by a scale of 120% the peak has all but disappeared.

By dilating the point and line features in the reference image as shown in Figure 6.21, the basic accuracy of the correlation decreases slightly, but it is found that both the rotation and scale tolerance is increased, and the correlation peak is still clearly visible for a rotation of 10° and a scale of 120%.

An image rotation of 10° at a range of 3000m corresponds to a lateral displacement of 520m which is larger than the image and so would never occur in practice with this system. Similarly a scale error of 120% implies a range measurement error of this magnitude which is also very unlikely.

In conclusion, it would appear that the correlation method is accurate and robust for small errors in scale and rotation but that as these increase, the magnitude of the peak decreases considerably and may be difficult to detect. Large peaks that appear elsewhere in the correlation output can be mistaken for the correct peak if the initial position uncertainty is the size of the radar image. Other problems include the fact that the technique is computationally intensive and does not function well for distributed targets unless the reference image is extremely accurate.

6.4.5. Application for Autonomous Navigation

For this technique to be used operationally by an autonomous vehicle, good *a priori* information about the area of interest is required so that a sufficiently accurate reference-image template can be created. However, the area of the template can be very small compared to the size of the subsequently generated radar image. It should be as small as possible, while still avoiding ambiguity problems, to minimise computational overheads.

Correlation with a full one-square-kilometre radar image need only be undertaken once as the beam can be steered to scan only a small area on subsequent scans. This will increase the frame rate by reducing both the scan and correlation times significantly.

6.5. Ground-Based Path Planning and Navigation

The ground-based path planning and navigation process can be divided into four distinct processes as shown in Figure 6.22. At the highest level the path-planning algorithm uses GPS and INS to identify the direction in which the vehicle should travel to reach its final destination, or a waypoint. Maps of the region (if they exist) can be used to direct the vehicle around obvious physical obstacles including untraversable terrain and other no-go areas defined by fences or walls. The second- and third-level processes are similar insofar as they attempt to determine the traversability of a region. The second level uses long-range gate-limited radar images to identify obvious obstacles by using changes of ground reflectivity that may indicate features to be avoided. The third level performs a similar function, albeit at higher resolution because the range is shorter. However in addition to measuring reflectivity, this level can use both gate and beamwidth-limited images to determine the terrain slope. Finally at the fourth level, short-range sensors are used to scan the path ahead for obstacles including rocks, tree branches and ditches that may impede the AGV progress.

This structural hierarchy is becoming a more common approach both to robotic terrain perception applications [141] and in SAR target discrimination [103].

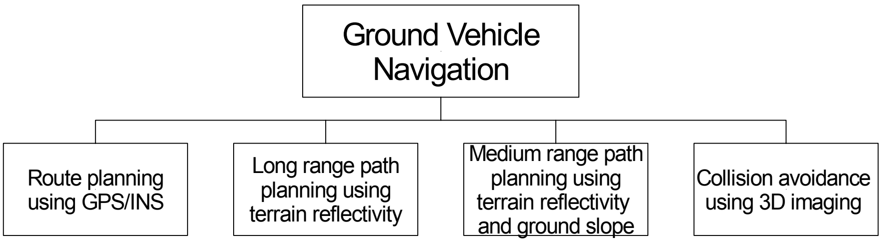
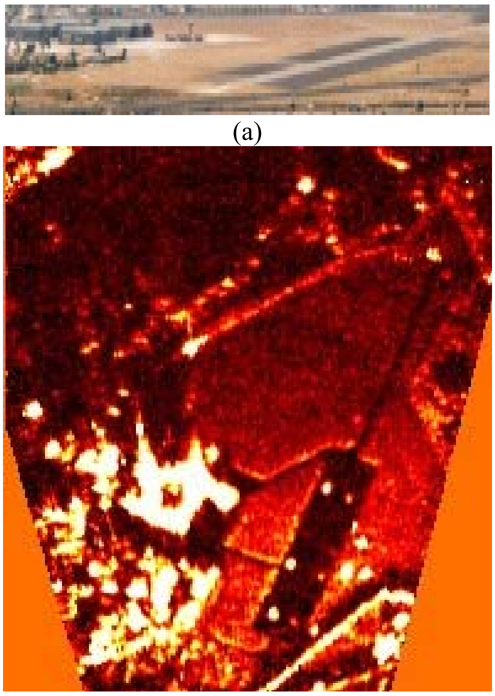


Figure 6.22: Hierarchy of route and path-planning processes required for AGV navigation

6.5.1. Advantages of using Radar Images

This section examines the use of long-range radar reflectivity images for third level processes. A range-gate-limited radar is used to produce ground reflectivity images from which some surface characteristic can be deduced. These images are superior to those made even by high-resolution video cameras if the grazing angle is very low because of their good resolution along the range axis.


(a)

(b)

Figure 6.23: Comparison between (a) a video image and (b) a radar image taken from the same position and covering the same range and angles. This shows that the radar down-range resolution is superior to that of the video image

In the example shown in Figure 6.23, the grazing angle from the radar to the ground varies between 2.67° and 1.3° for ranges between 1.5 and 3km and yet good quality images are obtained with a theoretical range-resolution of about 4m. In contrast, a 500 line video camera with an instantaneous field of view of 30° in azimuth and about 22° in elevation produces only 30 lines spanning the 1.5 to 3km range of interest, an effective range-resolution of about 50m.

For an AGV with a radar and camera mounted on a 2m mast such as the conceptual drawing shown in Figure 6.24, the grazing angle would be even smaller, and the resolution from the camera would be even worse than that calculated.

Figure 6.24: Conceptual drawing of an imaging radar mounted on the Argo AGV

It must be stressed that there will be significant differences between the images produced by the airborne radar, shown in Figure 6.23, and those that would be made from an AGV. In the first instance, though the grazing angle to the former is very low (≈1.3°), those from the latter may be even lower and may even be negative if the ground is falling away in front of the vehicle. In addition, from a 2m vantage point, even reasonably small objects including buildings and small trees cast long shadows and may even obscure sections of the image completely.

To some extent, in the terrain envisaged for the operation of an AGV, multiple overlapping images from different perspectives made as the vehicle moves, may counteract some of the shadowing problems. However very little can be done to remedy the problems of low grazing angles and a limited radar horizon due to the low vantage point.

From an image generation perspective, operating from a moving AGV will also be extremely challenging unless the scanner can be isolated from the violent and often unpredictable base motion that will occur when the vehicle is off road.

