Synthetic Aperture Imaging Algorithms: with application to wide bandwidth sonar

David W. Hawkins, B.Sc.

A thesis presented for the degree of
Doctor of Philosophy
in
Electrical and Electronic Engineering
at the
University of Canterbury,
Christchurch, New Zealand.

October 1996

ABSTRACT

This thesis contains the complete end-to-end simulation, development, implementation, and calibration of the wide bandwidth, low-Q, Kiwi-SAS synthetic aperture sonar (SAS). Through the use of a very stable towfish, a new novel wide bandwidth transducer design, and autofocus procedures, high-resolution diffraction limited imagery is produced. As a complete system calibration was performed, this diffraction limited imagery is not only geometrically calibrated, it is also calibrated for target cross-section or target strength estimation. Is is important to note that the diffraction limited images are formed *without* access to any form of inertial measurement information.

Previous investigations applying the synthetic aperture technique to sonar have developed processors based on exact, but inefficient, spatial-temporal domain time-delay and sum beamforming algorithms, or they have performed equivalent operations in the frequency domain using fast-correlation techniques (via the fast Fourier transform (FFT)). In this thesis, the algorithms used in the generation of synthetic aperture radar (SAR) images are derived in their wide bandwidth forms and it is shown that these more efficient algorithms can be used to form diffraction limited SAS images.

Several new algorithms are developed; accelerated chirp scaling algorithm represents an efficient method for processing synthetic aperture data, while modified phase gradient autofocus and a low-Q autofocus routine based on prominent point processing are used to focus both simulated and real target data that has been corrupted by known and unknown motion or medium propagation errors.

ACKNOWLEDGEMENTS

vi ACKNOWLEDGEMENTS

PREFACE

This thesis started out in 1993 as a masters project to build a new wide bandwidth synthetic aperture sonar system after the loss, at sea, of the system used by Mike Hayes. The system described in this thesis is called the Kiwi-SAS. At the end of 1993 the masters was transferred to a Ph.D., with the focus of the thesis changing to an investigation of the processing algorithms used to produce high resolution synthetic aperture radar (SAR) images and their applicability to wide bandwidth sonar. The investigation of these algorithms encompasses the main strip-map and spotlight mode imaging algorithms, autofocus algorithms, and post-processing operations such as calibration and multi-look processing.

The thesis is written in eight chapters. Chapter 1 begins with an overview of SAR imaging and ends with a discussion on the contributions of this thesis and an overview of the thesis organisation. A great deal of effort has gone in to developing a clear mathematical notation in this thesis and anyone familiar with signal processing and Fourier theory should be able follow the development of the algorithms.

Papers prepared during the course of this thesis are listed below in the order of presentation.

Hawkins, D.W. and Gough, P.T., 'Practical difficulties in applying the synthetic aperture technique to underwater sonar', *Image and Vision Computing, New Zealand*, August 1995.

Hawkins, D.W. and Gough, P.T., 'Recent sea trials of a synthetic aperture sonar', *Proceedings of the Institute of Acoustics*, Vol. 17, No. 8, December 1995.

Hawkins, D.W. and Gough, P.T., 'Multiresonance design of a Tonpilz transducer using the Finite Element Method', *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. 43, No. 5, September 1996.

Gough, P.T. and Hawkins, D.W., 'Image processing of synthetic aperture sonar recordings',30th *Asilomar Conference on Signals, Systems, and Computers*, November 1996.

Gough, P.T. and Hawkins, D.W., 'A unified framework for modern synthetic aperture imaging algorithms', accepted for publication in *The International Journal of Imaging Systems and Technology*, December 1996.

Gough, P.T. and Hawkins, D.W., 'Imaging algorithms for a synthetic aperture sonar: minimizing

the effects of aperture errors and aperture undersampling', accepted for publication in *IEEE Journal of Oceanic Engineering*, 1996.

Hawkins, D.W. and Gough, P.T., 'Efficient SAS image-reconstruction algorithms', to be presented at the Acoustical Society of America conference, College Park, Pensylvania, 16th-20th June 1997.

Hawkins, D.W. and Gough, P.T., 'An accelerated chirp scaling algorithm for synthetic aperture imaging', to be presented at the International Geoscience and Remote Sensing Symposium, Singapore International Convention and Exhibition Center, Singapore, 4-8th August 1997.

CONTENTS

Chapter 1

INTRODUCTION

This chapter provides an overview of synthetic aperture imaging from its origins in radar and astronomy where it is known as synthetic aperture radar (SAR) and inverse SAR (ISAR) to its application to sonar where it is known as synthetic aperture sonar (SAS). The appropriate background reading is indicated and an historical review is made of the major steps in the development of the synthetic aperture technique and in its application to both the radar and sonar fields. The chapter concludes with the contributions and organisation of this thesis.

1.1 IMAGING GEOMETRY AND TERMINOLOGY

Although the reader of this thesis is assumed to have a reasonable knowledge of the synthetic aperture technique, i.e., be familiar with the references recommended under Section 1.5, the imaging geometry and basic terminology used throughout this thesis is presented here so as to avoid confusion.

Figure 1.1 shows the general three-dimensional (3-D) imaging geometry of a typical real or synthetic aperture system. The platform (aircraft or spacecraft for SAR systems, towfish for SAS systems) travels along the u-axis which is commonly referred to as either the along-track, azimuth, or cross-range direction. In the case of a side-looking strip-map synthetic aperture system operating at broadside (zero squint angle), pulses are transmitted in a direction perpendicular to the platform path. The plane scribed by the boresight of the aperture is called the slant-range plane in slant-range coordinates of u and r. The ground range coordinate x is related to r through the depression angle of the synthetic aperture system. The final image along-track dimension y is parallel to the platform path u. Throughout this thesis, the synthetic aperture system models are developed in the (x, y)-domain. This development assumes that the platform is travelling in the same plane as the object being imaged. Section 4.6 discusses the transformation from the actual imaging domain in slant-range (r, y) coordinates to the ground-plane (x, y) coordinates.

The production of the final image from the raw data is generally referred to as the inversion of
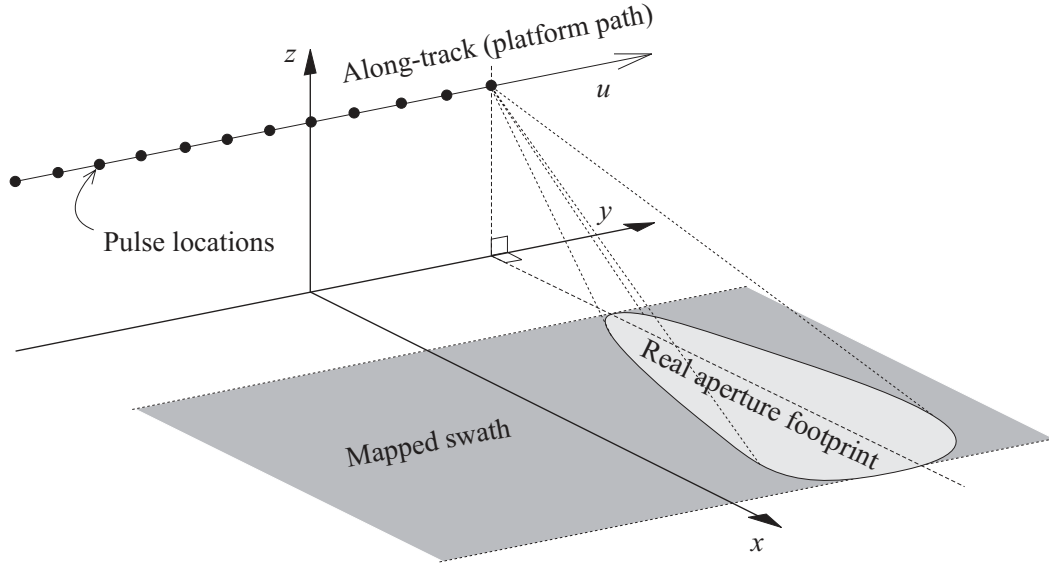
Figure 1.1 Basic imaging geometry appropriate for broadside (zero squint angle) real and synthetic aperture imaging (refer to the text for details).

the raw data or the synthetic aperture reconstruction. However, these terms are synonymous with matched filtering, correlation processing, or focusing. Each term has its origin in the interpretation of the inversion algorithm under consideration. The inversion schemes available for processing synthetic aperture data are presented in Chapter 4.

1.2   RANGE RESOLUTION

In real and synthetic aperture imaging systems, high range resolution is obtained by transmitting dispersed large time-bandwidth pulses that are compressed on reception using standard techniques known as *matched filtering* or *pulse compression*, or in some cases *deramp processing*. Each of these range resolving techniques is reviewed in Chapter 2. For a signal of bandwidth B_c and electromagnetic or acoustic wave propagation speed c, the achievable range resolution is typically referred to as

$$\delta x_{3\text{dB}} = \alpha_w \frac{c}{2B_c}, \tag{1.1}$$

where resolution is defined as the 3dB width of the range compressed pulse. The term α_w is a constant that is used throughout this thesis to reflect the effect of any weighting or windowing function that may be used to reduce range sidelobes. For example, α_w=0.88 for no (rectangular) weighting, α_w=1.30 for Hamming weighting [71].

1.3 REAL APERTURE IMAGING

In real aperture imaging, a platform (eg., aircraft or towfish) containing a moderately large real aperture (antenna) travels along a rectilinear path in the along-track direction and periodically transmits a pulse at an angle that is perpendicular to the platform path. This orientation of the radar or sonar system is termed side-looking, so these systems are typically referred to as *side-looking radars* (SLR) and sidescan or *side-looking sonars* (SLS) [143, 156]. These systems produce *strip-map* images. A strip-map image is built-up as follows; the imaging system operates such that the echoes from the current pulse are received before the next pulse is transmitted. As these echoes are received, they are demodulated, pulse compressed, and detected (i.e., only the magnitude information is retained). Each detected pulse produces a *range line* of the real aperture image. As the platform moves, these range lines are displayed next to each other at pixel spacings that scale relative to the along-track spacing of the pulses, i.e., $\Delta_u = v_p \tau_{\mathrm{rep}}$, where v_p is the platform velocity and τ_{rep} is the pulse repetition period. The final image is essentially a raster scan of a strip of the earth or sea floor, hence the name 'strip-map image'. Real aperture imaging is a *non-coherent* imaging technique, i.e., the *phase* of each echo is discarded; synthetic aperture imaging is a *coherent* imaging technique that exploits the extra information available in the phase of the real aperture data.

Chapter 3 shows that there exists a Fourier transform relationship between a real apertures' illumination (energizing) pattern and its radiation pattern. The radiation pattern of any dimension (width or length) of an aperture has an angular dependence that is referred to as the *beam pattern* of the aperture. Beam patterns are frequency dependent and have *beam widths* given by the 3dB response of their *main lobes*:

$$\theta_{\mathrm{3dB}} = \alpha_w \frac{\lambda}{D} = \alpha_w \frac{c}{fD}, \tag{1.2}$$

where D is the length of the aperture, and λ and f are the wavelength and frequency of the signal that the aperture is transmitting or receiving. The term α_w is a constant reflecting the main lobe widening due to weighting or *apodisation* of the aperture illumination function.

The constant angular response of the radiation pattern means that real aperture images have a *range variant* along-track resolution given by

$$\delta y_{\mathrm{3dB}}^{\mathrm{real}} = x\theta_{\mathrm{3dB}} = \alpha_w \frac{cx}{fD}. \tag{1.3}$$

This relationship also indicates that high resolution imaging requires the use of long apertures operating at high frequencies. The dependence of the along-track resolution in (1.3) and the range resolution in (1.1) on the wave propagation speed c enables real aperture sonars to achieve much higher along-track and range resolution than their real aperture radar counterparts.

The pulsed operation of both real and synthetic aperture systems leads to a limit on the *swath width* they are able to image without ambiguity [32]:

$$X_s = R_{\max} - R_{\min} = \frac{c\tau_{\mathrm{rep}}}{2} = \frac{c}{2 \cdot \mathrm{PRF}}, \tag{1.4}$$

where X_s is the swath width imaged, $R_{\min}$ and $R_{\max}$ are the minimum and maximum ranges, and $\mathrm{PRF} = 1/\tau_{\mathrm{rep}}$ is the pulse repetition frequency. Chapter 5 details the range and along-track ambiguities of real and synthetic aperture systems.

1.4   SYNTHETIC APERTURE IMAGING

The coherent nature of synthetic aperture imaging means that focused images can be formed from platform data that has been collected over an almost arbitrary path [106, 146]. The two main modes of synthetic aperture imaging discussed in this thesis are *strip-map mode* and *spotlight mode*. Other techniques such as inverse synthetic aperture radar (ISAR) and bistatic SAR are discussed in the references indicated in Section 1.5. Strip-map synthetic aperture systems operate in essentially the same manner as real aperture systems, except that the data received is stored or processed *coherently*. *Broadside operation* describes the situation where the main lobe of the radiation pattern points at an angle that is perpendicular to the platform path. Non-perpendicular operation is termed *squint-mode* and is only briefly described within this thesis. To synthesize the effect of a large aperture, the received echoes from the imaged target field are coherently integrated (summed) in an appropriate manner to produce an image that has an along-track resolution that is *independent* of range and wavelength, a result contrary to that of real aperture imaging. In spotlight mode, the real aperture is continuously steered or slewed such that it always illuminates the same ground patch. The continuous illumination of the same scene allows an image to be produced that has an along-track resolution that exceeds that achievable from a strip-map system. However, this increased resolution comes at the expense of a reduced area coverage. In military applications, the strip-map mode is typically used for mapping and reconnaissance, while the spotlight mode is used for weapon delivery and navigation system updating [13]. As with real aperture imaging, the pulsed method of image generation leads to range and along-track ambiguity constraints. These constraints are described in Chapters 4 and 5.

The key feature of the synthetic aperture technique is using the information introduced in the along-track direction due to modulation of the received signal by the relative motion of the platform to the target. In the SAR literature this modulation is often referred to as the targets' *phase history*. Many of the radar analyses of the phenomena refer to it misleadingly (as opposed to incorrectly) as a Doppler modulation, then use what is known as the *stop-start* assumption to determine the relevant mathematics. The stop-start assumption assumes that once each pulse is transmitted, and all echoes

have been received, that the platform instantaneously moves to the next along-track sampling location and the process is repeated. The stop-start assumption effectively removes movement from the system, therefore no *temporal* Doppler effect can exist. The correct term for the spectral content produced by the phase modulation in along-track is *spatial* Doppler, i.e., it is the rate of change of the relative platform-target distance that causes the *spatial* modulation of the return signal. The manipulation of the spatial bandwidth produced by this *geometrically* induced modulation is the essence of the synthetic aperture technique. True temporal Doppler effects cause a minor modulation within each transmitted pulse, however, for both radar and sonar this modulation can be ignored (see Sections 2.7 and 2.8). Given this brief overview of the (possibly) confusing aspects involved in analysing synthetic aperture theory, the following mathematics parallel the usual (basic) SAR development of the synthetic aperture technique. Chapter 4 presents a more detailed development of spatial and temporal based strip-map and spotlight synthetic aperture systems.

The following mathematical arguments are similar to those that led the initial investigators of the strip-map synthetic aperture technique to use the method for high-resolution image production. The method for achieving high along-track (azimuth) resolution can be described as a focusing operation, a matched filtering or correlation operation, or an along-track compression operation. To determine the possible system resolution, it is necessary to investigate the system response to a point target. This investigation gives the spatially induced Doppler bandwidth produced by the relative platform-target modulation and indicates the basic form of the matched filter. The along-track system resolution can then be determined.

Consider a real aperture system travelling along the u axis transmitting a modulated pulse $p_m(t)$. If the target field consists of a single point target located at $(x, y) = (x_0, 0)$ then the demodulated echo signal modelled using the stop-start assumption (see Section 2.8 for the validity of this assumption) and complex (real and imaginary) signal notation is given by

$$ss_b(t, u) = p_b\left(t - \frac{2R}{c}\right)\exp(-j2k_0 R), \tag{1.5}$$

where the baseband pulse $p_b(t) = p_m(t)\exp(-j\omega_0 t)$ has a narrow extent in time to achieve high range resolution (e.g. it resembles $\mathrm{sinc}(B_c t)$), the carrier wavenumber is $k_0 = \omega_0/c$, spreading effects have been ignored (it is assumed these are compensated for in the receiver), and the spatial-temporal response of the aperture is ignored. (Note: any function with a subscript 'm' implies modulated notation, i.e., the analytic representation of the function still contains a carrier term $\exp(+j\omega_0 t)$, where ω_0 is the carrier frequency. Demodulated functions are subscripted with a 'b' to indicate a *baseband* function.)

The changing platform-target range is

$$R = \sqrt{x_0^2 + u^2} \approx x_0 + \frac{u^2}{2x_0}, \tag{1.6}$$

where the binomial approximation is valid for target ranges $x_0 \gg u$.

The echo function in (1.5) becomes a two-dimensional function due to the u-dependence of the time delays $2R/c$. The along-track direction u is also known as the aperture dimension of the synthetic aperture. At this point it is prudent to comment on the use of the double functional notation. This notation is useful when describing synthetic aperture algorithms due to the fact that various steps in the processing algorithms often occur in pseudo-Fourier space. For example, the range-Doppler algorithm begins with the 2-D matrix of baseband compressed data given by $ss_b(t, u)$ (see Section 4.3.2) and performs a 1-D Fourier transform on the along-track dimension to produce the pseudo-Fourier or range-Doppler data given by $sS_b(t, k_u)$. The double functional notation allows the capitalisation of the character corresponding to the transformed parameter; in this case, the along-track parameter u transforms to wavenumber k_u. The consistent use of this notation throughout the thesis clarifies the mathematical description of the different synthetic algorithms considerably.

To investigate the bandwidth of the spatial Doppler modulation, consider the phase of the return signal in (1.5):

$$\phi(u) \approx -2k_0 \left(x_0 + \frac{u^2}{2x_0} \right). \tag{1.7}$$

To parallel the original SAR development of the mathematics, it is necessary to convert the along-track spatial coordinate to an along-track temporal coordinate using $u = v_p t$, where v_p is the platform velocity. Because confusion can result from the use of the temporal variable t to indicate range time and along-track time, it is often redefined as two quantities; range time or *fast-time*, t_1, is the *continuous* time ordinate for the echoes from the nth transmitted pulse, $t_1 \in [n\tau_{\text{rep}}, (n + 1)\tau_{\text{rep}}]$, while the along-track or *slow-time* ordinate, t_2, is discretized via the pulsed operation of the system to $t_2 = n\tau_{\text{rep}}$. Thus, our echo function in terms of the two temporal quantities becomes the new function $ss'_b(t_1, t_2)$, where $t_1 = t$ and $t_2 = u/v_p$ are both treated as continuous quantities. The effects of discretization are covered in Chapter 5.

The Fourier space of the aperture ordinate, u, used in this thesis is referred to as the aperture *Doppler wavenumber*, k_u, in radians per meter. In the classical development, the Fourier pair is slow-time, t_2, and spatial Doppler frequency, f_d, in Hertz. To determine the *instantaneous spatial Doppler frequency* (in Hz) produced by the point target, the phase function in (1.7) is differentiated with respect

to the slow-time variable:

$$f_{di}(t_2) = \frac{1}{2\pi}\frac{\partial\phi(v_p t_2)}{\partial t_2} \approx -\frac{k_0 v_p^2}{\pi x_0}t_2 = -f_{dr}t_2. \tag{1.8}$$

The instantaneous spatial Doppler frequency (within the validity of the quadratic approximation) can be seen as a linear FM (LFM) chirp of *Doppler rate* f_{dr}. The appropriate matched filter is then also LFM of rate f_{dr}.

The bandwidth produced by the Doppler modulation determines the maximum along-track resolution of the synthetic aperture. Due to the real aperture angles of acceptance, the target only produces a processable response during the time that the target lies within the 3dB width of the main lobe of the radiation pattern, that is for along-track times $t_2 \in [-\tau_d/2, \tau_d/2]$, where the *dwell time* for a broadside orientation is given by

$$\tau_d \approx \frac{x_0\theta_{3\text{dB}}}{v_p} = \frac{x_0\lambda_0}{v_p D} = \frac{2\pi x_0}{k_0 v_p D}. \tag{1.9}$$

This gives a *Doppler bandwidth* (in Hertz) of,

$$B_d = \Delta f_{di} \approx 2 \cdot \frac{k_0 v_p^2}{\pi x_0} \cdot \frac{\pi x_0}{k_0 v_p D} = \frac{2v_p}{D}, \tag{1.10}$$

which after matched filtering (along-track compression) produces an along-track spatial resolution of

$$\delta y_{3\text{dB}} \approx \frac{v_p}{B_d} = \frac{D}{2}, \tag{1.11}$$

which is independent of velocity, wavelength, and range to target. This also indicates that finer resolution can be achieved with a smaller real aperture, an achievement which is contrary to conventional real aperture imaging. These observations were first reported by Cutrona *et al* in 1961 [35], although the calculation had been made by Cutrona in 1954 [32, p340].

Due to the LFM nature of the along-track Doppler modulation, the along-track signal can be compressed using a LFM matched filter in a way that is similar to the pulse compression operation used to achieve high range resolution. This process is only *similar* because we have neglected to include the effect of the pulse envelope. In cases where the changing range to a target causes the envelope of the return signal to shift in range by more than (or a large fraction of) a range resolution cell, the along-track matched filter becomes two-dimensional and more complex methods of along-track compression are required. This effect is typically referred to as *range migration* or *range curvature*. The presence of the delay term $2R/c$ in the delayed versions of the transmitted pulse in (1.5) effects the modulation of the received echoes in realistic cases where the transmitted pulse itself contains some form of modulation

(typically LFM). The processing/inversion algorithms required for realistic situations that account for these and other effects are covered in detail in Chapter 4.

The along-track resolution shown in (1.11) can also be arrived at using a beam forming argument. The exploitation of the phase information allows a synthetic aperture to have an along-track resolving capability that is *twice* that of a non-coherent real aperture system; that is,

$$\delta y_{\mathrm{3dB}}^{\mathrm{synthetic}} = \frac{x_0 \lambda}{2 L_{\mathrm{sa}}}, \tag{1.12}$$

where the origin of the factor of 2 is covered in Chapters 3 and 4, and $L_{\mathrm{sa}} = v_p \tau_d$ is the distance over which a target at range x_0 remains in the real aperture radiation pattern. The substitution of L_{sa} using (1.9), again gives an along-track resolution of $D/2$.

1.5   ASSUMED BACKGROUND

This thesis assumes that the reader has a reasonable knowledge of the synthetic aperture technique. If not, there is a large selection of reference books available. For developments in strip-map SAR imaging prior to 1991 (range-Doppler inversion, sub-aperture/multi-look registration autofocus) see the book from the Jet Propulsion Laboratories (JPL), Curlander and McDonough, 1991 [31], for tomographic based spotlight SAR inversion and phase gradient autofocus (PGA) see the book from Sandia National Laboratories, Jakowatz *et al*, 1996 [82], and also the book from the Environmental Research Institute of Michigan (ERIM), Carrara *et al*, 1995 [21], and for wavenumber (Fourier synthesis) based inversion algorithms see Soumekh, 1994 [146] and Soumekh, 1995 [147].

1.6   HISTORICAL REVIEW

1.6.1   Origins of the synthetic aperture technique in radar applications

The synthetic aperture concept is attributed to C.A. Wiley of Goodyear Aircraft Corp. for his 1951 analysis of along-track spatial modulation (Doppler) in the returns of a forward looking (squinted), pulsed airborne radar [31, 139, 164]. He patented the idea under the name *Doppler beam sharpening* (DBS), the name reflects his frequency domain analysis of the effect. The first real-time airborne SAR processor flown in 1952 was based on the DBS concept. DBS exploits the fact that the repeated transmission of an FM pulse gives a signal with *discrete* spectral lines spaced by the pulse repetition frequency (PRF). The discrete nature of the spectrum means that temporal Doppler shifts that lie within PRF/2 of a spectral line can be measured quite accurately [7, p1035-6]. These Doppler shifts are due to Doppler effects occurring *between* pulses and they represent the relative platform-target speeds, Doppler shifts *within* pulses due to movement during pulse transmission are small and are typically

ignored. If the FM pulse is of wide bandwidth, the system has a high range resolving ability. The inversion of DBS data to produce a high resolution image is termed *velocity-induced* synthetic aperture radar [146, p317]. The original DBS concept was based on a similar plane wave assumption to that used in the tomographic formulation of spotlight SAR inversion schemes. Because of this plane wave assumption, DBS suffers from an impairment known as *motion through resolution cells* [5, 15, 146]. The recent approximation free FM-CW SAR inversion scheme developed by Soumekh does not suffer from this problem and represents a generalization of the DBS concept [146, pp317-336].

Independent of Wiley's work, L.J. Cutrona of the University of Michigan and C.W. Sherwin of the University of Illinois were developing the synthetic aperture principle from an aperture or *spatially-induced* point of view. These systems were the first to be referred to as *synthetic aperture radars* [31, 139, 164]. During the summer of 1953, at the University of Michigan, a combined effort with the University of Illinois team was undertaken under the guise of *Project Wolverine* to develop a high-resolution combat-surveillance radar system [17]. The ability of SAR to remotely observe a scene through cloud cover, observe day and night, and through light rain made it ideal for use as a tactical weapon. The development of the early work in SAR is retraced by Sherwin *et al* in reference [139], and a review of his personal involvement is made by Wiley in reference [164]. Curlander and McDonough [31] also retrace developments from an insiders point of view.

A spatially-induced synthetic aperture is based on the idea that a sequence of pulses recorded from a moving real aperture can, with suitable computation, be treated as the output of a much longer array. Early workers considered *unfocused* SAR [5], however, at the 1953 meeting, Sherwin indicated that a fully *focused* synthetic aperture would produce finer resolution [32, p339]. In 1961, Cutrona first published the fact that a fully focused synthetic aperture produces an along-track resolution that is independent of range and wavelength and depends only on the physical size of the illuminating real aperture [35, 139]. The major problem faced by the initial developers of a focused SAR processor was the range variance of the along-track focusing filters. This problem was overcome with the development of the first optical processor by Cutrona in 1957. In this system the pulse compressed echoes were recorded on photographic film for subsequent ground processing by a system of lenses and optical filters [34]. In the optical processor, the range variant focusing filter conveniently translated to a conical lens. The reviews by Brown and Porcello [17] and Tomiyasu [153] cover optical processing in detail.

The optical processor was the mainstay of SAR processing until the late 1960's when digital technology had advanced to the point where it could handle some sections of the processing [7, 31]. The first fully digital system was developed by Kirk in 1975 [87], the system also included a digital motion compensation system [88]. Since that time, digital-optical hybrids and fully digital processors have become common. Digital implementation of synthetic aperture systems also heralded the introduction of *squint mode* and *spotlight mode* radar systems [87].

The initial SAR systems were developed as airborne SARs; the first spaceborne SAR developed for

terrestrial imaging was the Seasat-SAR launched in 1978. Subsequent spaceborne SARs are described in Curlander and McDonough [31] and Elachi [47]. The spaceborne SAR differs from the airborne SAR in that the effect of range migration is more severe, earth rotation causes an effect known as *range walk*, multiple pulses are in transit at any one time, satellite motion and orbit eccentricity must be accounted for, and ionospheric and tropospheric irregularities cause defocus of the final images [31, 47, 153]. A processor that can handle most of these effects meant a more complex SAR processor was required. One of the first of these processors was developed by C. Wu of JPL [31], references [84,166] are the basis of most of the spaceborne-SAR processors described in 1991 by Curlander and McDonough [31, pp197-208], these processors are termed *range-Doppler processors*.

Walker is credited with the concept of *spotlight mode* SAR in 1980 [83,160] (the systems mentioned by Kirk in 1975 [87] and Brookner in 1978 [13] have a fixed squint angle, so are not true spotlight systems). In spotlight mode, the real aperture is steered or slewed so that the same small area of terrain remains illuminated while the platform traverses the synthetic aperture. In this mode, the along-track dimension of the image becomes limited by the beam width of the real aperture, but the along-track resolution of the final processed image is improved beyond the $D/2$ limit of conventional strip-map systems. Walker interpreted the processing of the spotlight synthetic aperture data in terms of a Fourier synthesis framework. This insight lead others such as Munson [114], Soumekh [146], and Jakowatz *et al* [82] to describe SAR processing in a more consistent signal processing framework. In the Fourier synthesis view, the raw data is shown to be samples of the Fourier transform of the image reflectivity at discrete locations in the 3-D Fourier space of the object. These discrete samples are interpolated onto an 2-D rectangular grid appropriate for inverse Fourier transformation by the inverse Fast Fourier Transform (iFFT). Initial developments in this and similar areas are reviewed by Ausherman *et al* [5].

Until the early 1990's the processing algorithms for strip-map SAR were largely based on the range-Doppler algorithm. During the early 1990's what is now referred to as *Fourier-based multidimensional signal processing* was applied to develop a more concrete theoretical principle for the inversion of SAR data [20, 146, 147]. The algorithms produced by this theory are now generically referred to as *wavenumber algorithms*. The development of these wavenumber algorithms represents a breakthrough as dramatic as the development of the original optical processor [6]. The initial wavenumber processors developed required interpolators to perform a remap of the raw data onto a rectangular grid suitable for Fourier processing. Developments of two groups at the International Geoscience and Remote Sensing Symposium in 1992 [30, 127, 134] led to the publication in 1994 of a new wavenumber inversion scheme known as *chirp scaling* [128]. Chirp scaling removes the interpolator needed in the previous wavenumber inversions and represents a further advancement of SAR processing. This thesis takes this algorithm one step further to produce the accelerated chirp scaling algorithm.

The first SAR images produced had basic *motion compensation* applied to them, but they still suffered along-track streaking due to the presence of residual timing errors. These initial motion compen-

sation schemes double-integrated the outputs of the accelerometers contained in *inertial measurement units* (IMUs), sometimes called *inertial navigation units* (INUs), to determine the bulk displacement errors. The residual timing errors that caused final image smears were most likely due to small uncorrected platform motions (due to IMU drift errors), transmitter/receiver timing errors, or medium inhomogeneity and as such could not easily be measured. The algorithms that can determine these errors use the data itself to iterate until an image with optimum focus is achieved. These algorithms are known as *autofocus* procedures. In the SAR community, there are three main autofocus algorithms that are routinely used to remove these residual errors. *Contrast Optimisation* algorithms [9,120] determine the final focused image on the basis that the image with highest contrast is the optimal image. *Subaperture*, *Map-drift* or *Multilook registration* algorithms [31,120] are based on the correlation of multilook images. *Phase Gradient Autofocus* (PGA) [44,82,158] is typically used in the processing of spotlight-SAR images, but has recently been adapted for strip-map systems [159]. PGA models the along-track error as a phase-only function of the Doppler wavenumber in the range-Doppler domain. By using redundancy over range in the range-Doppler data an estimate of this phase error can be obtained, and removed from the image. All of these algorithms are iterative in nature with the fastest algorithm, straight PGA with no enhancements, typically converging in 2-3 iterations. Chapter 6 develops these and several other more specialised SAR autofocus procedures, examines their applicability to wide bandwidth, low-Q systems and proposes an autofocus procedure for these low-Q systems.

Several other SAR inversion schemes are not dealt with in this thesis, however these methods are briefly mentioned as the principles can be applied to sonar applications. Interferometric SARs use multiple receiving antennas or multiple passes over an area to measure the height of terrain along with its reflectivity properties. The first application of this type of radar system is reported by L.C. Graham of Goodyear Aerospace Corp. [64,164], more recent processors are described in [82,146]. Inverse SAR (ISAR) uses a stationary antenna and uses the spatial Doppler information induced by a moving target to give cross-range resolution. Applications typically include missile and aircraft imaging [105] and planetary imaging [5].

The non-military applications of SAR are diverse and range from interplanetary imaging, terrestrial mapping, observing ocean current and wave spectra (oceanography), changing patterns of vegetation, and polar ice flow [5,31,47].

1.6.2 Applications of the synthetic aperture technique to sonar

The fundamental difference in propagation speeds between SAR and SAS have limited the application of the synthetic aperture technique in sonar. Although this difference makes higher resolution easier to achieve with smaller bandwidths for sonar, the problem of ambiguities introduced by PRF constraints and more sensitive motion compensation requirements far outweighs this gain [32, 33]. In airborne

SAR, the high speed of electromagnetic propagation, coupled with the modest standoff ranges from the platform to the target swath, allows many more finely spaced views of a scene than is possible with SAS. This higher along-track sampling rate is used to resolve ambiguities that a practical SAS system cannot [138]. When the standoff range becomes large, as is the case with spaceborne SAR, the along-track sampling requirements of SAR become similar to those faced by SAS systems. Chapter 5 covers the sampling requirements of these systems in detail.

The first synthetic aperture sonar systems to appear in the available unclassified literature (i.e., not in internal company or university reports) are the systems described in the patent by Walsh 1969 [161] and the experimental tank test system of the Tokyo Institute of Technology described in Sato *et al*, 1973 [137] and Sato and Ikeda, 1977 [135, 136]. Burckhardt *et al*, 1974 [19] details a synthetic aperture system for use in medical applications. In 1975 and 1977, Cutrona published two papers comparing conventional real aperture side-looking sonar (SLS) systems to their SAS counterparts [32, 33], a similar comparison was made by Bruce in 1992 [18]. Both authors concluded that the use of the synthetic aperture technique results in a significant improvement of the along-track resolution while maintaining comparable, and even greater mapping rates than is possible with conventional SLS.

The limiting factor in the early applications of the synthetic aperture technique to sonar was thought to be medium instability. In 1976, Williams published the results of coherency tests performed with a 400Hz carrier, 100Hz bandwidth system transmitting over one-way paths that varied from 100km to 500km. On an intermittent basis, the medium maintained coherency for up to 7.5 minutes and synthetic apertures close to 1km long were formed [165]. Lee, 1979 [94] and Rolt, 1989 [133] refer to an internal report by Stowe, 1974 [152] that describes a 10kHz system that operated over a 2.5km one-way path. The rms fluctuations over this acoustic path were on the order of one-hundredth of a wavelength at 10kHz over a 1 minute interval. Medium stability was confirmed by further coherency studies performed by Christoff *et al* in 1982 [26] with a 100kHz system that transmitted $140\mu s$ pulses over a one-way path of 48m, and by Gough and Hayes in 1989 [62] using their 22.5kHz carrier, 15kHz bandwidth system, over a two-way path using a reflecting target at a range of 66m. Both experiments concluded that the medium is stable over time periods of a minute or more and that the medium is coherent enough to form synthetic apertures. Section 4.3.2 of Rolt's thesis [133] contains further descriptions of medium stability experiments. The results of the previous University of Canterbury SAS system described in Hayes' 1989 thesis [75] and references [63, 74], the positive results of the SAS systems described shortly, complex simulation studies [22, 23, 78, 133, 167], and the positive results of this thesis, all confirm that the formation of a synthetic aperture in the ocean is viable.

The pulsed operation of any synthetic aperture system results in range and along-track ambiguity constraints that can be interpreted as limiting the mapping rate of the system, or limiting the along-track resolution of the system. Ambiguity constraints are easily satisfied in airborne SAR systems. However, the PRF and hence maximum range and along-track sample spacing of spaceborne SARs and SAS are

limited. In an attempt to circumvent these ambiguity constraints, multiple aperture systems, multiple beam systems, multiple pulse systems, and wide bandwidth systems have been proposed. Chapter 5 reviews these techniques and shows that the most viable option for a commercial synthetic aperture system is a system with a single transmit aperture coupled with a multiple aperture receiver.

To date, no unclassified SAR system uses the range or along-track spatial bandwidths employed by the Kiwi-SAS. The final SAS image resolution of 21cm×5cm is considerably finer resolution than that achieved by any SAR of equivalent carrier wavelength. The fine resolution is due to the correspondingly high spatial bandwidths covered by the system; that of range due to the chirp bandwidth coupled with the slow speed of sound in water and that of along-track due to the small real apertures employed. Access to this wide spatial bandwidth makes the applicability of normal SAR algorithms uncertain. One aspect of this thesis is to report on the wide spatial bandwidth properties of range-Doppler and wavenumber SAR algorithms for application to wide bandwidth SAS. Initial investigators of wide bandwidth SAS systems have typically only considered time-domain beam forming systems, time-domain correlators, or invariant matched filtering schemes. These schemes had similar aspects to the range-Doppler inversion scheme, but none were true wide bandwidth implementations. These investigators considered the SAR radar inversions to be applicable to narrow bandwidth systems only and so schemes were adopted whereby the wide bandwidth was split and processed as separate narrowband systems [22, 23, 75, 167]. Wide bandwidth systems have been determined by a number of investigators as ideal for applying the synthetic aperture technique. Wide bandwidth systems produce cleaner images in the presence of uncompensated motion errors, they provide richer information, via the bottom echoes, that could be used for sea floor classification, and the wide bandwidth smearing of the grating lobes in undersampled systems has been seen as a method for increasing the mapping rate [22, 23, 38, 74]. Given these observations, this thesis also reports on the effects of undersampling and presents methods for retrieving useful information from undersampled data and methods for suppressing ambiguous targets due to grating lobes. Many of the initial wide bandwidth investigations were performed on simulated data in which temporal Doppler had not been modelled [23], this thesis reports on the validity of this approximation and details Doppler tolerant waveforms available for use in realistic systems. For a review of synthetic aperture sonar development up until 1991 the reader of this thesis is directed to the thesis by Rolt [133] (The processor described in Rolt's thesis is an "exact" time-domain beam forming processor, similar to those described in Section 4.3.1). For a review of a broadband SAS system containing an "exact" processor analysed from a time-domain and frequency-domain fast correlation point of view (i.e., range invariant matched filtering) see Hayes and Gough, 1992 [74] (this is also discussed in Section 4.3.1).

During the conception of this thesis in 1993, there were few operational SAS systems reported in the unclassified literature. During the development of the University of Canterbury's Kiwi-SAS system, four other ocean-going SAS developments were discovered to be underway. These were the unclassified Euro-

pean ACID (ACoustical Imaging Development) system, now referred to as the SAMI (Synthetic Aperture Mapping and Imaging) system, the University of California at Santa Barbara (UCSB)/Sonatech multiple receiver system, and a system developed by Alliant Techsystems for the University of Hawaii and the National Defense Center of Excellence for Research in Ocean Sciences (CEROS) (an agency of the State of Hawaii), and the classified system of a combined French Navy/U.S. Navy cooperative venture. References to tank test systems were also found [4, 66, 104, 138]. The Loughborough University of Technology's (LUT's) tank system has also been used by LUT and University College London (UCL) to evaluate bathymetric or interferometric-SAS [4, 66, 104].

Some of the simulation studies leading up to the development of the wide bandwidth ACID/SAMI system are covered in the papers by Adams [1] and Chatillon, Zakharia, and Bouhier [22,23,167]. Further simulation studies and real sea floor results are covered in the papers by Lawlor *et al* [93], Riyait *et al* [130, 131], and Adams *et al* [2]. The ACID/SAMI project is a co-operative research project funded by the European Commission in the framework of the MAST (MArine Science Technology) project. The ACID/SAMI system consists of a single 2m aperture that operates at a carrier frequency of 8kHz, transmits linear period modulated pulses with a bandwidth of up to 6kHz, and typically operates out to a maximum range of 600m. The system has a theoretical range resolution of 13cm and a theoretical along-track resolution of 1m. The processor is time-domain based, operates on range-compressed data, and operates in real-time using a transputer based architecture, with motion compensation provided in real-time from IMU's (the inertial measurement data is currently redundant as it is not used by the current image processor [2]). The first successful tests of the system were performed in May 1993. During sea trials in July 1995 [2] and further sea trials in May 1996, multiple receiver data, bathymetric data, and long range SAS data was collected. Preliminary processing of the long range data has yielded low contrast images with 1m along-track resolution at a range of 3km. An image produced with the bathymetric data is reproduced in [2], however, images produced from the multiple receiver data are not available. No autofocus procedures have been applied to the ACID/SAMI system and although the ACID/SAMI simulation papers show the effects of along-track errors, none investigate autofocus procedures for removal of these errors. Even without the use of autofocus procedures, or in fact IMU measurements, the short range images produced by the ACID/SAMI system have high contrast (i.e., are well focused).

The UCSB/Sonatech system is described in B. L. Douglas' 1993 thesis [43]. The system consists of a multiple receiver SAS that produces well focused images using a time-domain beam forming processor and an autofocus procedure now referred to as *cascade autofocus*. Cascade autofocus utilizes the correlation between the physical aperture images available in a multi-receiver system to determine the towfish motion parameters. Each physical aperture image is formed in real-time using the multiple receiver outputs from the echoes received from each transmitted pulse. The system is inherently narrowband, transmitting burst waveforms with a bandwidth of 4kHz centered on a 600kHz carrier. The aperture

consists of ten, 10cm long, separately channeled elements. A single element is used on transmit, and all 10 channels are stored on receive. The theoretical resolution of the system in along-track is 5cm and in range is 19cm, 100m swaths are typically recorded for processing. Multiple along-track receivers have been seen as a method of increasing the allowable along-track platform velocity while satisfying along-track sampling requirements [93]. Alternatively, multiple receiver systems have been seen as a method of autofocusing SAS data [138]. By taking multiple measurements at the same along-track locations, phase differences between measurements can be measured. These phase differences correspond to the errors due to platform or medium errors and can be used to correctly focus images [138]. The UCSB system and others are discussed and analysed in more detail in Section 5.4.2.

Alliant Techsystems SAS system has been developed for the detection and classification of objects buried below the ocean floor. This type of system has applications including; environmental monitoring and cleanup tasks, geological surveying, and unexploded ordinance clearance. The system consists of a single small transmitter operating at 12.5kHz that transmits a 3 cycle Hanning weighted pulse (8kHz bandwidth) every 0.1s. Imaging of a 75m swath is possible at a range resolution of 10cm. The receiver array consists of four separately channeled, keel mounted, 20cm long elements. The longest individual aperture is a single receiver element, giving a theoretical along-track resolution of 10cm. Alliant outlines the system in their 1996 PACON paper [118]. Their paper contains images of targets (retro-reflectors, basketballs, cylinders, and spheres) located in a strip 6-18m from the towfish path. The processor described is based on a time-domain beam forming approach, motion compensation is provided via inertial measurement unit (IMU) output and via their proprietary autofocus technique; aperture position correction (APC).

The French and U.S. Navy have undertaken a joint program to produce a mine countermeasures SAS; the MUDSS SAS (mobile underwater debris survey system). The program has used the dual frequency (60kHz and 200kHz carriers) rail system of Guyonic [68](French Navy) and in July 1996, sea trialed another dual frequency SAS. This dual frequency SAS consists of a high frequency (HF)(180 kHz) and a low frequency (LF)(20kHz) system. The HF system has a theoretical resolution of 2.5cm (30kHz bandwidth) and the LF system has a resolution of 7.5cm (10kHz bandwidth) and has multi-aspect capability. It is designed to operate in shallow water in a high multi-path environment. The HF should produce shadows if the multi-path does not fill them in and the LF is designed to find and classify buried objects. It is designed to operate between 4 and 8 knots with a maximum range of 90 and 45 meters respectively. The primary objective of the sea tests was to test the hardware for proper operation and to test out motion compensation algorithms, especially the 2.5 cm resolution SAS. They are concentrating on motion compensation based on diffuse scattering algorithms [27]. Preliminary results from these sea trials produced 5-7.5cm imagery for the HF system and 7.5-10cm for the LF system. The comment regarding application of the simulator described in Huxtable and Geyer's paper [78] to the U.S. Navy's database of SAS data could imply that the processor described in [78] is one of the algorithms being

considered for the MUDSS processing algorithm. Huxtable and Geyer's paper [78] presents a complete sonar simulation of a multiple receiver, 150kHz carrier, 15kHz bandwidth system that includes navigation devices and autofocusing algorithms. Their paper demonstrates the more stringent motion compensation requirements of SAS relative to SAR and show that the motion compensation requirements for high-resolution SAS exceed off-the-shelf motion compensation sensors.

These recent synthetic aperture sonars have been developed for two main purposes; the non-military groups have developed systems specifically for sea floor and sub-bottom imaging, while the military designs are to be used to detect (and classify) mines (proud mines, tethered mines, buried mines). Objects lying proud of (above) the sea floor generate acoustic shadows which are used by classification algorithms in the detection of some of these mines.

1.7   THESIS CONTRIBUTIONS

There are a number of major and minor contributions to both the fields of radar and sonar contained within this thesis.

In terms of hardware, the wide bandwidth multiresonant Tonpilz transducer represents a unique and novel transducer for high-resolution sonar imaging. The design of these transducers using the Finite Element method is quick and inexpensive due to the ability of the Finite Element method to accurately model the physical and electrical characteristics of the transducer in software. This design method is detailed in Appendix B.

The introduction of the mapping operator in Chapter 2 and its application to both one- and two-dimensional imaging problems improves the clarity of the models encountered in imaging problems. This is especially true when dealing with the synthetic aperture imaging algorithms in Chapter 4. The derivation of the spatial-temporal domain algorithm, the fast-correlation algorithm, the range-Doppler algorithm, the wavenumber algorithm, the chirp scaling algorithm, the tomographic formulation of spotlight mode, the generalized form of spotlight mode, and FM-CW imaging in a unified mathematical framework allows anyone competent in mathematics and Fourier theory to follow what could be referred to as the 'black-art' of synthetic aperture imaging. The consistent mathematical framework of these system models allows a comparison between the different imaging modalities. A number of the algorithms derived are unique in that they are derived in their *wide bandwidth form*. No published SAR or SAS system uses the same spatial bandwidth as the Kiwi-SAS, so the derivation of the wide bandwidth versions of these algorithms has not previously been necessary. These wide bandwidth versions differ from the previous derivations in that deconvolution filters are necessary to remove wide bandwidth amplitude modulating terms. The compensation of these terms allows for the proper calibration of the final image estimate. These algorithms are all presented in such a way as to allow a direct digital implementation.