6.5.2. Multi-Level Terrain Map Description

The construction of a multi-level description of the terrain based on the outputs of all the available sensors as well as *a priori* knowledge of the terrain is illustrated in Figure 6.25. These maps are then used as input to the path-planning algorithms which will in turn direct the vehicle appropriately to traverse a specific region safely and quickly.

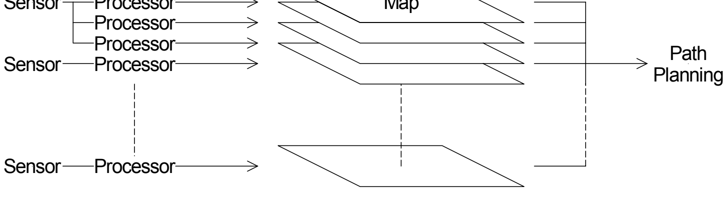


Figure 6.25: Structure required for the generation of a multi-level description of the environment used in ground vehicle path planning

Even using only a single radar sensor, different processes can extract intensity levels, discontinuities and textures to produce a rudimentary multi-level map. Gabor filters tuned for specific periodicity have been investigated by a number of groups both for the segmentation of radar images [16] and for general image processing applications [75], [49], [146]. These filters can be used to discriminate between homogenous and textured regions automatically and so would be capable of discriminating between treed and grassed regions.

Edge detection filters are generally not very effective when processing millimetre-wave radar images for the reasons discussed earlier in this chapter. However in high contrast situations, filters based on the Hough transform [104], Canny edge detector [53], Gabor filters [146] and deformable template techniques amongst others can be used to perform this function with some degree of confidence.

6.5.3. Terrain Map Interpretation

The difficulty with using radar-image information for autonomous navigation should not be underestimated. In this thesis only one of the operational modalities for a radar, as described in Figure 6.22, has been considered. This long-range imaging capability is used only for gross path-planning and traversability processes, while the finer obstacle avoidance functions proposed use higher resolution images generated by 3D radar or scanning laser systems.

In general, the path-planning process uses 2D radar image information to direct the vehicle using the following simple rules:

- Areas that exhibit a small amount of spatial variation in reflectivity probably represent grassland or small uniformly distributed scrubland (gorse, heather, bracken etc.). These areas are probably safe to traverse depending on the size of the vehicle.
- Areas that exhibit large spatial variations in reflectivity are probably trees or large bushes and are probably impassable.
- Edges with high reflectivity probably define borders of thick scrub if they are irregular, or walls of buildings or fences of they are linear and are definitely to be avoided.
- High reflectivity edges that cast shadows are walls, hedges or rows of trees and are also probably impassable.
- Abrupt changes in reflectivity are borders and should be avoided if possible because they may indicate changed conditions.
- Areas with uniformly low reflectivity and little spatial variation are probably paved areas, ditches, canals or rivers depending on the characteristics of their edges
    - Rivers will generally have irregular edges and so should be avoided
    - The opposite bank of ditches or canals will probably be visible as a diffuse region of high reflectivity and so should be avoided
    - The opposite edge of a paved area is sometimes defined by a line of high reflectivity if there is a corner present (curb or even high grass)

If the vehicle is already in a region of low reflectivity and it can still move, then the area is probably not water, and it would be wise to stay within such a region for as long as possible.

The inclusion of some *a priori* information such as the position of walls, fences, buildings and other non traversable obstacles that can be gleaned from aerial photographs, ordnance survey maps and other sources of intelligence adds a further map level for input into the path-planning process.

The application of some of these rules to the *a priori* no-go map and radar image of the terrain shown in Figure 6.26 is used to determine the effectiveness of the technique as it is applied for long-range images only. In an AGV navigation system, higher resolution short range images generated on the fly would be used to perform the slope estimation, obstacle avoidance and short range path planning functions.

6.6. Radar Reflectivity-Based Path Planning

In this example, the AGV starts at position A or B and has to navigate safely to position C. To make the problem more tractable, the radar image displayed in Figure 6.27 shows an artificial hole which has been placed in the perimeter wall to allow the vehicle to reach the end-point.

Figure 6.26: Aerial photograph of the terrain used in an autonomous path-planning exercise

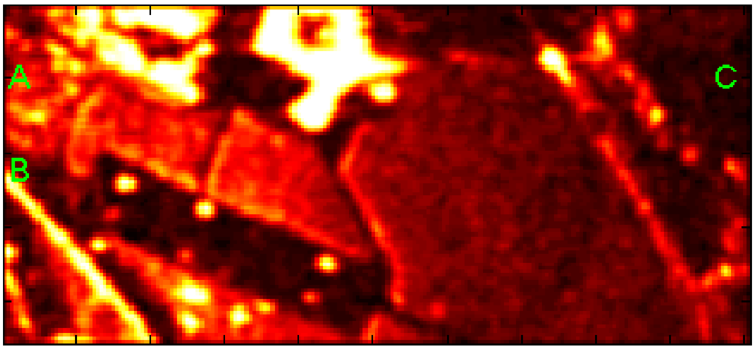


Figure 6.27: Filtered 2-D millimetre-wave image of the terrain

6.6.1. A Simplified Path-Planning Algorithm

As a precursor to developing the algorithms to reason about all of the features described in the previous paragraphs, a simple algorithm was developed to keep the AGV within a specific surface reflectivity band. To achieve this requires that a penalty be applied to the vehicle if it strays into regions of either lower or higher mean reflectivities.

Most path-planning techniques use occupancy grids with two, or at most three states (occupied, unoccupied and unknown) [79], [124], [82]. However such coarse quantisation is unsuitable for this application as the reflectivity map provides much more information about the environment and its traversability.

By converting the radar reflectivity map into a potential field, a gradient-descent technique [192] can be used to perform the path-planning function.

Because the radar-map reflectivity is encoded in 8-bits, a total of 256 possible reflectivities exist. However if the image is examined it can be seen that these vary on a pixel to pixel basis which would make the potential surface too rough to apply the gradient-descent technique, so the image is filtered using a 3×3 averaging low-pass function.

To transform the reflectivity surface into a repulsive potential-field, $U_{rep}(x,y)$ used by classical path-planning algorithms, any reflectivity values below the reflectivity at the start point are folded up into higher potential levels as shown in Figure 6.28.