This thesis has also increased the efficiency of the chirp scaling algorithm. The *accelerated* chirp scaling algorithm is elegant in that it minimizes the processing overhead required to focus synthetic aperture data, yet the whole algorithm consists of simple multiplies and Fourier transform operations, both of which are easily implemented on general purpose hardware. These features are likely to make it the processor of choice in future SAR and SAS systems.

Chapter 6 investigates the commonly used SAR autofocus algorithms. This chapter details the modified phase gradient autofocus algorithm. Phase gradient autofocus in its original form limits its applicability to high-Q systems that satisfy the tomographic formulation of spotlight mode SAR. The modifications presented in this chapter allow it to be applied to low-Q spotlight systems that need to be processed via the generalized spotlight algorithm. Phase curvature autofocus is also analysed in a low-Q form and a new autofocus routine for wide bandwidth, low-Q, systems is proposed.

The results chapter, Chapter 7, contains details on the end-to-end design, simulation, implementation, and calibration of the wide bandwidth, low-Q, Kiwi-SAS system. This chapter provides insights into the operation of the processing algorithms and shows that secondary range compression (SRC) is necessary in wide bandwidth systems to remove deterministic phase errors in the final image estimate. Autofocused, calibrated, diffraction limited images of a calibration target are presented. These images were obtained by processing echo data that had been collected during harbour trials of the Kiwi-SAS. During these sea trials, the Kiwi-SAS was not constrained in any way, and there was no inertial navigation system to provide motion estimates to the focusing algorithm.

Other contributions to the fields of SAS and SAR are found throughout the thesis; Chapter 2 contains an equation set from which the validity of the stop-start assumption used to develop the inversion schemes can be validated for a given radar or sonar system. Chapter 3 introduces the relatively new field of Fourier array imaging and relates it to conventional classical array theory. This chapter then presents the correct interpretation of grating lobe suppression. This interpretation reveals that the along-track sample spacing necessary to avoid aliasing of the mailobe energy of the radiation pattern is $D/4$, not $D/2$ as is quoted in most of the synthetic aperture literature. Chapter 5 shows how the $D/2$ sample spacings commonly used in spaceborne SAR applications are actually a trade-off of the final image dynamic range for a relaxed along-track sampling requirement. The level to which the along-track sampling requirement can be relaxed is quantified by the predicted along-track ambiguity to signal ratio (AASR) for a given *processed* along-track bandwidth. The discussion on AASR also gives insight into what is occurring when systems process a *reduced* along-track bandwidth in order to remove ambiguous targets from undersampled synthetic aperture images. Chapter 5 also analyses the published techniques for suppressing or avoiding ambiguities in synthetic aperture applications. Some of these methods are modified to apply to wide bandwidth, low-Q systems. An investigation of the use of wide bandwidth signals to smear grating lobe target levels yielded an accurate method of predicting the final image dynamic range. The level of the smeared grating lobes to the main target peak is accurately quantified

by the peak to grating lobe ratio (PGLR). The PGLR is also used to emphasise the fact that a $D/2$ sample spacing does not adequately sample the along-track dimension of the synthetic aperture. If it is assumed that $D/2$ along-track resolution is required in the final synthetic aperture image and that the highest image dynamic range is desired, then a sample spacing of at least $D/3$ is necessary. The discussion on multilook processing in Chapter 4 makes the important observation that any multilook processing in SAS applications should be applied to the range bandwidth in contrast to SAR systems that often perform multilook using the along-track bandwidth.

1.8   THESIS ORGANISATION

Chapter 2 introduces the relevant notation and signal processing concepts that are used within this thesis and within the SAR and SAS communities. Pulse compression techniques are reviewed by way of one-dimensional range imaging examples. The use of mapping operators in these examples is a precursor to their use in the two-dimensional imaging models of Chapter 4

Chapter 3 presents array theory in terms of Fourier array theory and classical array theory. This chapter shows how the wide bandwidth radiation patterns of the real apertures can be accurately modelled in the system models. The correct interpretation of grating lobe suppression in synthetic arrays is given.

Chapter 4 discusses the properties of the complex data encountered in coherent systems, presents the wide bandwidth forms of the commonly used synthetic aperture imaging algorithms, and presents the new accelerated chirp scaling algorithm. Multilook processing of synthetic aperture data is reviewed and the implementation details of a realistic processor are given.

Chapter 5 analyses the effects of sampling and undersampling on the along-track and range signals. Along-track and range ambiguities are discussed and quantified. Techniques for grating lobe suppression and avoidance are then analysed.

Chapter 6 reviews and analyses motion compensation techniques. Limitations on the applicability of SAR algorithms are discussed. Phase gradient autofocus, modified phase gradient autofocus, and phase curvature autofocus are developed and investigated for their application to low-Q systems.

Chapter 7 reviews the Kiwi-SAS system and contains an analysis of the algorithms presented in Chapter 4. The results of calibration sea trial tests are given. These results include the autofocused,

calibrated, diffraction limited images of a calibration target.

Chapter 8 contains the conclusions of the research, recommendations for future work, and recommendations for the Kiwi-SAS II.

Chapter 2

SIGNAL PROCESSING AND RANGE RESOLVING TECHNIQUES

This chapter introduces the relevant notation and signal processing concepts that are used within the scope of this thesis and within the SAR and SAS communities. This chapter begins with a brief review of sampling theory and its possible embodiments for synthetic aperture systems. The utility of mapping operators is then discussed. Pulse compression techniques are reviewed and applied to one-dimensional imaging examples as a precursor to the two-dimensional imaging algorithms presented in Chapter 4. The waveforms commonly used by synthetic aperture systems and their properties are then presented along with their tolerance to temporal Doppler effects. The discussion on temporal Doppler encompasses a discussion on the ambiguity function and details its analysis for wide bandwidth applications. This analysis of the wideband ambiguity function shows that the 'stop-start' approximation commonly employed in radar to develop system models is also applicable to sonar. Finally, methods for the accurate interpolation of complex valued data are given.

The final image produced by a synthetic aperture processor should be calibrated such that pixel intensity reflects the scattering cross-section of the location being imaged. In SAR this is known as *radiometric calibration* and the scattering parameter is known as the *radar cross-section* (RCS), in sonar this scattering parameter is characterised by its *backscatter cross-section* or its *target strength* [52, 92]. For accurate calibration, it is necessary to model the system completely, *including* the often neglected real or complex amplitude functions. This is especially important in wide bandwidth applications where these amplitude functions may *not* be constant. Calibration using these deterministic constants allows simulation errors or oversights to be quickly identified and easily corrected (these errors are missed if images are arbitrarily normalized to unity using the maximum image point). By including complex constants, comparisons of Fourier transforms calculated via the fast Fourier transform (FFT) can be made with their continuous or *analytic* forms.

2.1 SAMPLING THEORY

The following arguments are developed for signals parameterized by time, t, and angular or radian frequency, ω. These arguments are equally applicable to signals parameterized in terms of spatial quantities such as distance, x, and wavenumber k_x. Both temporal and spatial signals are analysed extensively throughout this thesis.

To digitally represent a continuous time domain signal $g(t)$ it is necessary to *sample* the signal. This (impulse) sampling operation is described by

$$
\begin{aligned}
g_t(t) &= g(t) \cdot \sum_{m=-\infty}^{\infty} \delta(t - m\Delta_t) \\
&= \sum_{m=-\infty}^{\infty} g(m\Delta_t) \cdot \delta(t - m\Delta_t),
\end{aligned}
\tag{2.1}
$$

where Δ_t is the sampling interval and $g_t(t)$ has a subscript 't' to indicate *temporal* sampling. It is important to note that (2.1) is still a *continuous* representation of the sampled signal. However, the sampled signal is fully characterised by the m samples given by $g(m\Delta_t)$. The continuous nature of (2.1) is exploited in Chapter 5 when developing a model for, and explaining the effects of along-track under-sampling.

The continuous Fourier transform of (2.1) is

$$
\begin{aligned}
G_t(\omega) &= G(\omega) \odot_\omega \frac{2\pi}{\Delta_t} \cdot \sum_{m=-\infty}^{\infty} \delta\left(\omega - m\frac{2\pi}{\Delta_t}\right) \\
&= \frac{2\pi}{\Delta_t} \cdot \sum_{m=-\infty}^{\infty} G\left(\omega - m\frac{2\pi}{\Delta_t}\right),
\end{aligned}
\tag{2.2}
$$

where $\odot_\omega$ represents convolution in ω and the Fourier transform operation is defined in (2.9). The effect of sampling the function $g(t)$ is to generate repeated copies of its scaled continuous spectrum $2\pi/\Delta_t \cdot G(\omega)$ every $m2\pi/\Delta_t$. The repeated nature of this continuous spectrum is important when dealing with the array theory presented in Chapter 3.

Equation (2.2) represents a *continuous* function and as such also needs to be sampled for use in

digital applications. This *spectrally* sampled signal is given by

$$G_s(\omega) = G_t(\omega) \cdot \sum_{n=-\infty}^{\infty} \delta(\omega - n\Delta_\omega)$$

$$= \sum_{n=-\infty}^{\infty} G_t(n\Delta_\omega) \cdot \delta(\omega - n\Delta_\omega),$$

(2.3)

which has the inverse Fourier transform,

$$g_s(t) = g_t(t) \odot_t \frac{2\pi}{\Delta_\omega} \cdot \sum_{n=-\infty}^{\infty} \delta\left(t - n\frac{2\pi}{\Delta_\omega}\right)$$

$$= \frac{2\pi}{\Delta_\omega} \cdot \sum_{n=-\infty}^{\infty} g_t\left(t - n\frac{2\pi}{\Delta_\omega}\right).$$

(2.4)

The effect of frequency sampling is to repeat copies of the scaled *temporally* sampled signal $2\pi/\Delta_\omega \cdot g_t(t)$ every $n2\pi/\Delta_\omega$. The repeated nature of this temporally and spectrally sampled signal seems to imply that the data is corrupted by the repetition. This observation is true if m or n are allowed to take on arbitrarily high values. A digital processor can not deal with an infinite number of samples, so that the values of m and n must be finite and as such there exists a value for m and n such that the sampled data suffers minimal corruption. If the temporal signal is repeated every $2\pi/\Delta_\omega$ and is sampled every Δ_t then m can take on the values $m \in [1, M]$, where

$$M = \frac{2\pi}{\Delta_t \Delta_\omega}.$$

(2.5)

Similarly, if the spectrum is repeated every $2\pi/\Delta_t$ and sampled every Δ_ω then $n \in [1, N]$, where N takes on the same value as M. The discrete representation of the time-limited temporal and frequency-limited spectrally sampled signals are the essence of the discrete Fourier transform (DFT) and its efficient implementation via the fast Fourier transform (FFT). The FFT of a temporal signal containing M samples gives a *single copy* of the *repeated* spectrum, and conversely the inverse FFT of a spectral signal containing N samples gives a single copy of the repeated temporal signal.

2.1.1   Real signals and Nyquist rate sampling

The signals used in synthetic aperture systems can be described in terms of *band-limited* functions. A band-limited real function, $g_r(t)$, having an amplitude function $g_0(t)$ and phase function $\phi(t)$ at a carrier

radian frequency ω_0 is mathematically described by (p89 [76])

$$g_r(t) = g_0(t) \cos \left[\omega_0 t + \phi(t) \right].$$ (2.6)

The amplitude function $g_0(t)$ is a slowly varying function also referred to as the envelope of the signal. In the range dimension of the synthetic aperture model, this amplitude function reflects the weighting of the transmitted pulse and in the along-track dimension it reflects the effect of the overall radiation pattern of the transmit and receive real apertures.

In the case of purely amplitude modulated (AM) pulses, the signal bandwidth is approximately given by the inverse of the temporal duration at the 3dB point of $g_0(t)$. For the more usual case of some form of phase modulation (PM), the system bandwidth is determined by the the modulating phase. In either case, if the spectrum of the real signal is zero for radian frequencies $|\omega| \geq \omega_{\max}$, then temporal samples taken at spacings $\Delta_t \leq \pi/\omega_{\max}$ are sufficient to reconstruct the continuous signal (see Appendix A in Curlander and McDonough [31]). Sampling at the minimum allowable temporal spacing $\Delta_t = \pi/\omega_{\max}$ is termed Nyquist rate sampling. The *sampling frequency* $f_s = 1/\Delta_t$ or $\omega_s = 2\pi/\Delta_t$ represents the extent of the Fourier domain that is sampled by the system, and also represents the distance over which this sampled information repeats in the Fourier domain.

For the real valued temporal signals encountered in synthetic aperture imaging applications, the *sampled bandwidth* is typically much larger than the transmitted *signal bandwidth*. Radar systems typically transmit a wide bandwidth signal (many MHz) about a high carrier frequency (many GHz), so Nyquist rate sampling is suboptimal. There are two more efficient sampling methods that can be employed. The first technique involves multiplying (demodulating) the received signal $g_r(t)$ with an *intermediate frequency* (IF) that acts to shift the signal bandwidth to a much lower carrier. This intermediate signal is then Nyquist rate sampled at a significantly lower rate. This IF method was employed by the Seasat-SAR processor (for example, see p183 of [31] and p1035 [7]). The second technique involves the sampling of the *complex baseband signal* that is described next.

2.1.2   Complex valued baseband signals

The real *band-limited* signal given by $g_r(t)$ in (2.6) can also be represented by the real part of its *complex pre-envelope* or *complex modulated representation*, $g_r(t) = \mathbf{Re} \left\{ g_m(t) \right\}$, where the complex modulated signal is given by (pp60-64 and p137 [28],pp83-91 [76],pp10-24 [129])

$$g_m(t) = g_0(t) \exp \left[j\omega_0 t + j\phi(t) \right].$$ (2.7)

Because only positive frequencies exist in the complex modulated signal it is also known as the one-sided form of $g_r(t)$. To generate the *complex baseband form* of the signal from the complex modulated signal

requires the removal of the carrier:

$$g_b(t) = g_m(t) \exp(-j\omega_0 t) = g_0(t) \exp\left[j\phi(t)\right]. \qquad (2.8)$$

The use of complex mathematics simplifies the modeling of the synthetic aperture system considerably. Also, as most practical systems perform operations on the complex basebanded samples, this is also the most appropriate format in which to develop the processing mathematics.

The conversion of a *real* Nyquist rate sampled signal into a complex signal is most easily performed by removing the negative frequencies in the signal. To achieved this, the signal is Fourier transformed and the negative frequency components are set to zero. This operation halves the signal power, so the one-sided spectrum is then multiplied by 2. This new spectrum is then inverse Fourier transformed to give the complex modulated form of the signal. The complex modulated signal is then demodulated to give the complex baseband signal (see p86 [76]). If this complex baseband, Nyquist rate sampled signal, is down-sampled (decimated) to the same sampling rate as quadrature sampling of the same signal, essentially identical samples are produced.

2.1.3   In-phase and quadrature (I-Q) sampling

In-phase and quadrature (I-Q) sampling splits the input signal into two paths. In the first path the signal is multiplied (mixed) with an in-phase version of the carrier, the result is low-pass filtered, and then sampled; this sample is assigned to the real part of a complex number. The other path is multiplied (mixed) with a quadrature (90° out-of-phase) version of the carrier, low-pass filtered and sampled; this sample is assigned to the imaginary part of a complex number (see Fig. 2.32 on p88 of [76]). If the spectrum of the baseband complex signal has a bandwidth B_c (Hz), then *complex* samples taken at any rate $\omega_s \geq 2\pi B_c$ are sufficient to reconstruct the continuous signal.

As stated above, it is not practical to Nyquist rate sample radar signals so either an intermediate frequency is employed, or the signal is I-Q sampled. Sonar systems operate at significantly lower frequencies than radar so they have a choice of either Nyquist rate or quadrature sampling. However, the I-Q method allows the use of a lower sampling frequency in most practical situations.

2.1.4   A comment on spectral notation

A baseband signal spectrum is a "window" of the signals continuous spectrum, this window has a width ω_s and is centered about a carrier ω_0. Though convention denotes the baseband frequencies $\omega \in [-\omega_s/2, \omega_s/2]$ the spectral samples actually correspond to samples from $\omega_0 + \omega$. When dealing with Nyquist rate sampled spectrums, the frequencies are also denoted $\omega \in [-\omega_s/2, \omega_s/2]$ (where ω_s is generally much higher than the baseband case), however, this time the frequencies correspond to the

correct spectral components. To avoid notational confusion, this thesis denotes the baseband frequencies as $\omega_b \in [-\omega_s/2, \omega_s/2]$ and the actual frequency components as ω, where $\omega = \omega_b + \omega_0$ when dealing with baseband signals.

2.2   MAPPING OPERATORS

In many of the mathematical models described in this thesis, it is often necessary to use a change of variables. When dealing with continuous quantities this presents few difficulties, however, a digital implementation requires a sampled representation. If the change of variables is non-linear, then interpolation is often necessary to map the sampled data from one rectangular sampling grid to another. The sampling grid is required to be rectangular as much of the processing involves Fourier transforms via the FFT. Many of the mapping operators involve a simple axis scaling. Explicitly stating the change of variable through the use of a mapping operator improves the clarity of many of the synthetic aperture algorithms. Forward mapping operators are represented by bold symbols such as the Stolt mapping $\mathcal{S}\{\cdot\}$ and the rotation operator $\mathcal{R}\{\cdot\}$, inverse mappings have the superscript '-1'. Fourier transforms are also mapping operations. Forward and inverse Fourier transformations are denoted by $\mathcal{F}\{\cdot\}$ and $\mathcal{F}^{-1}\{\cdot\}$ and have subscripts denoting the domain *from* which the operation is being performed. The definition of the forward temporal Fourier transform used in this thesis is

$$
\begin{aligned}
G(\omega) &= \mathcal{F}_t\{g(t)\} \\
&= \int_t g(t)\exp(-j\omega t)dt.
\end{aligned}
\tag{2.9}
$$

The forward transform maps signals to the ω (spectral) domain, and the inverse transform

$$
\begin{aligned}
g(t) &= \mathcal{F}_\omega^{-1}\{G(\omega)\} \\
&= \frac{1}{2\pi}\cdot\int_\omega G(\omega)\exp(+j\omega t)d\omega
\end{aligned}
\tag{2.10}
$$

maps the signal spectrum back to the time domain. The spatial Fourier transform used in this thesis has the same definition with the temporal variables t and ω replaced by their spatial counterparts, eg., x and k_x.

The definition of a mapping operator is best described using an example. The polar mapping operator is derived as follows: any integral in the (x, y)-domain can be represented in terms of an integral in the (r, θ)-domain via

$$
\int_x \int_y ff(x,y)dxdy \equiv \int_r \int_\theta ff[x(r,\theta), y(r,\theta)]JJ(r,\theta)drd\theta.
\tag{2.11}
$$

The *non-linear* polar coordinate transform is

$$r(x,y) \equiv \sqrt{x^2 + y^2}$$
$$\theta(x,y) \equiv \tan^{-1}\left(\frac{y}{x}\right) \tag{2.12}$$

and the *Jacobian* of the polar transform is [146, p156]

$$JJ(r,\theta) = \left|\frac{\partial(x,y)}{\partial(r,\theta)}\right|$$
$$= \left|\begin{matrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{matrix}\right| \tag{2.13}$$
$$= r.$$

Sometimes it is easier to calculate the Jacobian via

$$JJ(r,\theta) \equiv \frac{1}{JJ(x,y)}$$
$$= \left(\left|\frac{\partial(r,\theta)}{\partial(x,y)}\right|\right)^{-1} \tag{2.14}$$
$$= \left(\left|\begin{matrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{matrix}\right|\right)^{-1}.$$

The Jacobian reflects the scaling of the differential element $dxdy$ to its size in the new coordinate system, $JJ(r,\theta)drd\theta = rdrd\theta$.

This polar mapping operation can be neatly summarised by defining a new function,

$$gg(r,\theta) = \mathcal{P}\{ff(x,y)\}$$
$$= JJ(r,\theta) \cdot ff[x(r,\theta), y(r,\theta)] \tag{2.15}$$
$$= r \cdot ff[x(r,\theta), y(r,\theta)],$$

where the operator $\mathcal{P}\{\cdot\}$ changes variables and includes the effect of the Jacobian. Equation (2.11) then becomes

$$\int_x \int_y ff(x,y)dxdy \equiv \int_r \int_\theta gg(r,\theta)drd\theta \equiv \int_r \int_\theta \mathcal{P}\{ff(x,y)\}\,drd\theta\,. \tag{2.16}$$

The use of the *single* mapping operator to describe the coordinate transform minimizes the number of symbols required, and improves the clarity of the mathematics.

Table 2.1 contains the mapping operators used in this thesis and indicates where they are defined.

Table 2.1   Mapping operators used in the synthetic aperture processing algorithms in this thesis.

Operator	Description	Forward map	Inverse map	Definition
$\mathcal{C}\{\cdot\}$	2-D frequency/wavenumber rescaling		$(\omega_b, k_u) \rightarrow (k_x, k_y)$	(4.62)
$\mathcal{D}\{\cdot\}$	deramp/dechirp map	t or $t_s \rightarrow \omega_b$		(2.38), (4.69)
$\mathcal{F}\{\cdot\}$	Fourier transform	$t \rightarrow \omega$	$\omega \rightarrow t$	(2.9), (2.10)
$\mathcal{G}\{\cdot\}$	Slant-plane to ground plane conversion	$(x, y) \rightarrow (x_g, y_g)$		(4.129)
$\mathcal{H}\{\cdot\}$	2-D frequency/wavenumber rescaling		$(\omega_n, \Omega) \rightarrow (k_x, k_y)$	(4.124)
$\mathcal{K}\{\cdot\}$	1-D frequency/wavenumber rescaling	$k_x \rightarrow \omega$	$\omega \rightarrow k_x$	(2.27), (2.29)
$\mathcal{P}\{\cdot\}$	polar map	$(k_x, k_y) \rightarrow (\omega, u)$	$(\omega, u) \rightarrow (k_x, k_y)$	(4.82)
$\mathcal{P}_b\{\cdot\}$	baseband polar map	$(k_x, k_y) \rightarrow (\omega_b, u)$	$(\omega_b, u) \rightarrow (k_x, k_y)$	(4.79), (4.84)
$\mathcal{R}\{\cdot\}$	rotation operator		$(x_\theta, y_\theta) \rightarrow (x, y)$	(4.70)
$\mathcal{S}\{\cdot\}$	Stolt mapping	$(k_x, k_y) \rightarrow (\omega, k_u)$	$(\omega, k_u) \rightarrow (k_x, k_y)$	(4.36),(4.38)
$\mathcal{S}_b\{\cdot\}$	baseband Stolt mapping	$(k_x, k_y) \rightarrow (\omega_b, k_u)$	$(\omega_b, k_u) \rightarrow (k_x, k_y)$	(4.43)
$\mathcal{T}\{\cdot\}$	range migration correction		$(t, k_u) \rightarrow (x, k_y)$	(4.27)
$\mathcal{V}\{\cdot\}$	2-D frequency/wavenumber rescaling	$(\omega_n, \Omega) \rightarrow (\omega, k_u)$		(4.127)

Where appropriate, the mapping operators are defined such that the imaging system imposes the *forward mapping* and the processing algorithm undoes this mapping via the *inverse mapping*.