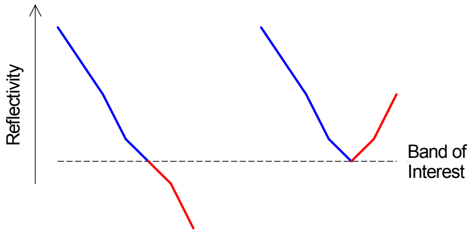


Figure 6.28: Folding the reflectivity map to place the vehicle band-of-interest at the minimum of the repulsive potential-field

For a vehicle starting at position B, which is already a low reflectivity point, there is not much change to the repulsive field image when this folding occurs as can be seen by the image in Figure 6.29.

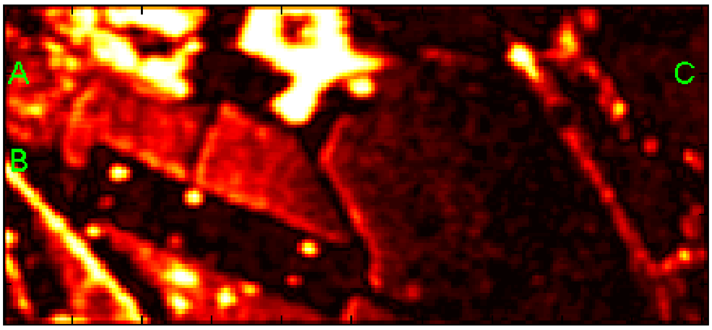


Figure 6.29: Repulsive potential-field map derived from reflectivity values, folded to produce a minimum at point B

As the objective is to guide the vehicle from the start point, B, to the finish, C, it is possible to introduce a potential-field sink at C at coordinates (x_{final},y_{final}). This produces an attractive field $U_{att}(x,y)$ with the potential increasing linearly with the distance from that point over the whole of the surface, a so-called conic well potential.

If $\rho(x,y)$ is the Euclidian distance to (x_{final},y_{final}) and K is the gradient, then the linear field can be defined

$$U_{att}(x,y) = K\rho(x,y). \tag{6.3}$$

This gradient field is added to the existing potential surface to produce a potential sum at each point on the surface

$$U(x,y) = U_{att}(x,y) + U_{rep}(x,y). \tag{6.4}$$

Guidance from point B to point C can then be achieved using a modified gradient-descent algorithm with some capability to escape from local minima. In this case, the negative gradient of U can be considered to be a force $F(x,y)$ acting on the vehicle to move it from A or B (x_{init},y_{init}) to the endpoint C (x_{final},y_{final}),

$$F(x,y) = -\nabla U(x,y) = -\nabla U_{att}(x,y) - \nabla U_{rep}(x,y)). \tag{6.5}$$

While it is possible to solve this equation for the attractive field, because the repulsive field is determined from the measured reflectivity map described earlier, it is more convenient to approximate the solution by searching the potential field $U(x,y)$ in the vicinity to find a minimum.

In formal terms the gradient descent based path-planning algorithm can be written as follows:

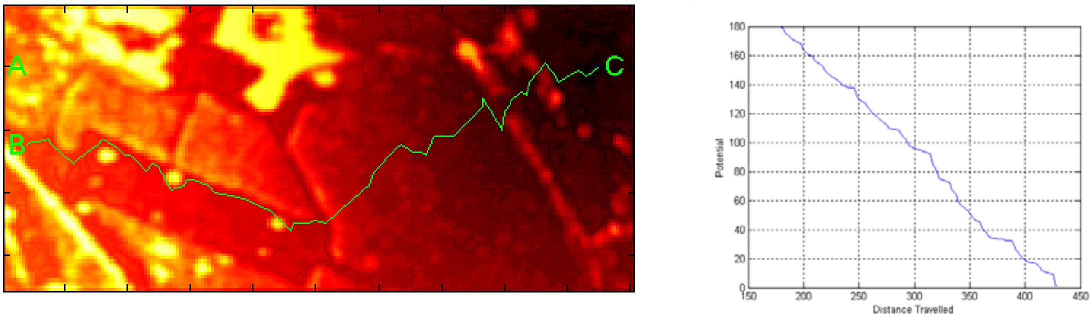
```
1.  $(x^0,y^0) \leftarrow (x_{init},y_{init}), i \leftarrow 0$
2.  IF $(x^i,y^i) \neq (x_{final},y_{final})$
        $(x^{i+1},y^{i+1}) \leftarrow (x^i,y^i) + \alpha^i \frac{F(x^i,y^i)}{|F(x^i,y^i)|}$
        $i \leftarrow i+1$
    ELSE return $<(x^0,y^0),(x^1,y^1),...(x_{final},y_{final})$
3   GO TO 2
```

In this algorithm, α^i is a scalar that determines the size of the step in the direction of the unit vector that points in the direction of the resultant force.

6.6.2. Basic Path-Planning Algorithm Interpretation

The first algorithm developed starts off by searching only the pixels adjacent to the occupied cell, a 3×3 area. This area increases with each iteration until a pixel is found at a lower potential and the vehicle starts to descend towards its goal. Because any intervening high potential pixels are ignored, this process is analogous to quantum-mechanical tunnelling. Once the vehicle has escaped from the local minimum, the search size is reset to the smallest search area where it remains until the vehicle gets stuck again.

One of the advantages of this algorithm is that it minimises the search area during normal operation and only widens it when necessary to escape from a local minimum. This is computationally extremely effective, but as can be seen from Figure 6.30 and Figure 6.31, can lead to some interesting, and not very appropriate behaviour.



(a) (b)

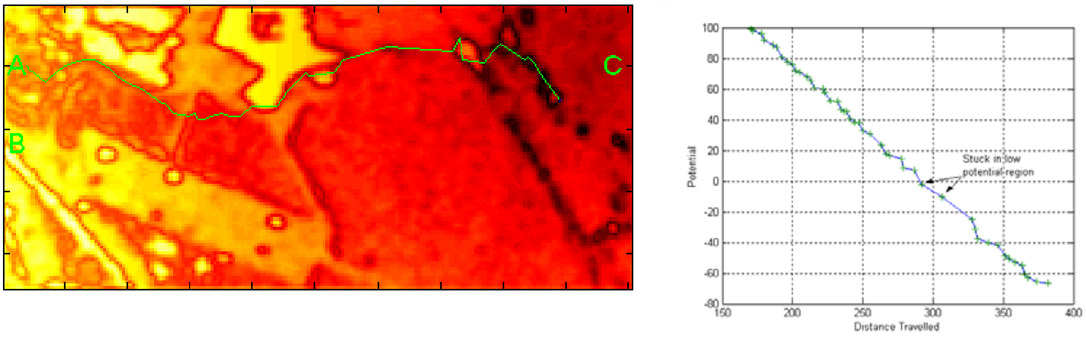
Figure 6.30: Simulation outputs of a modified gradient-descent algorithm with expanding search area and a sink gradient of 1 showing (a) the path traversed from B to C and (b) the potentials of the points passed through during the traverse

The actual route taken by the vehicle will be a function of both the magnitude of variations in the repulsive potential-field and the sink gradient. Figure 6.30 shows the path taken with a gradient of one. In this case it can be seen that the vehicle “tunnels” through the middle of a high repulsive field spot as it traverses along the edge of the runway.

If the sink gradient is made too small, and the end-point is at a high-field region, it is possible that the minimum will not fall on the end-point, but at some intermediate region, in which case no solution will be found and the vehicle will stop in this region. Other potential problems with this technique include regions with low field-potentials that may seduce the vehicle away from its goal. These are more common if the start-point occurs at a high potential such as that exhibited at point A.

In this case, both the low reflectivity areas, like the runway, and the high reflectivity areas, like the hangars, are mapped to high potential-fields, as is most of the plane through which the vehicle must pass. There are areas close to the end-point, C, where the potential is low because higher reflectivity targets have made up for the droop in measured reflectivity with range.

For a gradient of 1, the vehicle gets trapped in a minimum which is close to C as shown in Figure 6.31, and even increasing the gradient does not allow the vehicle to escape easily because the trap is too close to the end-point.



(a) (b)

Figure 6.31: Simulation outputs of a modified gradient-descent algorithm with expanding search area and a sink gradient of 1 showing (a) the path traversed from A to C that gets stuck in a local minimum and (b) the potentials of the points passed through during the traverse

A simple method to circumvent this problem is to re-calculate minimum-energy regions that correspond to the reflectivity region through which the vehicle is travelling at the time, and not to rely on a single reflectivity band for the complete traverse. This can easily be achieved by setting the vehicle waypoints to which it must navigate through regions of fairly homogenous clutter.

6.6.3. Path-Planning Algorithm with Waypoints

In the example shown in Figure 6.32, additional waypoints are selected at points D and E, (x_{way}, y_{way}), at which the whole process described in the previous section is re-initiated.

As with the previous example, local minima near the waypoints seduce the vehicle from its goal, so the modified technique utilised here is to incorporate a non-linear function that increases the gradient as the waypoint is approached. In this example the weighting function is proportional to the square-root of the distance from this point.

$$U_{att}(x,y) = K\sqrt{\rho(x,y)} \tag{6.6}$$

The main advantage of using this function is that, when the vehicle is far from the waypoint, the gradient is low, and it will be guided around most repulsive field peaks, but as it approaches the waypoint, no local minima will be sufficiently strong to seduce it away from its objective.

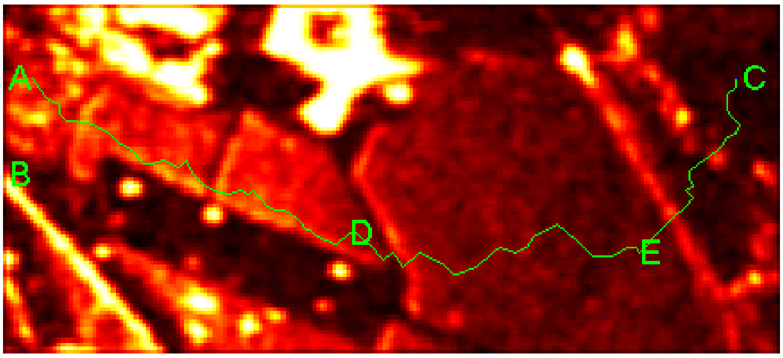


Figure 6.32: Simulation outputs of a modified gradient descent algorithm with expanding search area and a non-linear sink gradient showing the path from A to C via waypoints D and E

It can be seen from the path shown in Figure 6.32 that a combination of waypoints and non-linear gradient functions produces an effective guidance method that results in a reasonably direct path through the image. In addition, this path still follows a reasonably constant reflectivity contour.

This algorithm does not eliminate tunnelling through high potential barriers to the next low. This is behaviour which would be unsatisfactory in a real system.

Two alternative solutions are described in the following sections. In the first alternative, the vehicle examines the surrounding terrain and can move to a new position with a higher potential in an attempt to find a path. In the second, no-go areas are determined *a priori* and are encoded into another map of the area to eliminate tunnelling.

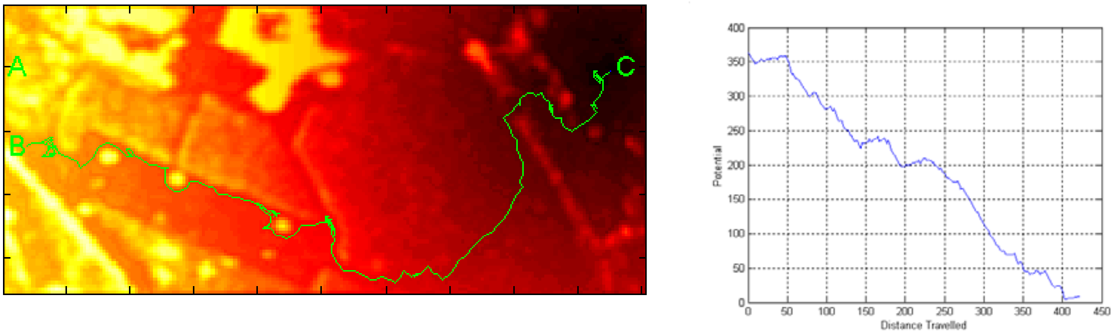
6.6.4. Path-Planning Algorithm Allowing Increases in Potential

In this algorithm, the search region is constrained to a small area, typically 3×3 or at most 5×5 pixels around the vehicle's present position. The vehicle then moves to the lowest potential point within that area excluding any points that it has already visited.

This algorithm overcomes the tunnelling behaviour exhibited by the previous algorithms while still allowing the vehicle to escape from local minima. This simple modification results in a range of apparently complex behaviours which are sensitive to the applied gradient, the smoothness of the surface being traversed (as determined by the low-pass filter coefficients), and the size of the local search area.