2.3   FOURIER TRANSFORMS FOR THE UNWARY

In this thesis, and often in synthetic aperture literature, the processing of synthetic aperture data is developed using Fourier theory via the continuous Fourier transform relationships given in (2.9) and (2.10). However, to develop a digital synthetic aperture processor requires the use of the fast Fourier transform (FFT). The differences between the continuous Fourier transform and the FFT are subtle, but they have an important impact on the phase functions appearing in the frequency domain filters used in many of the processors. This comment is especially true for the wavenumber algorithm developed in Section 4.3.3.

The following arguments are developed in 1-D, however, the same principles apply to 2-D FFTs. Due to its cyclic nature, an FFT of length N (where N is even) considers elements 0 to $N/2 - 1$ to correspond to positive temporal or frequency samples and elements $N/2$ to N to correspond to negative temporal or frequency samples. This ordering scheme is not how the data is recorded, and is not particularly desirable when trying to plot the temporal or frequency samples, so the samples are often left in the original or *logical* state and the data is reformatting or *shifted* to the FFT form prior to the FFT and then shifted back to the logical state after each implementation of the FFT or inverse FFT (iFFT). When using 2-D FFTs, this shifting operation is known as *quadrant shifting*. The mathematical

package MATLAB performs the correct 1-D and 2-D shifts via the function `fftshift` [101].

To explain the properties of the phase information obtained via an FFT relative to a continuous Fourier transform, is easier to consider that the data reformatting (shifting) is part of the FFT operation. To make this section clear, the shifting operation is referred to explicitly, however, through the rest of this thesis the term 'FFT' assumes that the shifting operations are implicit in the FFT calculation.

An FFT of a logically formatted sequence of data of length N considers element $(N/2 + 1)$ to be the axis origin (this element becomes the first element when the data is shifted), or alternatively this element can be interpreted as being the phase reference point, i.e., linear phase functions in the frequency/wavenumber domain, representing linear displacement in the time/spatial domain, are referenced to the $(N/2+1)$th element of the logically formatted data (this is why the origin is usually chosen as the center element—it gets the phase in the frequency/wavenumber domain right).

In strip-map or spotlight images, the raw data corresponds to echoes of the transmitted signal from locations out to *one side* of the imaging platform, eg., for a strip-map image, the valid data for processing are those temporal samples consistent with echoes from targets located within a swath of width X_s centered on some range r_0, i.e., the valid range data exists for times $t \in 2/c \cdot [r_0 - X_s/2, r_0 + X_s/2]$ (gating is sometimes used so that this is the only data collected). The implications of this one sided data on the mathematical description via a continuous Fourier transform versus the digital implementation via the FFT is now considered (here we will only consider a focused strip-map image and its spectrum, as this is adequate to point out the differences between the transforms).

If the reflectivity estimate of the imaged scene is given by the function $\widehat{ff}(x, y)$, then x is perpendicular to the platform path and corresponds to the range direction in a synthetic aperture image, while y is parallel to the platform path and corresponds to the along-track direction. The origin of the x-axis is defined from the platform location and the valid data exists out to one side of the platform for $x \in [r_0 - X_s/2, r_0 + X_s/2]$. The spectrum of this reflectivity function calculated via the continuous Fourier transform is

$$\widehat{FF}(k_x, k_y) = \mathcal{F}_{x,y} \left\{ \widehat{ff}(x, y) \right\} \tag{2.17}$$

(the functional notation is explained in more detail in Chapter 4). To digitally calculate this spectrum via the FFT requires that the x and y origins lie at the center of the logically formatted data, however, data has only been collected over a small swath *away* from the origin! The effect of this one sided data on processing via the FFT is best explained by way of an example. Figure 2.1(a) and (c) show a 10m by 10m simulated image 128×128 pixels wide. The image contains a single point target centered on $(x_0, y_0) = (5.39\text{m}, 0.39\text{m})$, i.e., centered on pixel (69,69). The bandwidth of the system is half that of the sampling frequency in either direction; for a sonar system, these parameters correspond to a sonar with an aperture of length $D = 0.3125\text{m}$ and chirp bandwidth $B_c = 4.8\text{kHz}$. The estimate of the target

reflectivity for this band limited point target is

$$\widehat{ff}(x,y) = \text{sinc}\left[\frac{B_{k_x}}{2\pi}(x - x_0)\right] \cdot \text{sinc}\left[\frac{B_{k_y}}{2\pi}(y - y_0)\right] \tag{2.18}$$

and the estimate of the target spectrum is

$$\widehat{FF}(k_x, k_y) = \text{rect}\left(\frac{k_x}{B_{k_x}}\right) \cdot \text{rect}\left(\frac{k_y}{B_{k_y}}\right) \cdot \exp(-jk_x x_0 - jk_y y_0) \tag{2.19}$$

where $B_{k_x} = B_{k_y} = 4\pi/D$. In Fig. 2.1(a), the negative x dimension of the image was zero padded with a 128×128 array of zeros so that the x origin was located in the center of the data (the y origin was already centered). The real part of the spectrum obtained via the FFT of this padded data is shown in Fig. 2.1(b), the sinusoidal oscillations of the real part of the exponential in (2.19) are clearly shown. To calculate the spectrum as defined in (2.17) it was necessary to *double* the size of the data set, in a more realistic imaging geometry, where the target is located a large distance from the x-origin, the padding required to center the x origin would far outweigh the amount of data collected!

To efficiently calculate the spectrum of the collected data, it is necessary to redefine the origin of the x-axis such that the center of the image corresponds to the origin of the new ordinate, i.e., $x' = x - r_0$ where $r_0 = 5$m is the scene center for the example in Fig. 2.1. This gives us the same reflectivity estimate with a modified coordinate system; that is,

$$\widehat{ff}(x', y) = \text{sinc}\left[\frac{B_{k_x}}{2\pi}(x' - x_0')\right] \cdot \text{sinc}\left[\frac{B_{k_y}}{2\pi}(y - y_0)\right] \tag{2.20}$$

where $x_0' = x_0 - r_0$ and the target location is now $(x_0', y_0) = (0.39\text{m}, 0.39\text{m})$. The estimate of the target spectrum then becomes

$$\widehat{FF}'(k_x, k_y) = \text{rect}\left(\frac{k_x}{B_{k_x}}\right) \cdot \text{rect}\left(\frac{k_y}{B_{k_y}}\right) \cdot \exp(-jk_x x_0' - jk_y y_0) \tag{2.21}$$

which is related to the original spectrum via

$$\widehat{FF}'(k_x, k_y) = \widehat{FF}(k_x, k_y) \cdot \exp(jk_x r_0). \tag{2.22}$$

The redefinition of the x-axis modifies the phase functions in the frequency domain such that they are now referenced to the point $(x, y) = (r_0, 0)$ instead of $(x, y) = (0, 0)$, thus in Fig. 2.1(d) the phase oscillations have a much lower frequency than those in Fig. 2.1(b). This lowering of the phase oscillations is extremely important when using interpolators to perform the mappings required in some of the synthetic aperture processing algorithms. Thus, the redefinition of the x-axis origin has two important

(a)

(b)

(c)

(d)
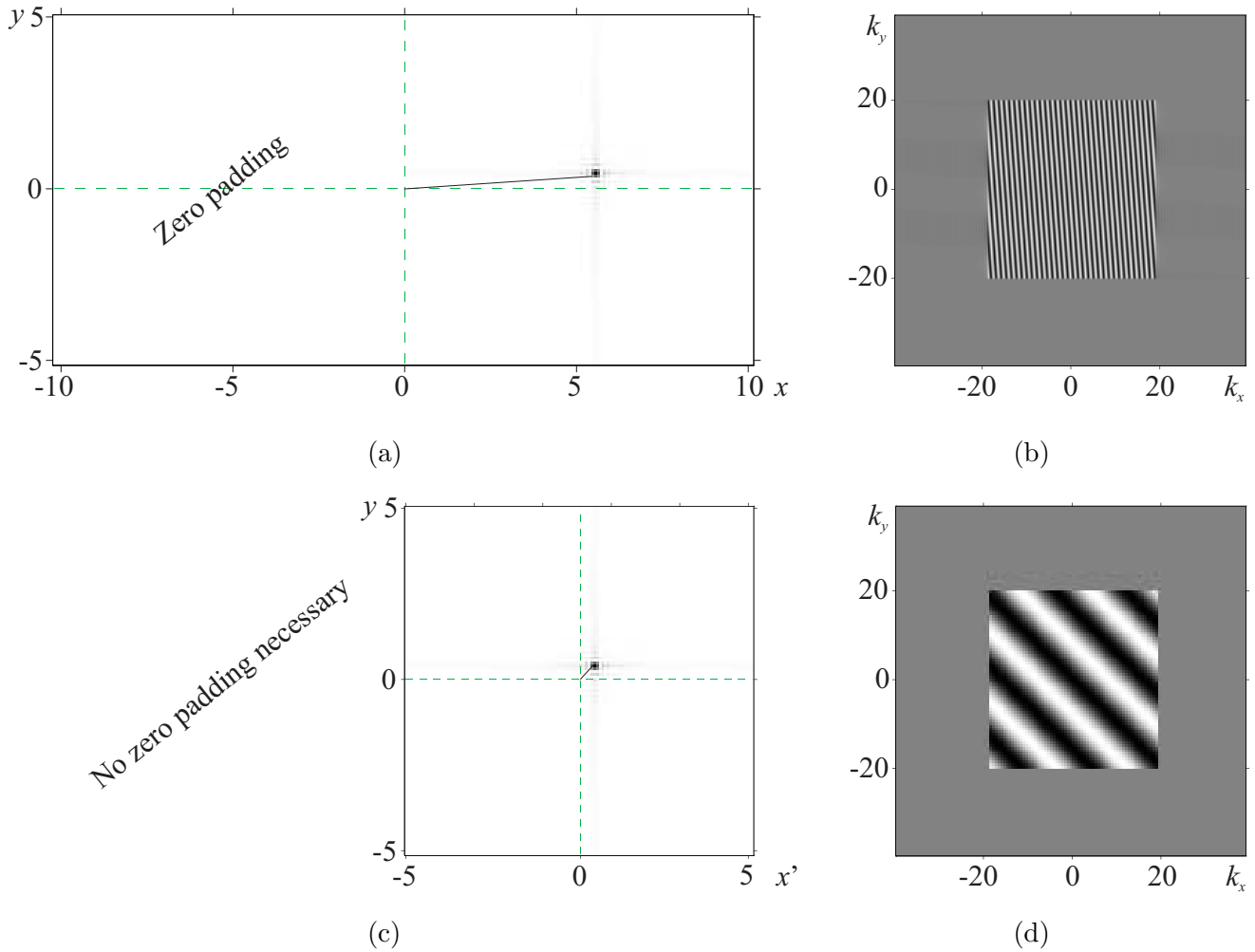
Figure 2.1 A point target simulation demonstrating the implications of the FFT. (a),(b) implementation of the continuous Fourier transform via the FFT; (a) the collected data $\widehat{ff}(x,y)$, zero padded such that the axis origin lies at the center of the data, (b) the real part of the image spectrum, real$\{\widehat{FF}(k_x,k_y)\}$. (c),(d) efficient calculation of the image spectrum by referencing phase to the center of the data instead of the axis origin; (c) the collected data with a redefined x-axis, $\widehat{ff}(x',y)$, (d) the real part of the redefined data spectrum, real$\{\widehat{FF}'(k_x,k_y)\}$. The dashed lines in (a) and (c) indicated the four quadrants that are shifted prior to the FFT, while (b) and (d) show the Fourier data obtained via the FFT after it has been shifted back into the logical format. The lines in (a) and (c) from the origin to the targets indicate the displacements that give rise to the phase functions in (b) and (d).

consequences; it represents the optimal method of calculating the image spectrum and obtains the best data for interpolation.

The relationship in (2.22) is analogous to generating baseband temporal data from modulated data, eg., the transmitted pulse

$$p_b(t) = p_m(t) \cdot \exp(j\omega_0 t). \tag{2.23}$$

If the echoes from the modulated pulse were sampled at the Nyquist rate (see Section 2.1.1), then the valid spectral data exists for frequencies $\omega \in [\omega_0 - \pi B_c, \omega_0 + \pi B_c]$. Data in the negative part of the spectrum, and from $\omega = 0$ to $\omega < \omega_0 - \pi B_c$, contains no useful information, therefore it is more efficient to produce a baseband signal at a lower effective sample rate (see Section 2.1.2). Alternatively, it is more efficient to use I-Q sampling and produce these samples directly (see Section 2.1.3) (the spatial analog to this last option is to gate the signal over the swath of interest).

2.4   PULSE COMPRESSION AND ONE-DIMENSIONAL IMAGING

Due to peak power limitations in both radar and sonar transmitters, the transmission of short, high power pulses is undesirable. In sonar, the additional problem of medium cavitation further precludes use of high power pulses. Initial developers of radar systems found that the effect of a short, high powered pulse could be obtained by spreading these wide bandwidth amplitude modulated (AM) pulses in time by modulating their carrier frequencies. This modulation allows the transmission of large bandwidth B_c (Hz) long pulses τ_c (s). These so called high *time-bandwidth product* $(\tau_c B_c)$ waveforms have the property that on reception, after pulse compression, the range resolution is dependent only on the signal bandwidth. The use of pulse compression also results in an increase in signal-to-noise (SNR) that is proportional to the time-bandwidth product of the transmitted pulse [31, p130]. Short pulses have a time-bandwidth product close to unity so their use results in no increase in SNR [31, Chapter 3].

The following subsections describe the commonly used methods of pulse compression and present a one-dimensional (1-D) imaging example showing its use. These 1-D examples are intended as a precursor to the two-dimensional (2-D) imaging techniques analysed in Chapter 4.

2.4.1   The matched filter or correlator

This was the most common method of pulse compression used in early SAR and SAS applications. The terminology still persists, however, the actual implementation of pulse compression is achieved in a variety of ways. The use of a matched filter decouples the transmitted pulse length from the range resolution equation, leaving the range resolution dependent only on the pulse bandwidth [31, p132]. The

process of pulse compression can be described as a correlation operation [28, Chapter 7] [31, p132] [129, p32]:

$$s(t) = \int_{-\infty}^{\infty} e(\tau)p^*(\tau - t)d\tau$$
$$= \mathcal{F}_{\omega}^{-1}\left\{E(\omega)P^*(\omega)\right\},$$

(2.24)

where $s(t)$ is the complex (modulated or baseband) output of the correlator (the pulse compressed signal), $e(t)$ is the complex received signal, $p(t)$ is the complex form of the transmitted pulse and $E(\omega)$ and $P(\omega)$ are the modulated or baseband spectrums. The frequency domain representation of the correlation operation in (2.24) is the operation typically used in a digital implementation of pulse compression. However, the digital implementation has a subtle difference; it is often only the phase of the transmitted pulse spectrum $P(\omega)$ that is used for pulse compression operation. Spectral shaping is then done using a window or weighting function at a later stage in the processing algorithm.

Consider a 1-D imaging situation where time t and range x are defined from the same origin and the function $f(x)$ describes the frequency and aspect independent target reflectivity function. The received signal for a modulated pulse $p_m(t)$; that is,

$$e_m(t) = \int_x f(x)\, p_m\left(t - \frac{2x}{c}\right) dx,$$

(2.25)

has the temporal Fourier transform

$$E_m(\omega) = P_m(\omega) \cdot \int_x f(x)\, \exp\left(-j\omega\frac{2x}{c}\right) dx$$
$$= P_m(\omega) \cdot \int_x f(x)\, \exp(-j2kx)dx$$
$$= P_m(\omega) \cdot \mathcal{K}\left\{F(k_x)\right\}$$
$$= P_m(\omega) \cdot \frac{2}{c} \cdot F[k_x(\omega)],$$

(2.26)

where wavenumber $k = \omega/c$ and $F(k_x)$ is the *spatial* Fourier transform of $f(x)$. The operator $\mathcal{K}\left\{\cdot\right\}$ describes the following coordinate transform that is induced by the measurement system:

$$\omega(k_x) = \frac{k_x c}{2} + \omega_0.$$

(2.27)

This coordinate transform has the Jacobian $J(\omega) = dk_x/d\omega = 2/c$. An estimate of the targets' *baseband* wavenumber domain can be obtained by matched filtering and inverse transforming the received signal;

that is,

$$
\begin{aligned}
\widehat{F}(k_x) &= \mathcal{K}^{-1}\left\{S_m(\omega)\right\} \\
&= \mathcal{K}^{-1}\left\{P_m^*(\omega)E_m(\omega)\right\} \\
&= \frac{c}{2}\cdot|P_m[\omega(k_x)]|^2\cdot\frac{2}{c}\cdot F(k_x) \\
&= \operatorname{rect}\left(\frac{k_x}{4\pi B_c/c}\right)\cdot F(k_x),
\end{aligned}
\tag{2.28}
$$

where the pulse compressed spectrum $|P_m(\omega)|^2$ has unity magnitude over the *modulated* bandwidth $\omega \in [\omega_0 - \pi B_c, \omega_0 + \pi B_c]$, and the inverse operator $\mathcal{K}^{-1}\{\cdot\}$ performs the coordinate transform:

$$
k_x(\omega) = \frac{2(\omega - \omega_0)}{c},.
\tag{2.29}
$$

producing baseband spatial wavenumbers $k_x \in 2\pi/c \cdot [-B_c, B_c]$. This coordinate transform has the Jacobian $J(k_x) = d\omega/dk_x = c/2$. The wavenumber estimate in (2.28) is then inverse *spatial* Fourier transformed and *detected* (modulus operator) to give an estimate of the target reflectivity:

$$
\begin{aligned}
|\widehat{f}(x)| &= \left|\mathcal{F}_{k_x}^{-1}\left\{\widehat{F}(k_x)\right\}\right| \\
&= \left|\frac{2B_c}{c}\cdot\operatorname{sinc}\left(\frac{2B_c}{c}\cdot x\right)\odot_x f(x)\right|.
\end{aligned}
\tag{2.30}
$$

The complex matched filtered image estimate $\widehat{f}(x)$ is seen to be a band-limited or smoothed version of the target function $f(x)$. Range sidelobes due to the sinc-function in (2.30) are usually suppressed by weighting the Fourier domain before performing the inverse spatial Fourier transform. The range resolution of each target in this weighted image estimate is $\delta x_{3\mathrm{dB}} = \alpha_w c/(2B_c)$, where α_w reflects the resolution loss due to weighting. The development and analysis of the matched-filter concept is covered in detail in Cook and Bernfeld [28].

2.4.2 Spectral synthesis

Sato, 1977 [136] presents a method of pulse compression for a synthetic aperture sonar that he calls *spectral synthesis*. This is an impractical method of pulse compression, but its description is useful for the interpretation of *deramp* or *dechirp* processing that is covered in the next section.

The sampled baseband spectrum of a pulse compressed signal is

$$
S_s(\omega_b) = \sum_n S_b(n\Delta_\omega)\cdot\delta(\omega_b - n\Delta_\omega).
\tag{2.31}
$$

Sato observed that this signal could be formed or "built-up" in the time-domain via the transmission of a multiplicity of sinusoidal signals each of baseband frequency $\omega_n = n\Delta_\omega$. To see how this works, consider the transmission of the sinusoid $p_m(t) = \exp[j(\omega_0 + \omega_n)t]$ towards a scene of target reflectivity $f(x)$. The demodulated return of this signal is associated with the baseband spectral sample via

$$
\begin{aligned}
S_b(\omega_n) &= \left(\int_x f(x) p_m \left(t - \frac{2x}{c} \right) dx \right) \cdot \exp[-j(\omega_0 + \omega_n)t] \\
&= \int_x f(x) \exp[-j2(k_0 + k_n)x] dx,
\end{aligned}
\tag{2.32}
$$

where $k_n = \omega_n/c$. If the baseband set of transmitted sinusoids is limited to $\omega_n \in [-\pi B_c, \pi B_c]$, the sampled spectrum is

$$
\begin{aligned}
S_s(\omega_b) &= \operatorname{rect}\left(\frac{\omega_b}{2\pi B_c} \right) \cdot \sum_n \left\{ \int_x f(x) \exp\left[-j(\omega_0 + n\Delta_\omega)\frac{2x}{c} \right] dx \right\} \cdot \delta(\omega_b - n\Delta_\omega) \\
&= \operatorname{rect}\left(\frac{\omega_b}{2\pi B_c} \right) \cdot \int_x f(x) \exp\left(-j2kx \right) dx \cdot \sum_n \delta(\omega_b - n\Delta_\omega).
\end{aligned}
\tag{2.33}
$$

The reflectivity estimate obtained from this synthesized spectrum,

$$
\begin{aligned}
\left| \widehat{f}(x) \right| &= \left| \mathcal{F}_{k_x}^{-1} \left\{ \mathcal{K}^{-1} \left\{ S_s(\omega_b) \right\} \right\} \right| \\
&= \left| \frac{2B_c}{c} \cdot \operatorname{sinc}\left(\frac{2B_c}{c} \cdot x \right) \odot_x f(x) \odot_x \frac{\pi c}{\Delta_\omega} \cdot \sum_n \delta\left(x - \frac{\pi c}{\Delta_\omega} \right) \right|,
\end{aligned}
\tag{2.34}
$$

is a scaled, repeated version of (2.30). If the inverse Fourier transform is performed using the iFFT, then the samples obtained for $\widehat{f}(x)$ in (2.34) represent $x \in [-\pi c/(2\Delta_\omega), \pi c/(2\Delta_\omega)]$. Thus, the frequency spacing of the transmitted sinusoids sets the maximum range over which this technique works.

Sato [136] synthesized a signal with 1MHz bandwidth about a 1.5MHz carrier, producing a signal with a quality factor of Q = 1.5 (equivalent to the Kiwi-SAS, but at a much higher carrier frequency). This bandwidth corresponds to a range resolution of 0.75mm or a temporal resolution of 1μs. The generation of an equivalently short pulse would be difficult using any other method [136]. Spectrum synthesis in a synthetic aperture system requires repeated passes over the same target scene. This was easily achieved by Sato with his tank set-up, however, this method is impractical for a synthetic aperture system such as an airborne SAR or an ocean-going SAS.