6.6.4.1. Gradient 2 and Search Area 5×5

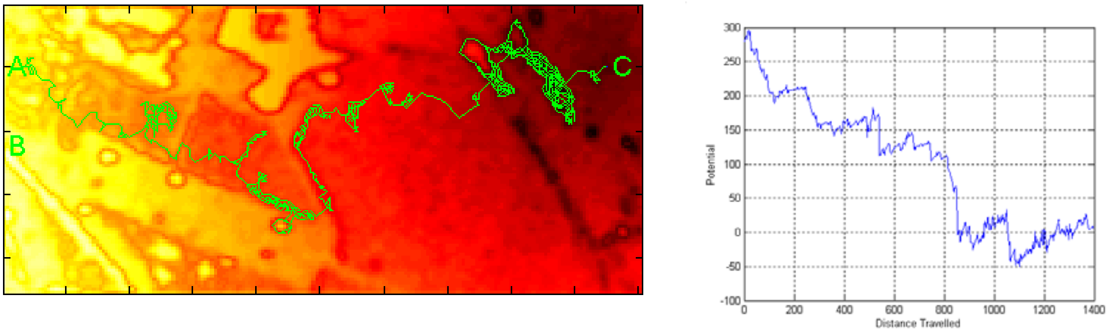
If the results of this simulation shown in Figure 6.33 are compared to those in the previous section it can be seen that under similar conditions, the total distance travelled is quite similar. In this case, however, where the vehicle approaches a higher potential region, it does not tunnel through but searches for a path around the obstruction.



(a) (b)

Figure 6.33: Simulation outputs of a modified gradient-descent algorithm with a fixed search area of 5×5 pixels, a facility to increase the potential and a sink gradient of 2 showing (a) the path traversed from B to C and (b) the potentials of the points passed through during the traverse

In the less homogenous example shown in Figure 6.34 where the vehicle starts out in a high-clutter region, it encounters many low potential areas where it remains for some time as it searches for the lowest potential escape route. It can be seen that the algorithm is sufficiently robust to extract the vehicle from these situations.

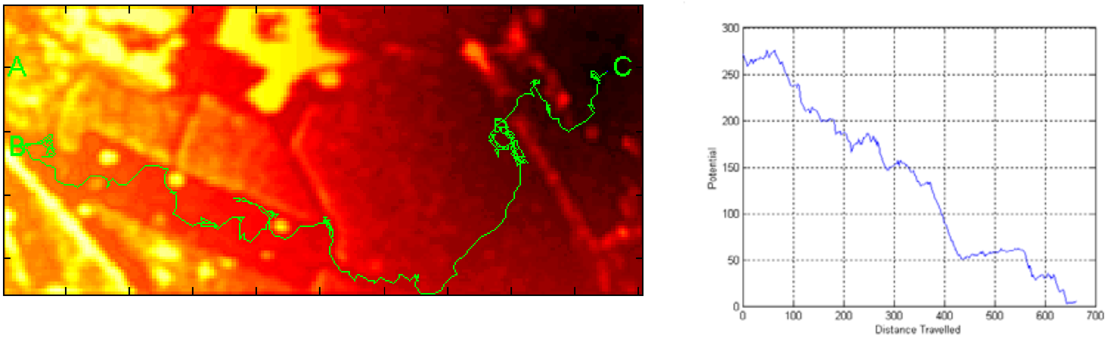


(a) (b)

Figure 6.34: Simulation outputs of a modified gradient descent algorithm with a fixed search area of 5×5 pixels, a facility to increase the potential and a sink gradient of 2 showing (a) the path traversed from A to C and (b) the potentials of the points passed through during the traverse

6.6.4.2. Gradient 1.5 and Search Area 5×5

If the gradient is decreased by a small margin as shown in Figure 6.35, then more local minima are encountered during the traverse which result in more local searches and hence a longer total distance travelled than that obtained for a larger gradient.

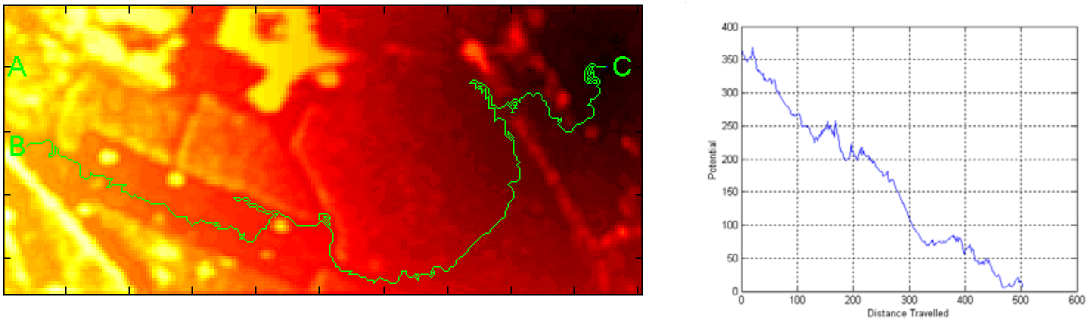


(a) (b)

Figure 6.35: Simulation outputs of a modified gradient-descent algorithm with a fixed search area of 5×5 pixels, a facility to increase the potential and a sink gradient of 1.5 showing (a) the path traversed from B to C and (b) the potentials of the points passed through

6.6.4.3. Gradient 2 and Search Area 3×3

If the gradient remains the same, but the search area is decreased to include only the pixels adjacent to the current position, then the total distance travelled increases as shown in Figure 6.36. Problems can arise because of the way the algorithm prohibits access to regions already visited as this can cut off the route to the end point.

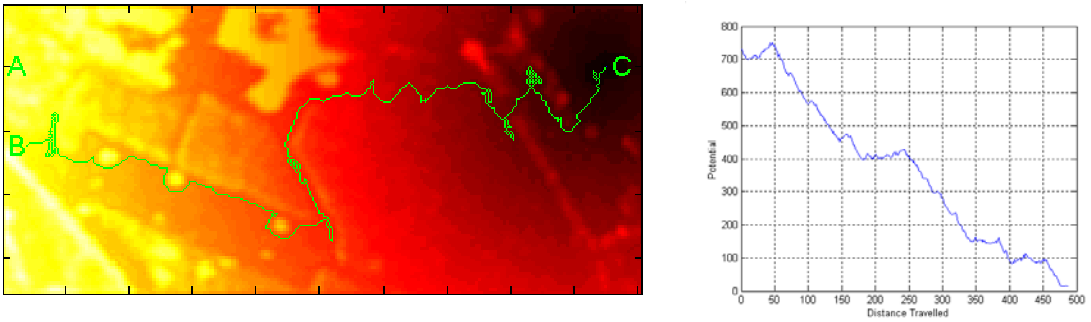


(a) (b)

Figure 6.36: Simulation outputs of a modified gradient-descent algorithm with a fixed search area of 3×3 pixels, a facility to increase the potential and a sink gradient of 2 showing (a) the path traversed from B to C and (b) the potentials of the points passed through during the traverse

6.6.4.4. Gradient 4 and Search Area 3×3

Increasing the gradient is not always a solution as shown in Figure 6.37 in which the path followed changes completely but the total distance travelled remains the same.



(a) (b)

Figure 6.37: Simulation outputs of modified gradient descent algorithm with a fixed search area of 3×3 pixels, a facility to increase the potential and a sink gradient of 4 showing (a) the path traversed from B to C and (b) the potentials of the points during the traverse

These examples have shown the rather convoluted paths generated by this algorithm to illustrate the process of path generation. In reality, when the complete path is available, a smoothing algorithm will generate a route that removes loops and filters the path.

6.6.4.5. Algorithm to Accommodate No-Go Areas

To ensure that the vehicle never attempts to pass through any impenetrable barriers, a second map of the area which defines no-go areas is defined as shown in Figure 6.38.

Figure 6.38: Map of the airfield compiled from aerial photographs and an ordnance survey map showing some of the no-go areas. The region of interest lies within the rectangular box

Special rules could be developed to cater for this additional information, but the simplest method is to let these barriers generate their own repulsive potential-fields [62] and incorporate them into the existing map.

6.7. Conclusions

This chapter has shown that good quality radar images can be used effectively for both navigation and guidance applications. It starts by looking at the image generation process which was found to produce particularly good images from the air. It shows that these images can be used to augment video images, and in some cases return additional information invisible in the optical image.

In Section 6.3, feature-based navigation techniques are used with very few features to produce results better than those obtained using GPS. In flight trials this technique confirms the premise that a few widely spaced point features can be used to navigate with an accuracy that is better than the resolution of radar system.

This feature-based technique is expanded to use 2-D correlation techniques also using very few features placed within a reference-image template. It is shown that sub-pixel navigation accuracies are possible even if the reference image has both scale and rotation errors.

Finally Section 6.6 shows that simple gradient-descent algorithms can be used with some modifications to guide an AGV through terrain mapped by the radar using variations in reflectivity only. It also shows that it would be practical to use multiple maps encoded with different characteristics of the same region to improve the process. Though the algorithms developed here are used for both short and long-range path planning, in reality any guidance will be augmented by short range sensors on the vehicle to overcome any ambiguity in the interpretation of the reflectivity.

Chapter 7.

Summary, Contributions and Future Work

7.1. Summary and Contributions

This thesis has addressed a problem with the range capability of sensors used for autonomous navigation of a robotic vehicle. It has proposed that a vehicle must be capable of sensing the surrounding environment over a broad range of scales in order to achieve full autonomy and that existing sensors do not have the long-range performance to achieve this goal.

The process of sensing an environment involves sweeping an appropriately shaped radar beam over the terrain, measuring the range to any reflecting surface and then building up an image of the environment based on the position of these surfaces and the intensity of the returns.

Objects in the immediate vicinity of the vehicle must be sensed with sufficient resolution that the system can determine whether the region is traversable or not. This is a non-trivial task as was discovered during the recent "DARPA Challenge". At the next level, the sensor range and resolution must be sufficiently good to allow for short-range path-planning to occur. For an AGV, this involves identifying individual trees, poles, fences and the ground slope. At long range the sensor must be capable of identifying general terrain characteristics, open as opposed to heavily-wooded regions and the existence of gullies and other impassable features.

The primary purpose of this long-range operational mode is to produce an overview of the environment that would save the vehicle entering and exploring every possible path as it navigates toward its objective. The actual scale of the information provided by the sensor depends on whether it is ground or air-borne, and on its size. In addition, the scale is also dependent on the distance from the objective.

This thesis considers the sensor requirements for the long-range imaging mode of an all-terrain AGV or a moderately sized UAV with a payload capacity of about 50kg. It is concerned with both the radar and the navigation aspects of this problem and examines the historical foundations of both and summarises the progress in the field up to the present.

Chapters 2 and 3 set the scene by discussing the techniques commonly used to produce radar images and examining the effects of the environment on the quality of these images. A summary of the state-of-the-art in radar development for autonomous guidance and radar-vision applications completes this background.

The contributions of these chapters include:

- Very low grazing-angle radar reflectivity measurements of grass and runways at 94GHz. These dispute some of the accepted models of radar reflectivity.
- Spatial decorrelation and pixel integration on the reflectivity distribution of millimetre-wave radar images. This leads to improved image quality and superior performance of feature extraction techniques.
- Effects and probability of mutual interference between two radars with overlapping frequency-bands operating in the same vicinity. This analysis shows that the interference will be common, and will have a major impact on the performance of the affected radar systems unless the appropriate precautions are taken.

Chapter 4 examines various radar modulations in the context of their viability as sensors for the long-range navigation application. It concludes that FMICW is the most practical and lowest cost of all of the possibilities considered.

The contributions of this chapter include:

- A novel and effective method to determine the gradient of a nearby hillside using a conically scanned range tracker.
- A function that optimises the range resolution at a particular range given a VCO with a quadratic linearity error.
- A simple method to determine the effects of transmitter leakage on the noise floor of an FMCW radar.
- An analysis of the effects of phase-noise around target echoes.

From a radar perspective, the main focus of this thesis is the analysis of various possible implementations of the FMICW technique and the derivation of a hybrid FMICW/Stretch system which is then implemented as part of the development of an airborne sensor. A comparison is made between the measured performance of the radar, and that predicted which verifies the received signal amplitude but finds that the range resolution is worse than expected.

Constraints on the speed of available processors and ADCs forces an implementation of the spectral analysis algorithm based on an analog coarse-gating process followed by digitisation and fine-gating using an FFT. The blocks of range gates are then combined to form a continuous spectrum.