2.4.3 Deramp or dechirp processing

Radar systems that transmit LFM or *chirp* signals with a high carrier to bandwidth ratio (high quality factor, Q) employ a very efficient method of pulse compression known as *deramp* [31,82] or *dechirp* [21]

processing. In deramp processing, the received signal is not passed through a matched filter, it is instead multiplied (mixed) with a delayed version of the transmitted LFM pulse. This demodulation operation results in an output signal that is associated with the *spectrum* of the pulse compressed pulse. In summary, the mixing operation employed by deramp processing *simultaneously* performs a Fourier transform and phase matching of the LFM signal.

The modulated complex form of the transmitted LFM chirp is

$$p_m(t) = \text{rect}\left(\frac{t}{\tau_c}\right) \cdot \exp(j\omega_0 t + j\pi K_c t^2), \tag{2.35}$$

where $K_c = B_c/\tau_c$ (Hz/s) is the *chirp rate*, B_c is the *chirp bandwidth* and τ_c is the *chirp length*. The properties of LFM are discussed in Section 2.5.1. The received reflected chirp signal is quadrature demodulated (*dechirped*) by mixing the signal with a reference signal representing the return from a hypothetical target at a range r_0 [82, p21 and p359]. In complex signal notation, this mixing operation is [31, p505]

$$
\begin{aligned}
e_b(t) &= \left(\int_x f(x) p_m\left(t - \frac{2x}{c}\right) dx\right) \cdot p_m^*\left(t - \frac{2r_0}{c}\right) \\
&= \left(\int_x f(x) p_m\left(t - \frac{2x}{c}\right) dx\right) \cdot \text{rect}\left(\frac{t - 2r_0/c}{\tau_g}\right) \cdot \exp\left[-j\omega_0\left(t - \frac{2r_0}{c}\right) - j\pi K_c\left(t - \frac{2r_0}{c}\right)^2\right] \\
&= \text{rect}\left(\frac{t - 2r_0/c}{\tau_g}\right) \cdot \int_x f(x) \exp\left\{-j\frac{2}{c}\left[\omega_0 + 2\pi K_c\left(t - \frac{2r_0}{c}\right)\right](x - r_0) + \frac{4\pi K_c}{c^2}(x - r_0)^2\right\} dx,
\end{aligned}
\tag{2.36}
$$

where τ_g is the *range gated* signal length which is discussed shortly. Assuming for the moment that the quadratic term known as the *residual video phase* can be ignored, the mixing operation can be shown to give directly samples of the pulse compressed spectrum:

$$
\begin{aligned}
S_b(\omega_b) &= \mathcal{D}\{e_b(t)\} \\
&= \frac{1}{2\pi K_c} \cdot \text{rect}\left(\frac{\omega_b}{2\pi K_c \tau_g}\right) \cdot \exp(j2kr_0) \cdot \int_x f(x) \exp(-j2kx) dx
\end{aligned}
\tag{2.37}
$$

The dechirp or deramp operator $\mathcal{D}\{\cdot\}$ performs the mapping:

$$\omega_b(t) \equiv 2\pi K_c\left(t - \frac{2r_0}{c}\right), \tag{2.38}$$

which has the Jacobian $J(\omega) = 1/(2\pi K_c)$.

The reflectivity estimate obtained from the deramped data is

$$
\begin{aligned}
\left|\widehat{f}(x)\right| &= \left|\mathcal{F}_{k_x}^{-1}\left\{\mathcal{K}^{-1}\left\{S_b(\omega_b)\right\}\right\}\right| \\
&= \left|\frac{\tau_g}{\pi c}\cdot \operatorname{sinc}\left(\frac{2K_c\tau_g}{c}\cdot x\right)\odot_x f(x)\odot_x \delta(x-r_0)\right|,
\end{aligned}
\tag{2.39}
$$

which is a scaled, band-limited version of the target reflectivity. The term $\delta(x-r_0)$ exists because range x is defined from the time origin; it is usually removed by redefining the range origin to the reference range r_0.

Deramped processing is popular in the SAR community due to its low sampling requirements. The sampling of the deramped data occurs at a much lower rate than normal quadrature baseband sampling requires. The process of deramping shown in (2.36) produces a baseband sinusoidal tone that is proportional to the distance of a particular target from the deramp reference range, i.e., $\exp[-j4\pi K_c(x-r_0)/c]$. In previous sonar applications this proportionality has been used to develop an aural output [39]. If the swath or *patch* being imaged is given by $x \in [r_0 - X_s/2, r_0 + X_s/2]$, where X_s is the swath (patch) width, then the required sampling rate is twice the highest baseband sinusoid expected:

$$
\begin{aligned}
\omega_s &= 2\cdot\frac{4\pi K_c}{c}\cdot\frac{X_s}{2} \\
&= \frac{4\pi K_c X_s}{c} \\
&= 2\pi B_c\cdot\frac{\tau_p}{\tau_c}.
\end{aligned}
\tag{2.40}
$$

The *patch time* $\tau_p = 2X_s/c$ and the chirp time are shown to emphasize the fact that when a small swath is being imaged and long chirps are employed ($\tau_p \ll \tau_c$), the required sampling rate is much lower than the normal complex baseband requirement of $\omega_s = 2\pi B_c$.

The range resolution for pulse compression via deramping; that is,

$$
\delta x_{\mathrm{3dB}}^{\mathrm{deramp}} = \alpha_w\frac{c}{2K_c\tau_g} =, \alpha_w\frac{c}{2B_{\mathrm{eff}}}
\tag{2.41}
$$

is controlled by the time gating of the signal. Two options are commonly employed; the method employed by Jakowatz *et al* [82] is to gate time for $t \in [2r_0/c - \tau_g/2, 2r_0/c + \tau_g/2]$ where $\tau_g = \tau_c - \tau_p$ removes the effect of *range skewing* that is caused by the relative delays of the targets across the patch (for more details see p396 [82]). This form of time gating gives a range resolution with an *effective bandwidth* of $B_{\mathrm{eff}} = K_c(\tau_c - \tau_p)$ which is close to B_c as $\tau_p \ll \tau_c$. By employing range *deskewing* or *deskew correction* it is possible to correct for the skew effect and use the entire chirp bandwidth for processing. This technique is covered in Appendix C.2 of Carrara *et al* [21]. Range deskewing also mitigates the effect of the quadratic residual video phase term in (2.36) [21, p505].

Further limitations of deramp processing are discussed in the next section, and when developing the tomographic formulation of the spotlight mode in Chapter 4. Chapter 10 of Curlander and McDonough has a further discussion on deramp processing [31]. Pulse compression via deramp processing is the method of choice for most radar altimeters and spotlight mode SARs [65].

2.4.4 The step transform

Complications arise in deramp processing when the swath width increases to the point that the patch time becomes comparable to, or larger than, the chirp length. The deramp sampling rate in (2.40) then exceeds the sampling rate required by standard quadrature demodulation, and the returns (echoes) from targets away from the reference range are not fully captured by the signal processor. The *step transform* is a variant of deramp processing that uses multiple reference range chirps. The full range swath is divided into subswaths such that each subswath width is smaller than the length of one of the multiple reference range chirps. In this way, the full returns from every target are captured by the signal processor and the sampling requirements of each subpatch are reduced. The analysis of the step transform and its use in range and along-track compression is covered in detail in Chapter 10 of Curlander and McDonough [31].

2.4.5 CTFM processing

The previous Canterbury-SAS processor performed pulse compression using the continuous transmission, frequency modulation (CTFM) dual-demodulation scheme described in Gough *et al* [60], de Roos *et al* [39] and Hayes and Gough [74]. In the same way as deramp and the step transform operate, this dual demodulation scheme produces samples of the echo spectrum directly. In references [39, 62, 63, 74], once the dual-demodulation had been performed, the resulting signal was forward Fourier transformed and range resolution was inferred from an analysis of the signal spectrum.

2.5 WAVEFORMS COMMONLY EMPLOYED IN SYNTHETIC APERTURE SYSTEMS

2.5.1 Linear FM (chirp) waveforms

The linear frequency modulated (LFM) waveform is the most commonly used waveform in SAR and SAS processing. If a LFM pulse is produced at audio frequencies over a time period of say, one second, the resulting sound is a chirp, thus these waveforms are commonly referred to as chirp waveforms. The

usual complex modulated form of a LFM waveform is [129, Chapter 7]

$$p_m(t) = \text{rect}\left(\frac{t}{\tau_c}\right) \cdot \exp(j\omega_0 t + j\pi K_c t^2). \tag{2.42}$$

Differentiation of the phase of the chirp gives an instantaneous radian frequency of

$$\omega_i(t) = \frac{d\phi(t)}{dt} = \omega_0 + 2\pi K_c t, \tag{2.43}$$

which is a linear function of time where ω_0 (rad/s) is the carrier radian frequency and K_c (Hz/s) is the LFM chirp-rate. The rect function in (2.42) limits the chirp length to $t \in [-\tau_c/2, \tau_c/2]$ giving a linearly swept FM signal over the instantaneous frequencies $\omega_i(t) \in [\omega_0 - \pi B_c, \omega_0 + \pi B_c]$ where the chirp bandwidth is $B_c = K_c \tau_c$ (Hz).

Using the principle of stationary phase [28, p39](see Appendix A), the approximate form of the Fourier transform of the modulated waveform in (2.42) is

$$P_m(\omega) \approx \text{rect}\left(\frac{\omega - \omega_0}{2\pi B_c}\right) \cdot \sqrt{\frac{j}{K_c}} \cdot \exp\left[-j\frac{(\omega - \omega_0)^2}{4\pi K_c}\right], \tag{2.44}$$

where the approximation is good for the large time-bandwidth pulses used in synthetic aperture imaging (the exact form of this spectrum involves Fresnel integrals [28]).

If a linear FM signal was transmitted in the one-dimensional range imaging example described by (2.28) and (2.30), and the matched filtering operation that forms $|P_m(\omega)|^2$ is replaced with the more common implementation of a *phase only* matched filtering, the image estimate becomes

$$\begin{aligned} |\widehat{f}(x)| &= \left| \mathcal{F}_{k_x}^{-1}\left\{ \mathcal{K}^{-1}\left\{ \exp\left[j\frac{(\omega - \omega_0)^2}{4\pi K_c} \right] \cdot E_m(\omega) \right\} \right\} \right| \\ &= \left| \frac{2}{c} \cdot \frac{B_c}{\sqrt{K_c}} \cdot \text{sinc}\left(\frac{2B_c}{c} \cdot x\right) \odot_x f(x) \right|. \end{aligned} \tag{2.45}$$

The *improvement factor* due to pulse compression of LFM given by

$$\text{IF}_{\text{pc}} = 20\log_{10}\left(\frac{B_c}{\sqrt{K_c}}\right) = 10\log_{10}(B_c \tau_c), \tag{2.46}$$

represents the increase in SNR due to pulse compression. The pulse compression factor can also be interpreted as the ratio of the length of the uncompressed pulse (τ_c) to the length of the compressed pulse ($1/B_c$). Note that the number of processor bits per digitized signal sample (processor word size) must eventually accommodate the full dynamic range afforded by the improvement factors due to pulse and along-track compression and the sampled signal level.

When a weighting function is used to minimize range sidelobes, it is also necessary to account for the effect of the *coherent gain* of the weighting function on the scaling of the image estimate. For example, if the image estimate is formed using a Hamming weighted spectral estimate, the image magnitude is scaled by 0.54, and the image range resolution is $\delta x_{3\text{dB}} = 1.30c/(2B_c)$ [71].

2.5.2   Hyperbolic (Linear period) FM

The modulated complex form of a hyperbolic frequency modulated (HFM) waveform is [129, p420]

$$p_m(t) = \text{rect}\left(\frac{t}{\tau_h}\right) \exp\left\{-j\frac{\omega_c}{K_\infty} \ln(1 - K_\infty t)\right\}. \tag{2.47}$$

Differentiation of the phase function yields an instantaneous frequency of

$$\omega_i(t) = \frac{\omega_c}{1 - K_\infty t}, \tag{2.48}$$

which is a hyperbolic function of time where ω_c plays the part of the 'carrier', and K_∞ (s^{-1}) sets the infinite frequency point at time $t = 1/K_\infty$. HFM is also known as linear period FM as the instantaneous period of the pulse, $T_i(t) = 2\pi/\omega_i(t)$, is linear with time, and due to its temporal Doppler tolerance, HFM is sometimes referred to as Doppler invariant FM [70].

The rect function limits time to $t \in [-\tau_h/2, \tau_h/2]$ where the HFM length is τ_h. Instantaneous frequency then sweeps over $\omega_i(t) \in [\omega_c/(1 + K_\infty\tau_h/2), \omega_c/(1 - K_\infty\tau_h/2)]$. By defining the required lower and upper frequencies, ω_a and ω_b, it is possible to characterize the pulse parameters (bandwidth, 'carrier' and infinite frequency point) via

$$B_h = \frac{\omega_b - \omega_a}{2\pi}, \qquad \omega_c = \frac{2\omega_a\omega_b}{\omega_a + \omega_b}, \qquad K_\infty = \frac{2}{\tau_h}\left(\frac{\omega_b - \omega_a}{\omega_b + \omega_a}\right). \tag{2.49}$$

Using the principle of stationary phase, the approximate spectrum $P_m(\omega)$ is

$$P_m(\omega) \approx \text{rect}\left(\frac{\omega - \omega_c}{K_\infty\tau_h\omega}\right) \cdot \frac{2\pi}{\omega}\sqrt{\frac{j\omega_c}{2\pi K_\infty}} \cdot \exp\left(-j\frac{\omega_c}{K_\infty}\left[\ln\left(\frac{\omega_c}{\omega}\right) + \frac{\omega}{\omega_c} - 1\right]\right). \tag{2.50}$$

Equation (2.50) shows that the magnitude of the HFM spectrum is amplitude modulated by a $2\pi/\omega$ term and that the envelope function $\text{rect}(t/\tau_h)$ undergoes a non-linear distortion during the transformation from the time to the frequency domain. This non-linear distortion does not affect the rect function, however, if the time domain signal was transmitted with an amplitude modulation given by a weighting function, then the weighting function is distorted in the frequency domain. The $2\pi/\omega$ weighting gives the lower frequencies in the spectrum of a HFM waveform a larger magnitude than the

higher frequencies. This non-symmetric and non-uniform spectral shape causes the range response in the pulse compressed HFM waveform to have higher range sidelobes than if the unweighted spectrum was uniform.

If a HFM signal was transmitted in the one-dimensional range imaging example described by (2.28) and (2.30), and the matched filtering operation that forms $|P_m(\omega)|^2$ is replaced with a *deconvolution* filter, the image estimate becomes

$$
\begin{aligned}
|\widehat{f}(x)| &= \left| \mathcal{F}_{k_x}^{-1} \left\{ \mathcal{K}^{-1} \left\{ \operatorname{rect}\left(\frac{\omega - \omega_c}{B_h}\right) \cdot \frac{\omega}{\omega_c} \cdot \exp\left(j\frac{\omega_c}{K_\infty} \left[\ln\left(\frac{\omega_c}{\omega}\right) + \frac{\omega}{\omega_c} - 1\right]\right) \cdot E_m(\omega) \right\} \right\} \right| \\
&= \left| \frac{2}{c} \cdot B_h \sqrt{\frac{j2\pi}{\omega_c K_\infty}} \cdot \operatorname{sinc}\left(\frac{2B_h}{c} \cdot x\right) \odot_x f(x) \right|.
\end{aligned}
\tag{2.51}
$$

The deconvolution filter matches the phase and generates a uniform spectrum of height $\sqrt{2\pi/(\omega_c K_\infty)}$ across the bandwidth B_h, inverse Fourier transformation of this limited uniform spectrum increases the height by B_h. Thus, the *improvement factor* due to pulse compression of HFM (using the deconvolution filter) is

$$
\mathrm{IF}_{\mathrm{pc}} = 20 \log_{10}\left(B_h \sqrt{\frac{2\pi}{\omega_c K_\infty}} \right).
\tag{2.52}
$$

Range sidelobes can again be suppressed using spectral weighting techniques. Instead of using a deconvolution filter, a uniform pulse spectrum is obtained if the pulse envelope is shaped in the time-domain [91]. This amplitude modulation of the transmitted pulse is not always desirable (see the discussion in Section 2.6).

Many radar systems employ bandwidths of many MHz, so they are referred to as wide bandwidth systems. Often this wide bandwidth signal is modulated to a carrier frequency of several GHz, so the carrier frequency-to-bandwidth ratio or quality factor, Q, of the system is high. These high-Q radar systems typically transmit LFM chirps and refer to them as Doppler tolerant waveforms. To determine why a LFM chirp transmitted from a high-Q system can be considered Doppler tolerant, it is necessary to expand the logarithm in the phase of the HFM waveform in (2.47) (the high-Q assumption implies $\omega_c/2\pi B_h \gg 1$). The expansion of the HFM phase (given on p423 of reference [129]) results in a LFM waveform with a carrier frequency $\omega_0 = \omega_c$ and a chirp rate $K_c = B_h/\tau_h$ [129, p423]. So the LFM waveform transmitted by the high-Q SAR can also be seen as the transmission of a small segment of a Doppler tolerant HFM waveform. This property explains why LFM is Doppler tolerant for high-Q SAR applications. (In the sense that the transmitted signal bandwidth is much less than the carrier frequency, high-Q systems can also be referred to as *narrow bandwidth* systems).

2.5.3   Phase-coded waveforms

Linear FM and hyperbolic FM belong to a wider class of waveforms known as phase modulated waveforms. The popular use of LFM and HFM in mapping or civilian applications is due in part to the fact that they are relatively easy to produce. Phase modulation using binary phase codes is a technique generally employed by military radar systems in an attempt to lower their probability of detection. The flexibility of digital systems means that the use of coded pulses for mapping is also a viable option. In a digital application, these waveforms are easy to generate, and easy to process. The transmission of multiple orthogonal codes can also be used as a method for increasing the effective along-track sampling rate of a synthetic aperture system, however, this is at the expense of degraded image quality due to the effects of self clutter (see the discussion in Chapter 5). The binary phase modulation schemes commonly used in SAR applications include; Barker codes, pseudo-random noise sequences (PRN), maximal length sequences (m-sequences), Huffman codes, and polyphase codes. The analysis of many of these phase codes is detailed in Chapter 8 of Cook and Bernfeld, 1967 [28] and in references [13, 129, 142].

2.6   PRACTICAL WAVEFORM CONSIDERATIONS

Range sidelobe levels in the pulse compressed data are directly related to the amplitude of the compressed pulsed spectrum [28, p174]. If a *matched filter* (i.e., including any amplitude effects) is used for pulse compression, then the pulse compressed spectrum is also the *energy spectrum* of the transmitted pulse. To obtain low sidelobe levels it is usual to add a weighting function to either the transmitted waveform, or apply a weighting function in the matched filter (or a combination of both). The pulse compressed spectra throughout this thesis are obtained via *phase only* matched filtering, or alternatively via *deconvolution* of the echo spectrum. The transmitted pulse is always a LFM waveform with a uniform transmission level in the time domain. The spectra produced after filtering ideally have uniform levels. Hamming weighting is then applied to reduce the characteristic sinc sidelobes in the image domain down to an acceptable level of -43dB [71]. See Chapter 7 of Cook and Bernfeld [28] for a detailed discussion of window functions and their effects (and transforms) in the time (space) and frequency (wavenumber) domains.

LFM waveforms are typically transmitted *without* amplitude weighting, this is because it is preferable to operate the transmit power amplifiers at a constant level. The weighting function is then applied on reception, during pulse compression [67] [28, p176,p189]. An additional method of reducing range sidelobes is to transmit non-linear FM waveforms (NLFM). Using a waveform with LFM over the majority of the pulse with small portions of higher FM rate signal at the beginning and end of the pulse produces a pulse compressed waveform with extremely low sidelobes [67].

In radar applications, an ideal broadband transmission channel is usually assumed. This assumption

is not valid for wide bandwidth sonar. The bandpass characteristic of most ultrasonic transducers limits the achievable pulse compression SNR gain of the sonar system. Because of this bandpass characteristic, some sonar applications require techniques that determine the 'optimum bandwidth' appropriate for the transducer within the sonar system. When a transducer has wideband characteristics (i.e. has a constant amplitude, linear phase transfer function over the transmitted signal bandwidth) the pulse-compression time-bandwidth gain is monotonic with increasing signal bandwidth. For the case of a bandpass transducer, the time-bandwidth gain has a maximum. Increasing the system bandwidth past this maximum does not increase the pulse compression gain [125]. Similarly, the transmitted waveform can be designed to achieve maximum power transmitted from the transducer. By a process of amplitude modulation and non-linear frequency modulation (AM-NLFM) the transmitted signal can drive the transducer to its peak power to produce a sound pressure waveform into the water with a desired power spectrum and maximum possible energy [124, 162].

Some radar systems are also adopting similar principles. Instead of modifying the transmitted pulse, it is the matched filter which is modified. By storing a replica of each transmitted pulse it is possible to generate an inverse filter that matches the phase modulation and *deconvolves* phase and amplitude errors [109, 112]. Chapter 11 of Cook and Bernfeld [28] discusses the effects of distortion on matched-filtered signals and other methods used for distortion compensation.

A further predictable waveform distortion is that caused by temporal Doppler. The transmitted and received pulse modulations are modified by the changing platform-target range during pulse transmission and reception. The effects of this distortion is described in the next section.

2.7 TEMPORAL DOPPLER EFFECTS—AMBIGUITY FUNCTIONS

Temporal Doppler causes a distortion of the modulation function *within* the pulses employed by SAR and SAS systems. The *spatial Doppler* shifts exploited by the imaging algorithms occur due to Doppler effects *between* pulses. If the relative platform-target range rate can be considered to be constant *over the length of the pulse*, then the velocity tolerances of the pulse need to be considered. If, however, the range rate can not be considered constant over the pulse length, then the acceleration tolerance of the pulse must also be investigated. Higher order platform-target range derivatives are usually not important in mapping applications.

The ambiguity surface or ambiguity function gives a measure of the correlator (matched filter) response to a return (echo) signal that is mismatched from the reference signal in delay and temporal Doppler, i.e., the ambiguity function indicates first order *range rate* effects.

If the transmitted waveform can be considered to be Doppler tolerant, then the receiver complexity drops as only one matched filter is required for pulse compression. The use of Doppler tolerant waveforms is useful in moving target identification and range velocity imaging. However, for the imaging

algorithms discussed in this thesis, the targets are not moving and so the Doppler shifts within pulses are deterministic (given the platform velocity). This deterministic nature allows any expected Doppler distortion to be corrected for during processing (if it is necessary).