The contributions of this section include:

- The evaluation of various FMICW modulation options in terms of their resolution, their suitability for an imaging application and their processing requirements.
- The design and implementation of a robust dual-loop lineariser that maintains a constant chirp-slope as well as a stable centre frequency.
- The design and implementation of a robust FMICW transceiver operating at 94GHz.
- The design and implementation of a wide-bandwidth spectrum analyser based on a stepped down-converter followed by a fine-resolution complex-FFT process.

A surface-reflectivity image is formed by sweeping the beam across the ground, and combining the returns spatially using the measured angle and range to each return. Chapter 6 develops and evaluates two techniques that use these images for autonomous navigation. Both the feature-based and the full 2-D image-based correlation techniques determine the platform position with sub-pixel accuracy.

The final section considers the navigation problem from a path-planning perspective and evaluates some modified gradient descent methods that allow an AGV to use a radar image to plan a safe route to an objective.

The contributions of this last section include:

- The use of features extracted from a millimetre-wave radar image in conjunction with those synthesised using *a priori* knowledge of the area, to perform feature-based navigation.
- The use of a raw millimetre-wave radar reflectivity image in conjunction with an image synthesized using *a priori* knowledge of the area to perform 2-D correlation-based navigation.
- The use of a millimetre-wave radar reflectivity image and the modified gradient descent technique to plan the path of a ground vehicle across the imaged terrain.

7.2. Future Work

7.2.1. Radar Systems

This thesis has discussed the development and implementation of a stabilised 94GHz FMICW front-end and a hybrid FMICW/Stretch process to produce high-resolution radar images from a helicopter. However, most of the navigation and guidance work conducted at the ACFR is based on smaller airborne carriers or ground-based vehicles.

To achieve the required angular resolution for long-range operation from the available airborne carriers only SAR or SLAR techniques would produce an appropriately narrow azimuth beamwidth.

The following section describes a number of projects that are underway which aim to address these issues:

The development of a light-weight 600×20mm microstrip antenna and an FMCW or FMICW radar to produce side-looking images. An investigation is underway to determine whether it is possible to perform image integration without mechanical stabilisation using available strap-down inertial and high-speed signal processing hardware. This approach will allow the use of light-weight hardware, similar to that used in the helicopter application, to be used in a UAV with a limited payload.

As the VSWR of a non-scanned array remains reasonably constant, it might be possible to dispense with temporal separation of the transmitted and received signals and use high-power FMCW. An investigation is being conducted into the feasibility of implementing light-weight robust reflected-power-cancellation techniques to achieve these ends.

For ground-based applications, where mass is not an issue, a more sophisticated front-end with less constrained signal processing is under development. In this case a coherent, bistatic 94GHz unit has been developed as shown in Figure 7.1. This will be used with a 1GHz 12-bit arbitrary waveform generator as a signal source and a 1GHz ADC to digitise the received echo.

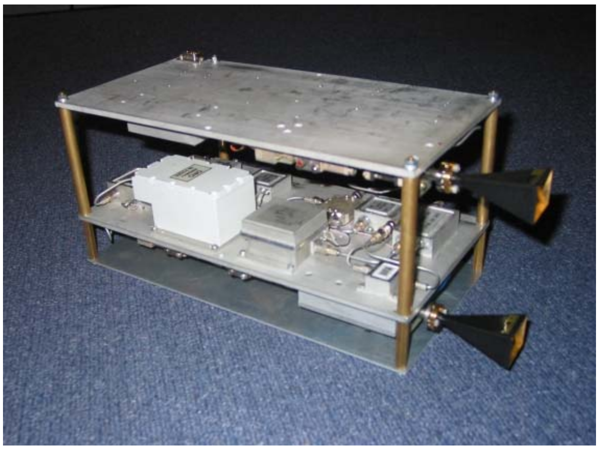


Figure 7.1: Coherent 94GHz front end developed for ground-based radar imaging

Options under consideration include the use of a pair of the microstrip antennas for isolation, with the radar configured for FMCW operation, or the use of a PIN-switch, a circulator and a single spinning-grating antenna (see Figure 7.2) with the radar configured for FMICW operation. The advantage of the latter configuration is that the antenna produces a rapid azimuth scan of the fan-beam without having to move the radar hardware.

The availability of high-speed programmable signal-processing hardware facilitates the investigation of novel modulation-techniques. A programme is underway to investigate the feasibility of using random or pseudo-random modulation for high-resolution radar applications. This investigation is particularly concerned with the elimination of mutual-interference which is an issue where the frequency bands of wide bandwidth radar systems overlap.

Figure 7.2: Spinning-grating antenna used to achieve a fast azimuth scan rate for range-gate limited radar images

To facilitate the development of radar units at 77GHz, a reconfigurable module shown in Figure 7.3 has been developed that allows the construction of radars for specific applications. For example, a single module with a VCO and an antenna produces an un-linearised FMCW or pulsed-FM radar. A pair of modules can be configured as a closed-loop system and a third can be used to add an orthogonal-polarisation channel. All of the 77GHz development undertaken at the ACFR uses these modules.

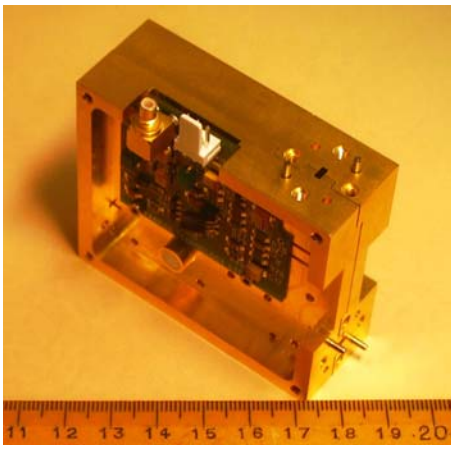


Figure 7.3: Reconfigurable module used for the development of 77GHz radar systems

7.2.2. Processing and Analysis

For both the airborne and ground-based applications, digitisation will occur much higher up the processing chain. With sample rates of 1GHz or higher, it is possible to dispense with the dual-step gating process discussed in this thesis and still achieve a realistic signal-bandwidth of 400MHz. This ability in conjunction with a programmable waveform-generator will form a versatile and powerful experimental radar system.

As an extension to the one-shot navigation correction processes discussed in this thesis, it would be advantageous to perform this process repeatedly with each new image acquired. This could be made particularly accurate as the coarse corrections possible with long-range operation are improved using finer corrections obtained from short-range high-resolution sensors.