2.7.1   Wideband formulation

The wideband ambiguity function (WAF) is defined as [99]

$$
\begin{aligned}
\chi_w(\eta, \tau) &= \sqrt{\eta} \int p(t) p^* \left[\eta(t - \tau) \right] dt \\
&= \frac{1}{\sqrt{\eta}} \int P(f) P^* \left(\frac{f}{\eta} \right) \exp(j 2\pi f t) df,
\end{aligned}
\tag{2.53}
$$

where τ is a delay relative to the platform-target range and the Doppler scale factor is

$$
\eta \approx 1 - \frac{2\dot{R}}{c},
\tag{2.54}
$$

where $\dot{R} = dR(t)/dt$ is the relative platform-target range rate and the approximation is valid for both sonar and radar mapping applications. Note that ambiguity functions are sometimes defined with the time-axis reversed, i.e. $\chi'_w(\eta, \tau) \equiv \chi_w(\eta, -\tau)$ [31, p134].

2.7.2   Narrowband formulation

The narrowband (Woodward) ambiguity function (NAF) is defined as [129, p119]

$$
\begin{aligned}
\chi_n(\nu, \tau) &= \int p(t) p^*(t - \tau) \exp(j 2\pi \nu t) dt \\
&= \int P^*(f) P(f - \nu) \exp(j 2\pi f \tau) df,
\end{aligned}
\tag{2.55}
$$

where the effect of a constant range rate is approximated by a simple frequency translation of the transmitted signal (i.e., carrier shift); that is,

$$
\nu \approx \frac{2\dot{R}}{c} f_0,
\tag{2.56}
$$

in essence neglecting the Doppler distortions of the modulation function [70].

2.7.3 Ambiguity function analysis and interpretation

The analysis of an ambiguity function is usually performed with reference to an ambiguity diagram which is the magnitude, or magnitude squared of the ambiguity function plotted on the $(\eta\text{-}\tau)$ or $(\nu\text{-}\tau)$ planes as contours or as a three-dimensional surface. The actual derivation of these diagrams is generally performed numerically, although analytic studies of wide bandwidth LFM and HFM waveforms have been performed using the principle of stationary phase [90, 99]. Narrowband ambiguity analyses for a number of waveforms can be found in [28, 129].

Additional insight as to the effects of the filter mismatch between the Doppler distorted received pulse and the matched filter (reference pulse) can be obtained by plotting the frequency-time characteristics of the two waveforms on the same $f\text{-}t$ plot [90]. Correlator (matched filter) loss can then be explained in terms of signal overlap loss (due to translations), signal slope differences and points of intersection (due to FM rate changes).

2.7.3.1 Ambiguity analysis of Linear FM

Radar analyses of the effect of a constant range rate and thus temporal Doppler effects on a LFM pulse are typically performed using the NAF. The effect of temporal Doppler on a LFM waveform is to shift the carrier and alter the sweep rate. In a narrowband analysis this modulation distortion is approximated as a simple carrier shift. This carrier shift has two effects; it causes the signals to correlate with a slight delay error, and the $f\text{-}t$ lines of the two pulses no longer overlay perfectly which results in overlap loss (loss of correlator magnitude). A wideband analysis of the temporal Doppler modulation distortion in LFM is necessary for sonar applications. An analysis of the WAF for wide bandwidth LFM shows that the actual temporal Doppler tolerance is closely governed by the *slope* difference between the received and reference $f\text{-}t$ characteristic lines [70, 90]. As a result, wideband LFM sonars have temporal Doppler tolerances far smaller than conventional narrow-band theory predicts [70, 90].

A plot of both the narrowband and wideband ambiguity diagrams with contours at $|\chi_n(\nu, \tau)|^2 = 0.5$ and $|\chi_w(\eta, \tau)|^2 = 0.5$ allows the range rate tolerances to be calculated as [77, 90]

$$|\dot{R}^n| \leq 0.15c \cdot \frac{2\pi B_c}{\omega_0} \qquad \text{and} \qquad |\dot{R}^w| \leq 0.87c \cdot \frac{1}{B_c \tau_c}. \tag{2.57}$$

The platform-target range rate $|\dot{R}| \approx v_p^2 t_2/x_0$ (see (1.6)) is a maximum for a target entering the leading edge of the beam, or leaving the trailing edge of the beam, so that the maximum range rate encountered in practice is given closely by $|\dot{R}_{\max}| \approx 2\pi c v_p/(\omega_0 D)$. This maximum range rate sets the following limit

on the along-track velocity:

$$v_p^n < 0.15 \cdot B_c D \qquad \text{and} \qquad v_p^w < 0.87 \cdot \frac{\omega_0 D}{2\pi B_c \tau_c}. \tag{2.58}$$

For the Kiwi-SAS with $B_c = 20\,\text{kHz}$, $f_0 = \omega_0/(2\pi) = 30\,\text{kHz}$, $D = 0.325\,\text{m}$, and $\tau_c = 50\,\text{ms}$, the narrowband theory predicts that Doppler does not reduce the filter output substantially for speeds below $975\,\text{m/s}$, however, wideband theory sets the velocity limit at $8.5\,\text{m/s}$ (actually both tolerances are probably higher than this as the constants in (2.58) vary between references, eg. see [70]). As the along-track velocity is determined by the along-track ambiguity constraints of the synthetic aperture system, it is more appropriate to rearrange the wideband velocity constraint in (2.58) to be a constraint on the chirp length; that is,

$$\tau_c < 0.87 \cdot \frac{\omega_0 D}{2\pi B_c v_p} = 0.87 \cdot \frac{2\omega_0 \tau_{\text{rep}}}{\pi B_c}, \tag{2.59}$$

where the along-track velocity is constrained to $v_p \leq D/(4\tau_{\text{rep}})$. Using the Kiwi-SAS parameters above and the repetition period $\tau_{\text{rep}} = 0.3\text{s}$, the maximum chirp length predicted by (2.59) is 1.6s. Given that this value is less than the repetition period, the Kiwi-SAS system has no problems with temporal Doppler even when using CTFM ($\tau_c = \tau_{\text{rep}}$). Thus the Kiwi-SAS LFM pulse can be considered Doppler *tolerant.*

Even when a temporal Doppler analysis indicates that a particular waveform is tolerant, the output of the matched filter for LFM is a maximum *away from the actual target location.* The peaks of the ambiguity surfaces for both narrowband and wideband formulations follow the same path through range rate-delay ($\dot{R}$-τ) space given by $\dot{R} = \pi B_c c/(\omega_0 \tau_c) \cdot \tau$ [90]. By inserting the maximum range rate and the velocity ambiguity constraint, the maximum range measurement error is

$$\Delta x_{\text{max}} \approx \frac{c\tau}{2} = \frac{1}{2} \cdot \frac{c}{2B_c} \cdot \frac{\tau_c}{\tau_{\text{rep}}} = \frac{\delta x_{\text{3dB}}}{2} \cdot \frac{\tau_c}{\tau_{\text{rep}}}. \tag{2.60}$$

The maximum peak shift is only half a range resolution cell at the beam edge when using CTFM. The Kiwi-SAS images presented in this thesis typically used a 50ms chirp, so the shift in the correlation peak can be ignored.

A wideband analysis of the acceleration tolerance of LFM waveforms gives $|\ddot{R}^w| < 18c/(\pi B_c \tau_c^2)$ [90]. The maximum platform-target change of range rate is approximately $|\ddot{R}| \approx v_p^2/x_{\text{min}}$, where x_{min} is the minimum target range. Using the velocity ambiguity constraint, the limit on the chirp length due to acceleration is

$$\tau_c < \frac{12\tau_{\text{rep}}}{D} \cdot \sqrt{\frac{2cx_{\text{min}}}{\pi B_c}}. \tag{2.61}$$

The maximum chirp length for the Kiwi-SAS is 3.4s for a minimum range of 2m.

The results of the velocity and acceleration tolerances of this section, indicate that the LFM pulse employed by the Kiwi-SAS can be considered Doppler tolerant.

2.7.3.2   Ambiguity analysis of Hyperbolic FM

If a hyperbolic FM waveform undergoes a temporal Doppler distortion the resulting signal still overlays the original pulse on the f-t diagram. So the only correlation loss is due to the slight overlap loss at the pulse extremities. For HFM this property holds for all reasonable values of range rate so is considered velocity invariant [70, 91] and hence Doppler invariant. An analysis of the acceleration tolerance (Eq. (42) in [70]) can be used to give a limit on the HFM pulse length:

$$\tau_h < \frac{2\pi c x_{\min}}{v_p^2} \left(\frac{1}{\omega_a} + \frac{1}{\omega_b} \right). \tag{2.62}$$

Using $v_p = D/(4\tau_{\mathrm{rep}})$, $f_a = \omega_a/(2\pi) = 20\mathrm{kHz}$, $f_b = \omega_b/(2\pi) = 40\mathrm{kHz}$, $x_{\min} = 2\mathrm{m}$ and the other Kiwi-SAS parameters, the HFM pulse must be less than 3s. Thus, the HFM waveform is tolerant to the accelerations expected during Kiwi-SAS operation.

2.8   TEMPORAL DOPPLER TOLERANCE—VALIDITY OF THE 'STOP-START' ASSUMPTION

When developing synthetic aperture system models, it is common to assume that the 'stop-start' or 'stop-and-hop' approximation holds. In the 'stop-start' approximation, it is assumed that the platform carrying the measurement system can be considered to transmit each pulse and receive its returns (echoes) at *exactly* the same along-track sample location before *instantaneously* moving to the next sampling location. This assumption is equivalent to ignoring the *temporal* Doppler effects *within* a pulse, while exploiting the *spatial* Doppler shifts *between* pulses to develop the synthetic aperture model.

The applicability of the 'stop-start' assumption relies on the velocity and acceleration tolerances of the transmitted waveform. In SAR systems these tolerance requirements are always met, however, some SAS systems that transmit pulses with wide bandwidths and/or long pulses may only just meet these tolerance requirements. The analyses contained in the previous sections show that the waveforms employed by the Kiwi-SAS system can be considered Doppler tolerant under all operational conditions.

2.9   INTERPOLATION OF COMPLEX DATA

Interpolation of complex data is an important step in many of the synthetic aperture system models. Interpolation of the non-uniformly spaced samples collected by the synthetic aperture system onto a rectangular grid is necessary to utilize the FFT for efficient processing. This section presents three of the most commonly used *one-dimensional* interpolation techniques. Of the interpolations required in the synthetic aperture models; the range-Doppler and wavenumber algorithm require a one-dimensional interpolator, and the tomographic formulation of spotlight mode requires a two-dimensional interpolator. This two-dimensional interpolation is efficiently achieved by implementing two one-dimensional interpolations. Two-dimensional interpolators are discussed in depth by Soumekh, 1994 [146].

2.9.1   Sinc interpolation

Texts that present sampling theory refer to the sinc function as the perfect interpolator of a uniformly sampled signal. This is because the sinc interpolator has an ideal low-pass filter characteristic in the reciprocal Fourier domain. What these texts often neglect to mention, or comment on, is the fact that the sinc function is defined over an ordinate of infinite extent; $t \in [-\infty, \infty]$ for time domain interpolation, or $\omega \in [-\infty, \infty]$ for frequency interpolation. One reason this fact is not mentioned is that one of the most common applications requiring interpolation is the increasing or decreasing of a signals' *sampling rate* by some rational ratio $(p/q)\,\omega_s$, the actual *location* of the new samples is irrelevant. The procedure for *increasing* a signals sampling rate by $(p/q)\,\omega_s$ $(p > q)$ is as follows; the signal is Fourier transformed, padded with zeros to $p\,\omega_s$, inverse transformed, then decimated, keeping only every q^{th} sample (this resampling method is part of a more general field referred to as *multirate sampling* [29,51]). For synthetic aperture systems, it is the sample *location* that is important and these sample locations are *not always uniform*. To specify input and output sample locations explicitly requires the use of digital filters applied as time domain convolutions, i.e., requires a one-dimensional finite impulse response (FIR) digital filter.

The interpolation of time domain information is identical to frequency domain interpolation, so this section discusses interpolation from the time domain. To digitally implement a sinc interpolator, it is necessary to *truncate* the sinc function. The mathematical operation representing this truncation operation is a multiplication of the sinc function with a rect function. This multiplication in the time domain represents a convolution with the ideal low-pass filter characteristic of the sinc (a rect function in the frequency domain) with a sinc function representing the Fourier transform of the truncation function. This convolution has three detrimental effects in the Fourier domain; it produces ripples in what should ideally be a uniform pass-band, produces sidelobes in the stop-band, and it increases the bandwidth of the Fourier domain (this may then cause aliasing of the Fourier signal). Two operations are necessary to mitigate these effects; bandwidth reduction and time domain weighting. Reduction

of the truncated interpolating filter bandwidth is achieved by broadening the truncated and weighted interpolating sinc function by 4 to 5 percent. This new sinc function is then weighted in the time-domain to reduce the pass-band and stop-band ripple in the frequency domain. This thesis weights the truncated sinc with a Hanning window. The weighted, truncated, and reduced bandwidth interpolating filter then has a uniform pass-band over 95 percent of the sampled bandwidth (the outer 5 percent of the final image may need to be discarded).

The implications of the application of this interpolator in each of the appropriate synthetic aperture algorithms are as follows. The range-Doppler algorithm contains a time domain mapping operation that requires *time domain interpolation*. Because the frequency domain of this time domain signal has zero magnitude outside of the bandwidth B_c it is not necessary to discard 5 percent of the interpolated bandwidth (that 5 percent had zero magnitude anyway). Both the wavenumber and the tomographic mode require *frequency domain interpolation*. After the interpolation has been completed and the data inverse Fourier transformed back to the time or image domain, the outer 5 percent of the image pixels (in the interpolated direction) should be removed.

Images produced throughout this thesis for algorithms requiring an interpolator used a Hanning weighted sinc function truncated at 8 one-sided zero crossings. Further details of this interpolator can be found in Chapter 3 of Jakowatz *et al*, 1996 [82]. Jakowatz *et al* [82] also contains an application of this form of filter to a spotlight SAR system that uses the tomographic formulation of the spotlight model to process data.

2.9.2 Polyphase filters

The resampling requirements of an interpolator depend on a wide variety of parameters such as; the image formation algorithm, the collection geometry, and the synthetic aperture system parameters. As it is not practical to design an ensemble of resampling filters to handle each case, *polyphase filtering* is employed as an efficient method of generating the required filter from a *prototype filter* template. The prototype filter is a sampled version of the interpolator, such as the sinc interpolator described above. The filter weightings required for a particular application are obtained by resampling the prototype filter using the multirate sampling technique described above. If the exact filter weight for a particular location can not be obtained, a nearest neighbour approach is used. The term 'polyphase' refers to the efficient manner in which these filter weights are then applied to the raw data to produce the interpolated output.

Polyphase filters are used in Carrara *et al*, 1995 [21] to perform the interpolation steps required in spotlight systems. Application of these filters to both the tomographic formulation and the wavenumber algorithm are presented. For further details of polyphase filters see Chapter 4 of Carrara *et al* [21].

2.9.3   Chirp scaling

The principle of chirp scaling is exploited in the chirp scaling algorithm to avoid the interpolation step required in the range-Doppler algorithm. Alternatively, the chirp scaling operation can be interpreted as an approximation to the Stolt map of the wavenumber algorithm. The chirp scaling principle applies to large time-bandwidth LFM signals [123, pp203-204]. The baseband return of a LFM pulse of chirp rate K_c from a target at range $r_a = ct_a/2$ is

$$e_b(t) = \text{rect}\left(\frac{t - t_a}{\tau_c}\right) \cdot \exp\left[j\pi K_c(t - t_a)^2\right]. \tag{2.63}$$

In the chirp scaling algorithm, the modulating phase of the LFM chirps in range are perturbed such that the *phase centers* of all targets in the scene have congruent range migration loci. The phase structure of the chirp signal in (2.63) is reshaped by a multiplication with another signal with a chirp rate that is a small fraction of the original FM rate. Multiplication and re-arrangement gives

$$\begin{aligned} m_b(t) &= \text{rect}\left(\frac{t - t_a}{\tau_c}\right) \cdot \exp\left[j\pi K_c(t - t_a)^2\right] \cdot \exp(j\pi C_s K_c t^2) \\ &= \text{rect}\left(\frac{t - t_a}{\tau_c}\right) \cdot \exp\left[j\pi K_s(t - t_b)^2\right] \cdot \exp\left(j\pi K_c \frac{C_s}{(1 + C_s)} t_a^2\right), \end{aligned} \tag{2.64}$$

where dimensionless C_s is the *chirp scaling multiplier*, the *scaled chirp rate* is

$$K_s = K_c(1 + C_s) \tag{2.65}$$

and the new phase center is located at

$$t_b = \frac{t_a}{(1 + C_s)}. \tag{2.66}$$

The second exponential term on the final line of (2.64) is called the *phase residual*. During the final step of the chirp scaling algorithm, this phase term is eliminated via multiplication by the complex conjugate of the phase residual.

To determine the effect of the chirp scaling on the range response requires the spectra of the signals. The baseband spectrum of (2.63) is

$$E_b(\omega_b) \approx \text{rect}\left(\frac{\omega_b}{2\pi B_c}\right) \cdot \sqrt{\frac{j}{K_c}} \cdot \exp\left(-j\frac{\omega_b^2}{4\pi K_c}\right) \cdot \exp(-j\omega_b t_a) \tag{2.67}$$

and the baseband spectrum of the chirp scaled signal in (2.64) is

$$M_b(\omega_b) \approx \mathrm{rect}\left(\frac{\omega_b - 2\pi C_s K_c t_a}{2\pi B_c(1 + C_s)}\right) \cdot \sqrt{\frac{j}{K_s}} \cdot \exp\left(-j\frac{\omega_b^2}{4\pi K_s}\right) \cdot \exp(-j\omega_b t_b) \cdot \exp\left(j\pi K_c \frac{C_s}{(1 + C_s)} t_a^2\right),$$

(2.68)

where it can be seen that the effect of chirp scaling is to produce a spectrum of bandwidth $B_c(1 + C_s)$ with a modified magnitude about a new carrier $\omega_0 + 2\pi C_s K_c t_a$. The system sampling frequency ω_s must be such that it supports this spectrum shift. Pulse compression via phase matching of the quadratic term in (2.67) and inverse Fourier transformation produces the following time domain response:

$$s_b(t) = \sqrt{j\tau_c B_c} \cdot \mathrm{sinc}[B_c(t - t_a)].$$

(2.69)

Similarly, compression and phase residual matching of (2.68) gives

$$n_b(t) = \sqrt{j\tau_c B_c(1 + C_s)} \cdot \mathrm{sinc}[B_c(1 + C_s)(t - t_b)] \cdot \exp[j2\pi C_s K_c t_a(t - t_b)].$$

(2.70)

where it can be seen that the effect of chirp scaling is to modify the width and magnitude of the range response and to introduce a modulation term. The shifts required in the chirp scaling algorithm are kept to a minimum by choosing the center of the swath as the reference range. The shifts typically experienced require small values for C_s and so to a good approximation

$$|n_b(t)| \approx \sqrt{\tau_c B_c} \cdot |\mathrm{sinc}[B_c(t - t_b)]|.$$

(2.71)

Thus, the chirp scaling multiplier produces a shift of the range response function. In the range-Doppler algorithm this shift is achieved using a much slower time-domain interpolator. The modulation term in (2.70) distorts the 2-D wavenumber domain, producing the same curved spectrum shape as produced by the Stolt mapping operation of the wavenumber algorithm.

2.10   SUMMARY

The mapping operators, range imaging techniques, and interpolators presented in this chapter along with the wide bandwidth radiation patterns developed in the next chapter, form the basis of all the synthetic aperture imaging techniques in Chapter 4. The use of the mapping operator to emphasize the coordinate transforms, and hence digital interpolations, required by the synthetic aperture models is a key element in the clear presentation and direct digital implementation of the synthetic aperture models and inversion schemes.

Chapter 3

PRINCIPLES OF ARRAY THEORY

The synthetic aperture principle can be presented from two equivalent points of view; the spatial domain formulation (array theory) or in spatial Fourier transform space (the spatial Doppler formulation). This chapter begins by developing a wide bandwidth formulation of aperture theory using Fourier array theory. This wide bandwidth formulation is then related to classical aperture theory. Real and synthetic array theory is presented and *beam forming* techniques such as array steering, focusing and de-focusing are covered. Some of the classical aspects of beam forming are interpreted in terms of the wide bandwidth theory and some of the misconceptions of synthetic aperture theory are resolved. In particular, the correct interpretation of the effects of null suppression is given.

The effect of the aperture radiation patterns in synthetic aperture systems are often incorrectly modelled. Especially in wide bandwidth, low-Q, systems where the radiation patterns have often been considered too difficult to model. This chapter shows how the spatial-temporal response of wide bandwidth active imaging systems can easily be incorporated into imaging models and inversion schemes.

A transmitter (or receiver) that generates (or receives) with an equal response, independent of angle is known as an *omni-directional* element. An omni-directional element is (mathematically) considered to have an infinitesimally small physical extent. To construct an aperture (or antenna) with a response over a given solid angle in a given direction a number of these omni-directional elements can be spaced over a two-dimensional plane to produce a *real array*. If all of the elements in this real array are powered up (energized) simultaneously then the omni-directional response from each element interferes constructively and destructively in the medium in front of the array. Given a *simultaneous* excitation of all elements, the element patterns interfere *constructively* only over angles centered about the normal to the array face (the array *boresight*) and produce a main beam or main lobe over a solid angle that depends on the dimensions of the planar array and the frequency of the energizing signal. If the elements in the array are energized at different times, then it is possible to *steer* and *focus* the main lobe of the array to any given direction, this type of array is known as a *phased array*. The interference patterns of the energized elements with respect to angle from the boresight of the array are known as *beam patterns*,

directivity patterns or *radiation patterns* and reflect the angular response of an array.

The term *aperture* generally refers to a mathematically continuous quantity, whereas the term *array* refers to the physical implementation of an aperture by a sampling operation where elements are located at each sample location. However, the terms aperture and array can be used interchangeably.

In the following sections, the development of array theory and the radiation patterns of apertures is based on approximations of the equations governing multidimensional scattering theory. Many of these approximations are valid for both sonar and radar. The order of presentation of this approximate formulation follows more of an intuitive path than an "exact" path. The result is, however, valid for the situations dealt with in the context of this thesis. For a more accurate analysis of multidimensional scattering theory, see the excellent review by Bates *et al*, 1991 [8], the book by Soumekh, 1994 [146], or the book by Goodman, 1968 [58]. The implicit assumptions made in this development are that the transmitted wavelength and the aperture length are much smaller than the range to the nearest target or spatial location under consideration.

3.1   THE SPATIAL-TEMPORAL RESPONSE AND RADIATION PATTERN OF AN APERTURE

The two-dimensional apertures encountered in synthetic aperture systems can be adequately treated by separating the radiation patterns in along-track and height into two separable one-dimensional patterns [12, Ch. 13]. An aperture is considered to be a reversible system, so that the development of either a transmitting or receiving aperture is adequate to describe both. With reference to Fig. 3.1, the signal received at a point (x, y) from a *transmit* aperture of length D_T, located at the origin of the (x, y) domain, that has been energized via an illumination function $i_T(v)$ and radiating a signal $p_m(t)$ is given by (Huygens' principle [58, p31])

$$h_T(t, x, y) \approx \int_{-D_T/2}^{D_T/2} \frac{i_T(v)}{\sqrt{x^2 + (y - v)^2}} \cdot p_m\left(t - \frac{\sqrt{x^2 + (y - v)^2}}{c}\right) dv, \qquad (3.1)$$

where v is the axis of the face of the aperture, parallel to the scene y-axis. Temporal responses and radiation patterns are denoted using the single functional notation, h or H, as the functions required by this thesis are related via a temporal Fourier transform alone.

A temporal Fourier transform of (3.1) gives the *radiation pattern* of the aperture:

$$H_T(\omega, x, y) \approx P_m(\omega) \cdot \int_{-D_T/2}^{D_T/2} i_T(v) \cdot \frac{\exp\left(-jk\sqrt{x^2 + (y - v)^2}\right)}{\sqrt{x^2 + (y - v)^2}} dv. \qquad (3.2)$$

We can reference spatial and temporal quantities to the center of the aperture, thus removing the v
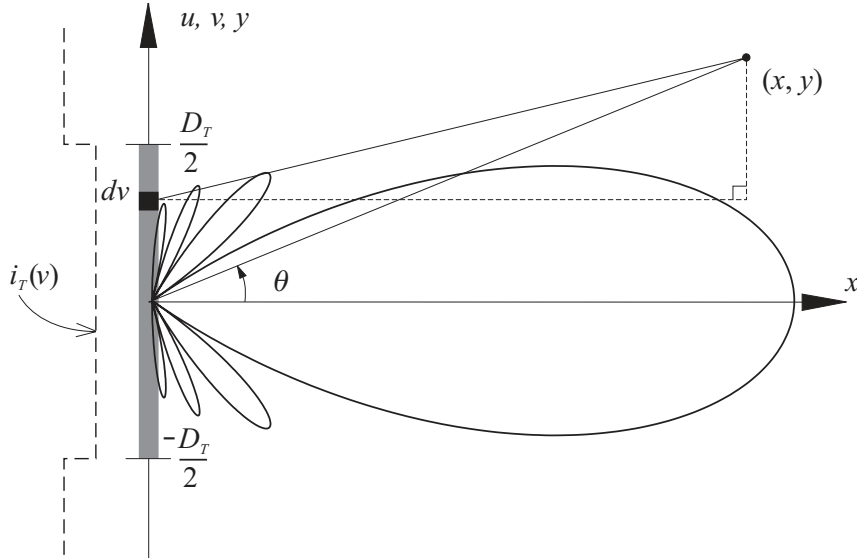
Figure 3.1   Geometry used for the derivation of an aperture's radiation pattern. The sinc-like function plotted with respect to θ is the radiation pattern of a uniformly illuminated aperture for a single frequency within the signal transmitted by the aperture.

dependence, by replacing the exponential in (3.2) with its Fourier decomposition and by changing the order of integration [146, p152 and pp192-195]:

$$
\begin{aligned}
H_T(\omega, x, y) &\approx P_m(\omega) \cdot \int_{-D_T/2}^{D_T/2} i_T(v) \int_{-k}^{k} \frac{\exp\left(-j\sqrt{k^2 - k_v^2} \cdot x - jk_v(y - v)\right)}{\sqrt{k^2 - k_v^2}} dk_v dv \\
&= P_m(\omega) \cdot \int_{-k}^{k} \left\{ \int_{-D_T/2}^{D_T/2} i_T(v) \exp(jk_v v) dv \right\} \frac{\exp\left(-j\sqrt{k^2 - k_v^2} \cdot x - jk_v y\right)}{\sqrt{k^2 - k_v^2}} dk_v \\
&= P_m(\omega) \cdot \int_{-k}^{k} I_T(k_v) \frac{\exp\left(-j\sqrt{k^2 - k_v^2} \cdot x - jk_v y\right)}{\sqrt{k^2 - k_v^2}} dk_v,
\end{aligned}
\tag{3.3}
$$

where $I_T(k_v)$ is a Fourier-like transform of $i_T(v)$ with the normal forward Fourier kernel of $\exp(-jk_v v)$ replaced by $\exp(+jk_v v)$. In Fourier array imaging, the Fourier parameter of the aperture, k_v, is known as the Doppler wavenumber, in units of radians per meter, of the array. In general $I_T(k_v)$ is complex with a slowly varying amplitude $A_T(k_v)$ and a phase that determines whether the real aperture has any steering or focusing power. So,

$$
I_T(k_v) = A_T(k_v) \exp[-j\psi(k_v)].
\tag{3.4}
$$

For an unsteered and unfocused real aperture, the illumination function $i_T(v)$ is real and mostly just

tapers or apodises the real aperture. (This means that the *effective* length of the aperture D_{eff} is smaller than its physical length D_T). Thus,

$$A_T(k_v) = \int_v i_T(v) \exp(+jk_v v) dv, \tag{3.5}$$

where the $+j$ Fourier kernel is usually associated with an inverse Fourier transform (the use of an inverse Fourier transform as defined in (2.10) would, however, also scale this function by $1/(2\pi)$). In most cases, $i_T(v)$ is symmetric and a forward Fourier kernel suffices. The integral in (3.3) can be solved via the principle of stationary phase (see Appendix A) to give

$$\begin{aligned} H_T(\omega, x, y) &\approx P_m(\omega) \cdot \int_{-k}^{k} A_T(k_v) \frac{\exp\left(-j\sqrt{k^2 - k_v^2} \cdot x - jk_v y\right)}{\sqrt{k^2 - k_v^2}} dk_v \\ &\approx P_m(\omega) \cdot A_T(k\sin\theta) \cdot \frac{\exp\left(-jk\sqrt{x^2 + y^2}\right)}{\sqrt{x^2 + y^2}} \\ &= P_m(\omega) \cdot A_t(\omega, x, y) \cdot \frac{\exp\left(-jk\sqrt{x^2 + y^2}\right)}{\sqrt{x^2 + y^2}}, \end{aligned} \tag{3.6}$$

where the amplitude function or *amplitude pattern* scales from the k_v-wavenumber domain to the spatial (x, y)-domain via

$$A_t(\omega, x, y) \equiv A_T(k_v) \tag{3.7}$$

where the wavenumber $k_v = k\sin\theta$ and $\theta = \sin^{-1}\left(y/\sqrt{x^2 + y^2}\right)$ is the aspect angle from the center of the aperture to the measurement location. The magnitude of the aperture's amplitude pattern $|A_t(\omega, x, y)|$ dictates its power distribution in the spatial domain at frequency ω [147]. The aperture *beam pattern* is this magnitude pattern plotted versus angle θ, spatial frequency $\sin\theta/\lambda$, or wavenumber $k_v = k\sin\theta$. Note that the phase delay (exponential function) in (3.6) is now referenced from the array center.

The *spatial-temporal response* of the aperture for the transmitted signal $p_m(t)$ is then

$$h_T(t, x, y) \approx \frac{1}{\sqrt{x^2 + y^2}} \cdot \left\{ a_t(t, x, y) \odot_t p_m(t) \odot_t \delta\left(t - \frac{\sqrt{x^2 + y^2}}{c}\right) \right\}, \tag{3.8}$$

where the *impulse response* of the aperture

$$a_t(t, x, y) = \mathcal{F}_\omega^{-1}\left\{A_t(\omega, x, y)\right\}, \tag{3.9}$$

reflects the aperture response to an impulsive plane wave incident on the aperture at an angle θ to the main response axis.

An example:

Figure 3.1 shows a transmit aperture of length D_T with a constant illumination function $i_T(v) = \text{rect}(t/D_T)$. This aperture has a radiation pattern with an amplitude pattern given by

$$A_T(k_v) = \text{sinc}\left(\frac{k_v D_T}{2\pi}\right),$$

(3.10)

or equivalently,

$$A_t(\omega, x, y) = \text{sinc}\left(\frac{k D_T \sin\theta}{2\pi}\right).$$

(3.11)

This amplitude pattern, for a single frequency ω, is plotted versus aspect angle θ in Fig. 3.1. The amplitude pattern has a wavenumber bandwidth $B_{k_v} = 2\pi/D_T$ and a 3dB main lobe width of

$$\theta_{3\text{dB}} = \alpha_w \frac{2\pi}{k D_T} = \alpha_w \frac{\lambda}{D_T},$$

(3.12)

where $\alpha_w = 0.88$ for the uniformly illuminated aperture.

3.1.1   Correlated beam patterns

To determine the angular response of a wide bandwidth system in the time domain, it is necessary to determine the *correlated beam pattern*. The correlated beam pattern is the inverse Fourier transform of the pulse compressed spectrum weighted by the radiation amplitude pattern at time $t = 0$; that is,

$$a_{\text{corr}}(0, x, y) = \frac{1}{2\pi} \int_\omega |P_m(\omega)|^2 \cdot A_t(\omega, x, y) d\omega.$$

(3.13)

The correlated beam pattern can also be interpreted as the weighted sum of all the monochromatic radiation patterns produced by the aperture. For example, a uniformly illuminated aperture has the amplitude pattern $A_t(\omega, x, y) = \text{sinc}[k D_T \sin\theta/(2\pi)]$. If the transmitted signal $p_m(t)$ is say, a LFM chirp, then $|P_m(\omega)|^2 \approx \text{rect}[(\omega - \omega_0)/(2\pi B_c)]$ and the overall angular response in the time domain is given as the uniformly weighted sum of the scaled (changing) sinc patterns.

An interesting comparison of this correlated beam pattern can be made to higher order windowing or weighting functions. Many higher order window functions are made up of multiplied or *shifted-scaled(in height)*-and-summed sinc functions (eg. Triangle, Hanning, Hamming, Blackman and Blackman-Harris windows [71]), whereas the correlated beam pattern of the uniform illumination aperture is made up of *scaled(in angle)*-and-summed sinc functions. The zeros of the monochromatic sinc patterns act

to suppress the sidelobes in the spatial-temporal response to an extent that is proportional to the signal bandwidth. This effect has been suggested as a method for smearing the grating lobes in an undersampled synthetic aperture and is discussed in Chapter 5.

3.2   CLASSICAL APERTURE THEORY

Classical aperture theory also begins with the radiation pattern shown in (3.2). Instead of performing a Fourier decomposition of the exponential, the following approximations are made; the square root term in the exponential (phase) of (3.2) is replaced by a truncated Taylor series expansion [150, p9]:

$$\sqrt{x^2 + (y-v)^2} \approx r - v\sin\theta + \frac{v^2}{2r}, \tag{3.14}$$

where only up to the quadratic term has been retained, and the denominator is replaced by

$$\sqrt{x^2 + (y-v)^2} \approx r, \tag{3.15}$$

where the polar coordinates of the point (x, y) are given as $(r, \theta) = \left[\sqrt{x^2 + y^2}, \tan^{-1}(y/x)\right]$. The *classical radiation pattern* is then

$$
\begin{aligned}
H_T^c(\omega, x, y) &\approx P_m(\omega) \cdot \left\{ \int_{-\infty}^{\infty} i_T(v) \exp\left[-jk\left(v\sin\theta - \frac{v^2}{2r}\right)\right] dv \right\} \cdot \frac{\exp\left(-jk\sqrt{x^2+y^2}\right)}{\sqrt{x^2+y^2}} \\
&= P_m(\omega) \cdot A_t^c(\omega, x, y) \cdot \frac{\exp\left(-jk\sqrt{x^2+y^2}\right)}{\sqrt{x^2+y^2}},
\end{aligned}
\tag{3.16}
$$

where the *classical amplitude pattern*, $A_t^c(\omega, x, y)$, is the scaled version of

$$
\begin{aligned}
A_T^c(k_v) &= \int_{-\infty}^{\infty} \left\{ i_T(v) \cdot \exp\left(-j\frac{kv^2}{2r}\right) \right\} \exp(jk_v v) dv \\
&= I_T(k_v) \odot_{k_v} \exp\left(j\frac{k_v^2 r}{2k}\right).
\end{aligned}
\tag{3.17}
$$

A comparison of the amplitude pattern predicted by a classical aperture analysis in (3.17) with the amplitude pattern predicted by Fourier array theory in (3.5) shows that there is a difference involving a quadratic term (actually, there are higher order terms that were ignored in the initial Taylor series expansion in (3.14)). The classical form of the amplitude pattern can be split into two spatial regions in range; the near field or *Fresnel zone* of the aperture where the quadratic and higher order terms are significant, and the far field or *Fraunhofer zone* of the aperture where the quadratic and higher order terms are no longer significant and the Fourier kernel dominates [150, p10] [25]. The transition between

these two zones is given by

$$r_{\text{far field}} \geq \frac{2D_T^2}{\lambda} = \frac{kD_T^2}{\pi}. \tag{3.18}$$

Fourier array theory does not divide the range spatial domain into different different regions. Fourier array theory always treats the wavefronts as spherical, so there is no need to differentiate between regions. Classical theory approximates the spherical wavefronts as 'almost' quadratic in the near-field, and as plane waves in the far-field. It may seem then, that classical aperture theory is redundant, however, classical and Fourier theory are still both necessary. Fourier array theory is based on the spatial frequency (wavenumber) decomposition of the spatial energy incident or transmitted across an array. When dealing with phased arrays or synthetic apertures, this spatial information is available due to the multiplicity of elements that make up the array. In these two cases, so long as the spatial locations imaged are in the far-field of the array elements, Fourier array theory provides a more accurate inversion of the wave field than classical array theory. When dealing with a real aperture made up of elements that are connected in parallel, the resulting array responds like a continuous aperture. By allowing the spatial information to coherently interfere before recording the signal from this continuous aperture, some of the spatial information is destroyed and the Fresnel and Fraunhofer regions exist. A mathematical analysis of a hypothetical continuous aperture is always possible, it is the *physical* implementation of a continuous aperture that creates the two zones.

In summary, classical aperture theory is applicable to systems employing continuous apertures that have no access to the spatial information across the aperture. Fourier array theory is applicable to phased arrays and synthetic arrays where the spatial information is available. Systems that apply classical (near field) methods to phased and synthetic arrays are performing an unnecessary approximation.

3.2.1   Illumination function modulation–Radiation pattern focusing, steering, broadening, and invariance

To produce a broadside radiation pattern, the illumination function, i.e., the voltage or current weighting across the aperture, is *real*, and all elements of the array are energized *simultaneously*. If time delays are inserted across the aperture then we obtain the ability to control the shape of the radiation pattern. These delays translate to a complex illumination function:

$$i_c(v) = i_T(v) \exp\left[j\phi(v)\right]. \tag{3.19}$$

In the simplest application $\phi(v)$ can be used to image in the near-field of the aperture by *focusing* the radiation pattern. Focusing is achieved by the insertion of a quadratic (hyperbolic if the exact expression is used) delay of the transmitted signal to the elements across the array. If a linear delay

is also inserted, then the beam pattern can be *steered* or *scanned*. For example, if $\phi(v) = k(\alpha v + \beta v^2)$ where β is the focusing parameter and α is the steering parameter, the near-field amplitude pattern predicted by classical aperture theory is

$$
\begin{aligned}
A^c_{\alpha\beta}(k_v; r) &= \int_{-\infty}^{\infty} i_c(v) \exp\left(jk_v v - \frac{kv^2}{2r}\right) dv \\
&= \int_{-\infty}^{\infty} i_T(v) \exp\left[j(k_v + k\alpha)v + jk\left(\beta - \frac{1}{2r}\right)v^2\right] dv \qquad (3.20) \\
&= A_T(k_v + k\alpha) \qquad \text{for} \quad \beta = \frac{1}{2r}.
\end{aligned}
$$

At the focal point r, the amplitude pattern is the same as the far-field pattern centered on spatial frequency (wavenumber) $k_{v0} = -k\alpha$. If the aperture is steered to large angles off boresight, higher order phase terms cause the amplitude pattern to broaden [150, p34-37].

The process of illumination function modulation is more general than it may first appear. The function $i_T(v)$ is always a slowly varying amplitude function, so the spatial bandwidth of the aperture can be *increased* by the modulation function. This is *exactly* the same principle as is used in pulse compression. Looking at the effect of the focusing and steering phase on the near-field pattern in (3.20), we see that the near-field pattern has a linear spatial frequency modulation (LSFM) (i.e. a quadratic spatial modulation) imposed on it by the approximations used in developing the classical formulation of the amplitude pattern. The LSFM is *matched* by the focusing phase controlled by β, then the pattern is then shifted (modulated) by the steering parameter α.

Instead of performing this beam compression operation in the near-field we can broaden (*defocus*) the far-field pattern. In this way, we can use a long real array of length L_{ra} and quadratically modulate its illumination function to produce the spatial response expected from a *much smaller array* (the phase functions generated in the spatial frequency domain can easily be incorporated into the inversion). For example, if an aperture of length L_{ra} has an illumination function given by $i_c(v) = \mathrm{rect}(v/L_{\mathrm{ra}})\exp(j\pi k_D v^2)$, where the 'chirp rate' $k_D = 1/(L_{\mathrm{ra}}D)$ the aperture produces a wavenumber bandwidth of $B_{k_v} \approx 2\pi k_D L_{\mathrm{ra}} = 2\pi/D$, which is the same spatial frequency response as an aperture of length D (say the response of a single element in the array!). This property is useful in high-resolution synthetic aperture systems where short apertures are needed for high along-track resolution. Short apertures can only transmit low power; however, a larger defocused array can transmit more power into the same solid angles as a much smaller array. This concept has been applied to medical synthetic aperture systems, see reference [86].

The defocused beam just described had the same spatial frequency response as an array of length D, but its angular response still changes with frequency in a wide bandwidth application as $\theta_{3\mathrm{dB}} \approx \lambda/D$. It is possible to create *invariant* beam patterns by employing a frequency dependent 'chirp rate'. To

make the above example have a constant angular resolution of θ_c, the approximate 'chirp rate' required is $k_{\theta_c} \approx k \sin \theta_c / (2\pi L_{\mathrm{ra}})$. Previous efforts to generate angular invariant beam patterns have focused on keeping the length of the array constant with wavelength [61, 163]. For example, if an upswept LFM pulse is applied to an array initially of length L_{ra} then outer elements of the array are turned off as the LFM instantaneous frequency increases. Reference [157] suggests a method for defocusing using a random modulating phase function.

3.2.2 Depth of focus (radial beam width)

The depth of focus of a finite length aperture is the distance about the focal point r over which fixed focusing parameters cause a tolerable decrease in target response. This effect is exploited in synthetic aperture systems that treat the system impulse response as invariant over the depth of focus of the synthetic array (eg., see Section 4.3.1).

For a focused aperture, the radial beam width can be determined by considering points around the focal point r, shifted in range by some small amount. Analysis of the amplitude pattern produced with respect to range gives the *depth of focus* of the focussed array of length D_T as

$$\mathrm{DOF} \approx \alpha_{\mathrm{DOF}} \frac{r^2 \lambda}{D_T^2}, \tag{3.21}$$

where the constant $\alpha_{\mathrm{DOF}} = 2, 4, 7$ depending on the criteria you choose for acceptable loss of output amplitude from the array [146, p204] [74] [150, p53] (Note r is used here, not x, as the aperture may be steered off boresight). For a synthetic aperture, the real aperture length D_T is replaced with the synthetic aperture length $L_{\mathrm{sa}} = r\lambda/D$ and the depth of focus becomes

$$\mathrm{DOF}_{\mathrm{synthetic}} = \alpha_{\mathrm{DOF}} \frac{D_T^2}{\lambda}, \tag{3.22}$$

which is independent of range due to the fact that the synthetic aperture length scales with range. The four-look Seasat-SAR processor described in Elachi, 1988 [47, p108] needed to update the along-track focusing parameters across 3 depths of focus.

3.3 ARRAY THEORY

An array is an aperture excited only at points or in localized areas [150, p71]. The practical generation of a continuous illumination function $i_T(v)$ via the placement of omni-directional elements every Δ_v

along a real aperture of length L_{ra} is written

$$i_{\mathrm{af}}(v) = i_T(v) \cdot \sum_n \delta(v - n\Delta_v). \tag{3.23}$$

This array has the amplitude pattern or *array factor*

$$A_{\mathrm{af}}(k_v) = I_T(k_v) \odot_{k_v} \frac{2\pi}{\Delta_v} \cdot \sum_n \delta\left(k_v - n\frac{2\pi}{\Delta_v}\right). \tag{3.24}$$

If the illumination function is uniform over the array length, then the array factor is

$$
\begin{aligned}
A_{\mathrm{af}}(k_v) &= \mathrm{sinc}\left(\frac{k_v L_{\mathrm{ra}}}{2\pi}\right) \odot_{k_v} \frac{2\pi}{\Delta_v} \cdot \sum_n \delta\left(k_v - n\frac{2\pi}{\Delta_v}\right) \\
&= \frac{2\pi}{\Delta_v} \cdot \frac{\sin(k_v N \Delta_v/2)}{N \sin(k_v \Delta_v/2)},
\end{aligned}
\tag{3.25}
$$

where $N = L_{\mathrm{ra}}/\Delta_v$ is the number of elements in the array. Sampling of the aperture has caused the radiation pattern to contain repeated versions of the main lobe of the continuous aperture radiation pattern. These repeated lobes are known as *grating lobes*. The aperture has a wavenumber bandwidth of $B_{k_v} = 2k$ as $k_v \in [-k, k]$, thus sample spacings of

$$\Delta_v = \frac{2\pi}{B_{k_v}} = \frac{\pi}{k} = \frac{\lambda}{2} \tag{3.26}$$

are adequate to force the grating lobes out of the 'visible' spectrum. If the array is to be operated at broadside (i.e., no beam steering) then this sample spacing is often relaxed to slightly less than λ. As the sample spacings approach $\Delta_v = \lambda$ grating lobes move into the visible region [150, p99]. In the Kiwi-SAS system, the transmitter and receiver apertures are made up of arrays of closely packed elements. This allows the overall arrays of length D_T and D_R to be treated as continuous apertures.

An omni-directional model is appropriate for the microstrip or slotted waveguide elements employed to fabricate the electromagnetic arrays used in radar [31, p274]. In sonar systems, the elements or transducers used to fabricate the array have a physical extent and thus have their own radiation patterns. The element response, $i_e(v)$, is incorporated into the array model via a convolution with the sampled array illumination function in (3.23). The radiation pattern of this form of array is

$$A_{\mathrm{array}}(k_v) = A_{\mathrm{af}}(k_v) \cdot I_e(k_v). \tag{3.27}$$

If the element has a uniform response, then it has a radiation pattern with nulls every $k_v = m2\pi/D_e$, where D_e is the element length and $|m| = 1, 2, \cdots$. If the elements are closely packed in the array, then

$\Delta_v = D_e$ and the nulls of the element radiation pattern *suppress* the grating lobes in the array factor. Section 3.5 shows how this aspect of array theory has been incorrectly applied to synthetic aperture theory.