In the absence of any *a priori* information, the process of identifying natural features as they come into view, and then tracking them from one observation to the next, can be used to navigate and to construct a map. This is a process known as Simultaneous Localisation and Mapping (SLAM), and it has been well investigated for features observed from sensors with similar modalities and resolutions. Of interest, in the context of this thesis, would be the association of features based on sensors with vastly different resolutions, and observed from different perspectives.

Kernel-based feature detection, identification and tracking methods could be applied to the radar data to improve overall navigation performance.

In addition to using only features based on differences in reflectivity, it would be interesting to extract and use some of their other characteristics. For example, the target polarisation-state can be used to determine whether a feature is natural or manmade, and fine Doppler-spectra can be used to determine whether an object is moving. The latter is an effective way to differentiate between a large rock and a wind-blown shrub, a distinction that is critical for an AGV.

The use of multi-dimensional information from the environment can be a powerful tool in the identification and characterisation of target types. An interesting project would be to use the reflectivity, polarisation and Doppler information from millimetre-wave radars as inputs to the multi-dimensional feature characterisation algorithms currently used to process visible images.

As part of the feature characterisation work undertaken at the ACFR, an investigation is underway to characterise the millimetre-wave radar signatures of trees in terms of backscatter and penetration. This work should help with ground traversability issues. Figure 7.4 shows a processed radar image of a row of trees.

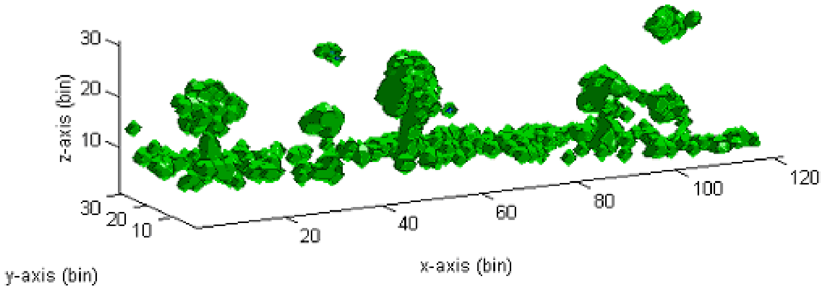


Figure 7.4: 3-D radar reflectivity image of a row of trees

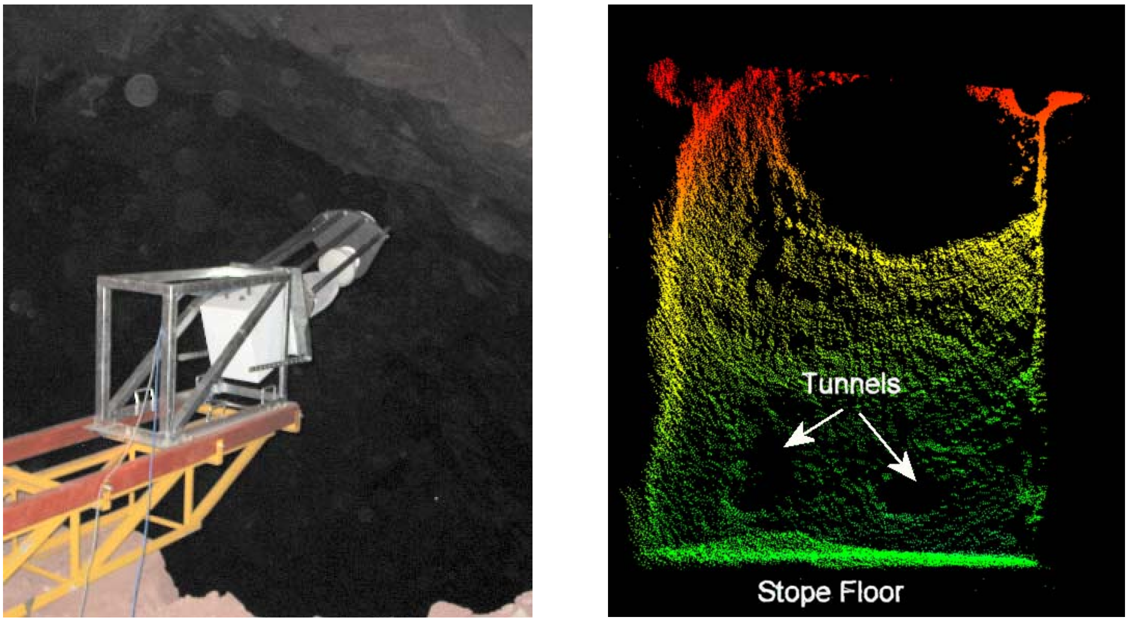
7.2.3. Applications

Existing applications of the classical wide-bandwidth FMCW processing techniques discussed in the thesis include a miniature mirror-scan radar mounted in a UAV. The system shown in Figure 7.5 consists of a mirror scan 77GHz radar mounted in 2-axis gimbals. A program is underway to use both artificial and natural features extracted from the radar data for airborne SLAM research.

Figure 7.5: Miniature 77GHz radar and mirror scanner mounted on gimbals in the nose of the Brumby Mk-II

The use of radar in mines is a reasonably new field driven by the requirement to see through dust, and it in these applications that much of the current work is undertaken.

Radar units based on the 77GHz reconfigurable modules have been used with good results in a number of mine-visualisation applications. One of the most successful has been to use the radar to map the interiors of vapour-filled underground cavities called stopes. Figure 7.6 shows the radar installation and an example of the data produced. This radar technology has also been used to produce images of the bench for open-cut mining in an ongoing program to improve dragline performance.



(a) (b)

Figure 7.6: Underground application showing (a) the installation of the 77GHz radar with a 3-D mirror scanner and (b) a point-cloud image of an underground stope made through vapour and dust while the stope was being refilled

A number of new high-speed applications is under development. A bucket-fill profiling radar for use by large mining shovels is one, and the visualisation radar for the Argo AGV is another. These radar systems operate at 94GHz and with their associated scanners will be capable of producing up to 15 image frames per second.

7.2.4. Summary

This thesis has been concerned with the implementation of a long-range millimetre-wave radar for use in autonomous guidance and navigation applications. As part of this process an interrupted FMCW radar has been designed and built, and the images produced by the radar have been used to evaluate a number of autonomous navigation algorithms. This implementation has extended the sensing horizon available to autonomous vehicles by an order of magnitude from 300m to 3km. In turn this will enable enhanced capability and wider future application of these systems, facilitating the introduction of fully autonomous vehicles capable of successfully traversing unstructured outdoor environments.