When determining the number of elements in a sonar array, the following must be considered; the sampled illumination function may need some form of amplitude taper to reduce the sidelobes in the radiation pattern, the closer the elements are to omni-directional, the closer the sampled illumination function approximates the continuous form. The number of elements in an array is also constrained by fabrication limitations. For example, the wide bandwidth design of the Kiwi-SAS sonar transmitters resulted in elements with a square face of 27mm each side, only after these transducers were built was the array illumination function considered.

3.4   ACTIVE ARRAYS

Figure 3.2 shows the four types of arrays commonly employed in active imaging systems. The following sections develop models for, and/or discuss the similarities and differences between, these real and synthetic arrays.

3.4.1   Focus-on-receive real array

Figure 3.2(a) shows an active imaging system with a single transmitter and multiple receivers. The multiple receivers of this array have separate channels with time delay and summing networks. Each time a pulse is transmitted, these networks act to steer and focus the receive array to the point under consideration. The transmitter-to-target phase path remains constant, whereas the receiver-to-target distance varies. The receiver array exploits the spatial Doppler effect of this *one-way* varying phase to obtain an ultimate angular resolution of

$$\theta_{3\mathrm{dB}} \approx \frac{\lambda}{L_{\mathrm{ra}} \cos\theta}, \tag{3.28}$$

where θ is the steer angle of the array [146, p202].

An unsteered (broadside) aperture acts as a low-pass spatial filter, while a steered aperture acts as a band-pass filter. The bandwidth of these spatial filters is inversely proportional to the aperture length. In the set-up shown in Fig. 3.2(a), the transmit aperture insonifies or illuminates an angular region $\theta \in [-\lambda/(2D_T), \lambda/(2D_T)]$. Targets within this region are imaged by steering and focusing the higher resolution (narrower band-pass filter) receive aperture and recording the echoes. Generally only the magnitude information is stored by these systems.
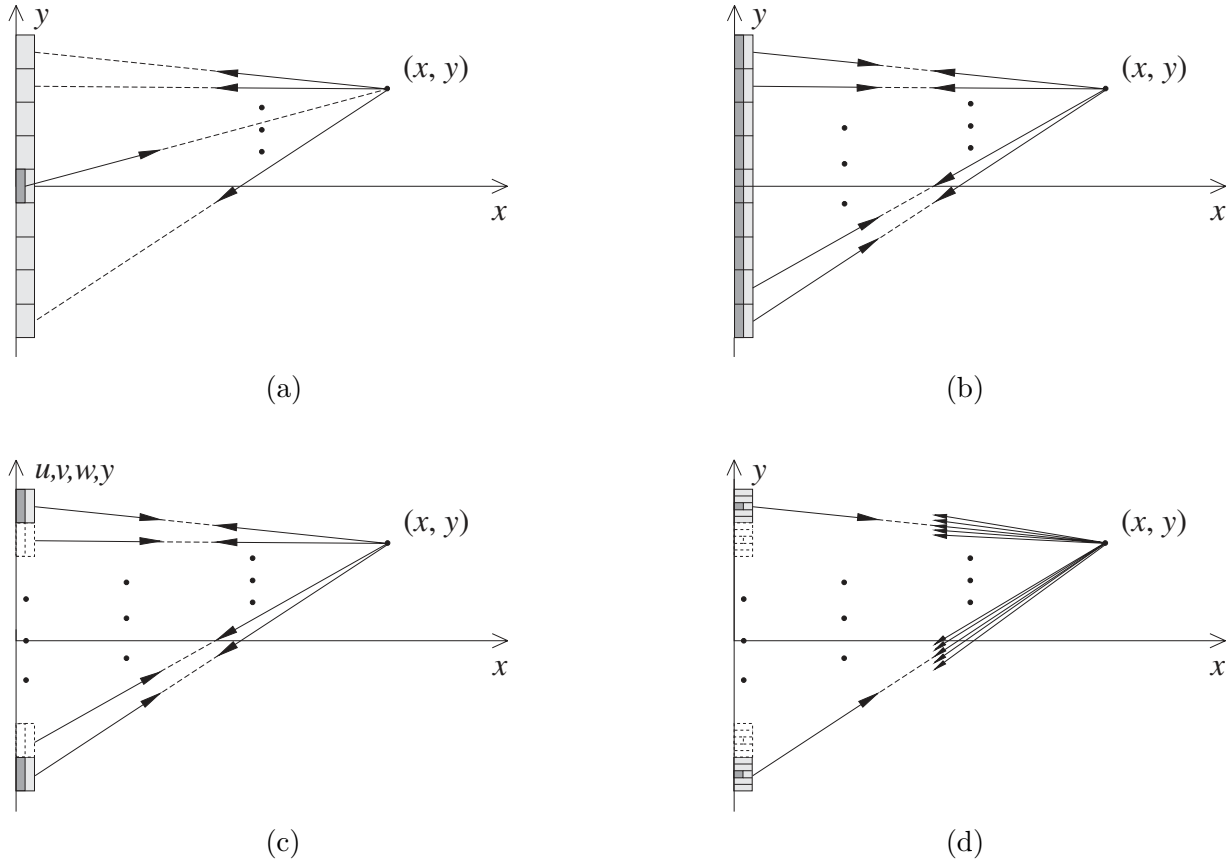
Figure 3.2   Active arrays. (a) phased array with focus-on-receive only, (b) phased array with focus-on-transmit and receive, (c) single receiver synthetic array, (d) multiple receiver synthetic array. Transmitter elements are indicated by the darker shaded boxes, while the receiver elements are indicated by the lighter shaded boxes. Note that the synthetic arrays shown are *schematic* only and the locations of the apertures does not reflect the actual along-track sample spacings required by practical systems.

3.4.2   Focus-on-transmit and receive real array

Figure 3.2(b) is the more general form of a *phased array*. With a phased array, focusing on transmission and reception is possible [150]. Focusing of a phased array has its origin in the approximation used to derive the classical radiation pattern of an aperture (see (3.17)). The more accurate theory of phased arrays developed in Chapter 3, of Soumekh, 1994 [146], shows that a better reconstruction of the target field is obtained via the inversion of the coherently stored signals than is obtained using the classical analog beam forming technique.

A classical beam forming imaging system coherently detects the echo signals [146, p199]. This approach limits the angular resolution of the phased array to (3.28). In the phased array inversion

scheme afforded by Fourier array imaging, focusing is not necessary, and the angular resolution of the array becomes dependent on the steer angles, target locations, and transmitted signal wavelength, i.e., reconstructed targets have a shift-variant point spread function. Images formed using the Fourier array technique always have superior signal-to-noise ratios than scenes imaged using analog beam forming techniques [146, p199].

Radiation patterns of steered and focused arrays can be found on pp207-208 of Soumekh [146].

3.4.3 Single receiver synthetic array

If we have transmit and receive apertures co-located at the origin of the (x, y) plane with real valued illumination functions $i_T(v)$ and $i_R(w)$, where v and w are the axes of the faces of the apertures parallel to the y-axis, then the combined radiation pattern is

$$
\begin{aligned}
H(\omega, x, y) &\approx P_m(\omega) \cdot \int_{-D_T/2}^{D_T/2} i_T(v) \frac{\exp\left(-jk\sqrt{x^2 + (y-v)^2}\right)}{\sqrt{x^2 + (y-v)^2}} dv \\
&\quad \cdot \int_{-D_R/2}^{D_R/2} i_R(w) \frac{\exp\left(-jk\sqrt{x^2 + (y-w)^2}\right)}{\sqrt{x^2 + (y-w)^2}} dw \\
&\approx P_m(\omega) \cdot A_t(\omega, x, y) \cdot A_r(\omega, x, y) \cdot \frac{\exp\left(-j2k\sqrt{x^2 + y^2}\right)}{(x^2 + y^2)},
\end{aligned} \tag{3.29}
$$

where the two-way phase path is due to the time taken for the transmitted signal to travel from the center of transmit aperture, out to the point (x, y) and back to the center of the receive aperture. The apertures of length D_T and D_R act as low-pass spatial filters that limit the spatial bandwidth transduced by the active imaging system. The overall low-pass effect of the two apertures is dominated by the longest aperture, throughout this thesis, unless otherwise indicated, the symbol D refers to the longest array, i.e., $D = \max(D_T, D_R)$, or refers to the aperture length in systems that employ a single aperture for transmission and reception. The dispersion term, $(x^2 + y^2)$, in the denominator of (3.29) is generally accounted for by time varying gain circuits in the receiver, so is dropped in the following mathematical development.

If the active imaging system is moved to a point u, then the radiation pattern is given by $H(\omega, x, y - u)$. The output of the system at point u, to the echoes from a frequency independent, aspect independent

target field of reflectivity $ff(x, y)$, is modelled as

$$
Ee_m(\omega, u) \approx \int_x \int_y ff(x, y) H(\omega, x, y - u) dx dy
$$

$$
= P_m(\omega) \cdot \int_x \int_y ff(x, y) \cdot A_t(\omega, x, y - u) \cdot A_r(\omega, x, y - u) \cdot \exp\left(-j2k\sqrt{x^2 + (y - u)^2}\right) dx dy.
$$

(3.30)

Alternatively, in the time domain this is

$$
ee_m(t, u) \approx \int_x \int_y ff(x, y) \cdot \left[a_t(t, x, y - u) \odot_t a_r(t, x, y - u) \odot_t p_m\left(t - \frac{2\sqrt{x^2 + (y - u)^2}}{c}\right) \right] dx dy.
$$

(3.31)

In the Fourier domain of the aperture, or u-axis, (3.30) becomes

$$
EE_m(\omega, k_u) \approx P_m(\omega) \cdot A(k_u) \cdot \int_x \int_y ff(x, y) \cdot \sqrt{\frac{\pi x}{jk}} \cdot \exp\left(-j\sqrt{4k^2 - k_u^2} \cdot x - k_u y\right) dx dy.
$$

(3.32)

The inversion of this model is the basis of the wavenumber processors discussed in Chapter 4, Section 4.3.3. The sampled nature of this synthetic aperture model is discussed briefly in Section 3.5 and in detail in Chapter 5.

The appearance of the *two-way* phase factor in the delay exponential of (3.29) means that the rate of change of the phase in (3.30)—i.e., the instantaneous spatial frequency (Doppler wavenumber)—increases at twice the rate of the *one-way* system. This is why a moving active imaging system has twice the spatial bandwidth of a bistatic arrangement where only one aperture moves or a passive system where only the receiver moves. The overall amplitude pattern of the apertures scales as $k_u = 2k \sin \theta = 2ky/\sqrt{x^2 + y^2}$ and

$$
A(k_u) = A_T(k_u/2) \cdot A_R(k_u/2).
$$

(3.33)

The extra factor of two doubles the spatial bandwidth (Doppler wavenumber bandwidth) of each aperture to $B_{k_u} = 4\pi/D$ where $D = D_T$ or D_R. For example, the amplitude pattern of a uniformly illuminated transmitter, $A_t(\omega, x, y) = \text{sinc}[kD \sin \theta/(2\pi)]$, scales to $A_T(k_u/2) = \text{sinc}[k_u D_T/(4\pi)]$.

The doubling of the spatial bandwidth of the two-way system means that a focused synthetic array has twice the angular resolution of a real array; that is,

$$
\theta_{3\text{dB}}^{\text{synthetic}} = \alpha_w \frac{\lambda}{2L_{\text{sa}}},
$$

(3.34)

where the constant α_w reflects the amplitude weighting due to the combined effect of the amplitude patterns of the real apertures.

3.4.4 Multiple receiver synthetic array

Figure 3.2(d) shows a synthetic aperture that employs a single transmit aperture and multiple receivers. The use of multiple receivers has been seen as a method for satisfying the along-track sampling requirements of synthetic aperture systems. The transmit aperture sets up a spatial wave field which is sampled by the multiple receivers. For any given pulse, the transmitter-to-target phase path remains fixed; however the target-to-receiver paths vary for each receiver. These phase differences must be accounted for in the synthetic aperture processor; either by adjusting the phase of each receiver to that of an equivalent single receiver system [78], or by forming preliminary partially focused images using the outputs of real receive array for each transmitted pulse [43].

Multiple receiver systems are discussed in more detail in Chapter 5.

3.5 SYNTHETIC APERTURE GRATING LOBE SUPPRESSION

In 1978 Tomiyasu published a tutorial paper on strip-map SAR [153]. In his paper, he described several methods which all determined that along-track samples spaced every $\Delta_u = D/2$ were adequate to sample a synthetic array (D being the length of the transmit/receive array). To date, no publications have appeared which refute his analysis, even though it has been obvious to a number of investigators that sample spacings of $D/2$ are barely adequate in many situations [31, 132].

The spatial sampling requirement is classically referred to as $D/2$ in the SAR literature due in part to an incorrect interpretation of the grating lobe suppressing effect of the real aperture radiation pattern [153]. Figure 3.3 clearly explains the origin of this confusion; Fig. 3.3(a) can be interpreted as the array factor of the synthetic aperture at frequency ω, this figure is formed by the along-track compression (phase matching) or focusing of the along-track signal at the top of Fig. 3.3(b). This along-track compression forms a main lobe at $u = 0$ and grating lobes spaced at $n\Delta_g$ where $|n| = 1, 2, \ldots$ and $\Delta_g = x\lambda/(2\Delta_u)$. The first null of the real aperture radiation pattern occurs at $u = x\lambda/D$. The sample spacing required for this null to suppress the grating lobe is $\Delta_u = D/2$.

This interpretation assumes an *incorrect* sequence of events. The correct sequence is shown in Fig. 3.3(b), the along-track phase function at the top of Fig. 3.3(b) is weighted by the real aperture radiation pattern, giving the signal shown in the center. To determine the point at which aliasing occurs requires the instantaneous position of the phase function in the (ω, k_u)-domain (3.32), $u_i = \partial\phi(k_u)/\partial k_u \approx k_u x/(2k)$. At the folding Doppler wavenumber, $k_u = \pi/\Delta_u$ the instantaneous signal is at spatial location $u_i = x\lambda/(4\Delta_u) = \Delta_g/2$, after this point the signal aliases. The along-track signal passes
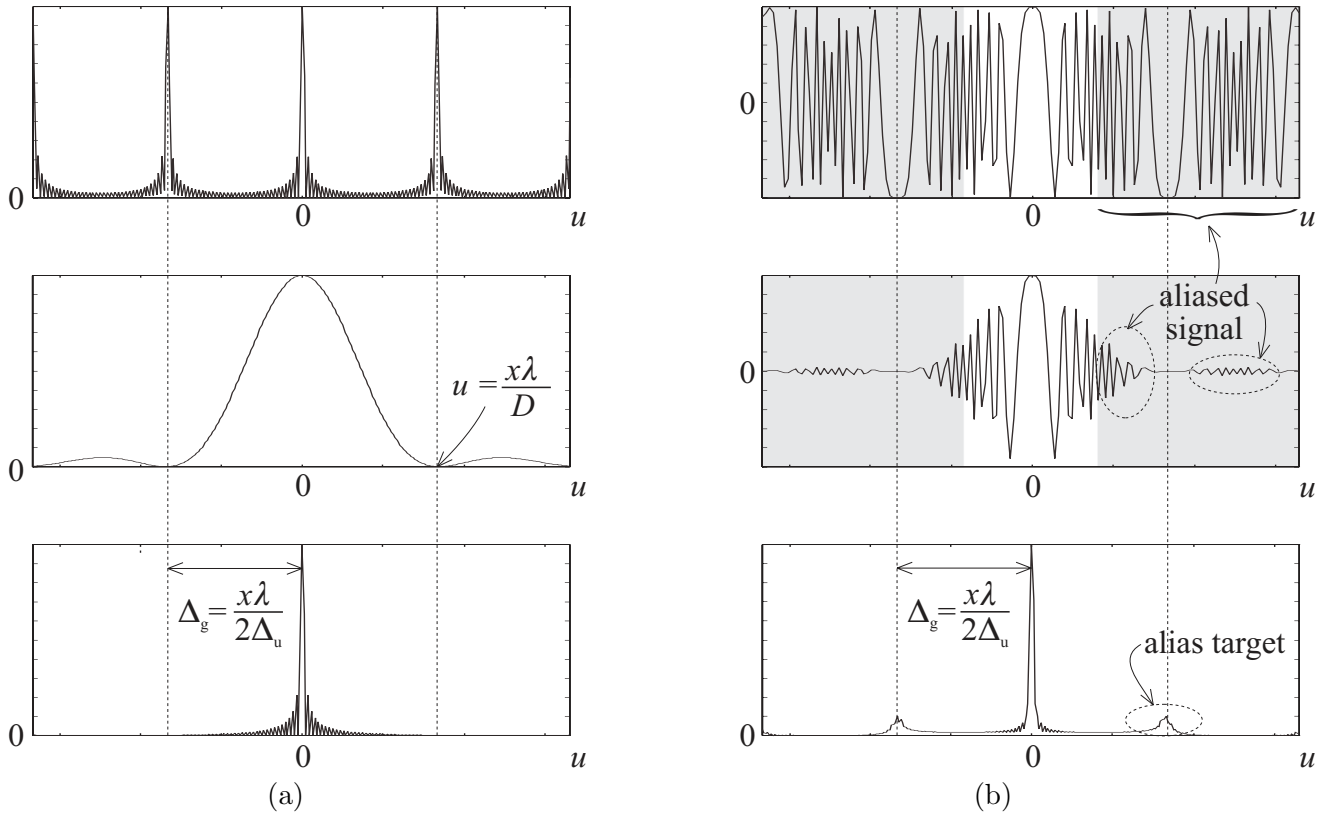
Figure 3.3   The effect of the real aperture radiation pattern on along-track undersampling, (a) classical grating lobe suppression interpretation, (b) correct interpretation. The aliased signal indicated in grey in (b) corresponds to along-track locations where the PRF is not high enough to adequately sample the along-track phase (the repeated low frequency components away from $u = 0$ are where the along-track signal aliases to $k_u = 0$).

through $k_u = 0$ again at multiples of $u = \Delta_g$. With an along-track sample spacing of $D/2$, the effect of the real aperture pattern on the aliased signal is to suppress the terms at $|u| = n\Delta_g$ and 'attenuate' the aliased signals near the grating lobes. When this signal is along-track compressed (focused), the aliased energy near the first grating lobe is compressed into an alias or grating lobe target that is 10 percent of the main lobe height. To reduce this alias level, the along-track sample spacing should ideally be $\Delta_u = D/4$. The classical interpretation incorrectly assumes that *all* the aliased energy is compressed into the grating lobe peak before the null of the real aperture radiation pattern suppresses it. The interpretation is correct for a real aperture array because all the elements are energised simultaneously, however, for a synthetic aperture system this is not the case.

Aperture sampling from a Doppler wavenumber bandwidth perspective, covered in Chapter 4, also results in the same $D/4$ along-track sampling requirement. The more detailed investigation of along-track sampling in Chapter 5, using the along-track ambiguity to signal ratio (AASR) and the peak to grating lobe ratio (PGLR), shows that if $D/2$ along-track resolution is required in a final image estimate

that has the *highest possible dynamic range*, then a sample spacing of at least $D/3$ is required to force the level of the grating lobe targets beneath the sidelobe levels generated by a point target. This $D/3$ or $D/4$ sampling requirement does not mean that $D/2$ sampled systems do not work, what it does imply is that $D/2$ sampled systems have aliased part of the mainlobe energy during data collection. This aliased energy results in grating lobe targets either side of the main target response which limit the dynamic range of the final image estimate. In airborne SAR systems, this never happens as the PRF is so high that the spatial bandwidth of the aperture is always well sampled. In spaceborne SARs and in SAS systems that operate at a sample spacing of $D/2$, the reduced sampling rate is based on the commercial requirement to map as fast as possible, however, this increase in mapping rate comes at the expense of reduced dynamic range in the final image estimate.

3.6 ADVANTAGES/DISADVANTAGES OF SEPARATE TRANSMITTER AND RECEIVER ARRAYS

Advantages

Lower sidelobes in the angular response; in a radar system employing the same array for transmission and reception the monochromatic amplitude pattern for uniform illumination of an aperture of length D is $A(\omega, x, y) = \text{sinc}^2(kD \sin\theta/(2\pi))$. When separate arrays of length D_T and length D_R are used the amplitude pattern is $A(\omega, x, y) = \text{sinc}(kD_T \sin\theta/(2\pi)) \text{sinc}(kD_R \sin\theta/(2\pi))$, if we choose the ratio of the array lengths to be 0.717 then the two highest sidelobes in the combined monochromatic pattern are 5.2dB lower than the -26.8dB sidelobe in the sinc-squared monochromatic pattern of the single transmitter/receiver array. Combining the multiple array arrangement with wide bandwidth pulse compression results in a spatial-temporal response with very low sidelobes in the correlated beam pattern.

Continuous illumination; with a separate transmitter and receiver it is possible to continuously illuminate an area of interest. This continuous illumination may smooth out the aspect dependence of some targets and hopefully ensures that no targets are missed.

Disadvantages

Cross-talk; while the transmitter is 'on' the transmitted signal follows a fairly direct path to the receiver. The signal received due to this cross-talk path has a dynamic range which may exceed the dynamic range of some target returns. Adaptive filters that cancel this cross-talk are an active area of research.

Chapter 4

SYNTHETIC APERTURE IMAGING ALGORITHMS

Synthetic aperture imaging algorithms can be broadly classified into three groups; spatial-temporal domain processors [7, 31, 47], range-Doppler processors [31, 84, 166], and wavenumber processors [6, 20, 128]. The mathematical notation used in synthetic aperture references to describe each system model is wide and varied. By applying multi-dimensional signal processing techniques, each of these synthetic aperture algorithms is described within a coherent and unified mathematical framework. The use of mapping operators to emphasize the coordinate transforms involved in each algorithm is an important step in clarifying their operation. By representing each system model in a consistent framework a comparison of each algorithm can be made and the relationship between strip-map and spotlight modes can be determined. It also allows a comparison of similar inversion schemes found in fields such as holography, computerized tomography (CT), magnetic resonance imaging (MRI), x-ray crystallography, radio astronomy, and seismics. Another objective of this chapter is to represent these algorithms to the accuracy required for direct digital implementation in realistic synthetic aperture processors.

This chapter begins by giving an overview of what a synthetic aperture system is trying to achieve. The *spatially-induced* strip-map synthetic aperture model and the various inversion algorithms are developed in what is effectively a *chronological* order. This order is; the spatial-temporal domain processor, the range-Doppler algorithm, the wavenumber algorithm, and the chirp scaling algorithm. This is followed by the spotlight synthetic aperture model and its inversion schemes. *Velocity-induced* FM-CW imaging inversion is presented and related to the spatially-induced model. Finally, slant-to-ground plane conversion, multilook processing, and the practical implementation of a parallel processing unit is discussed.

(a)   (b)   (c)   (d)
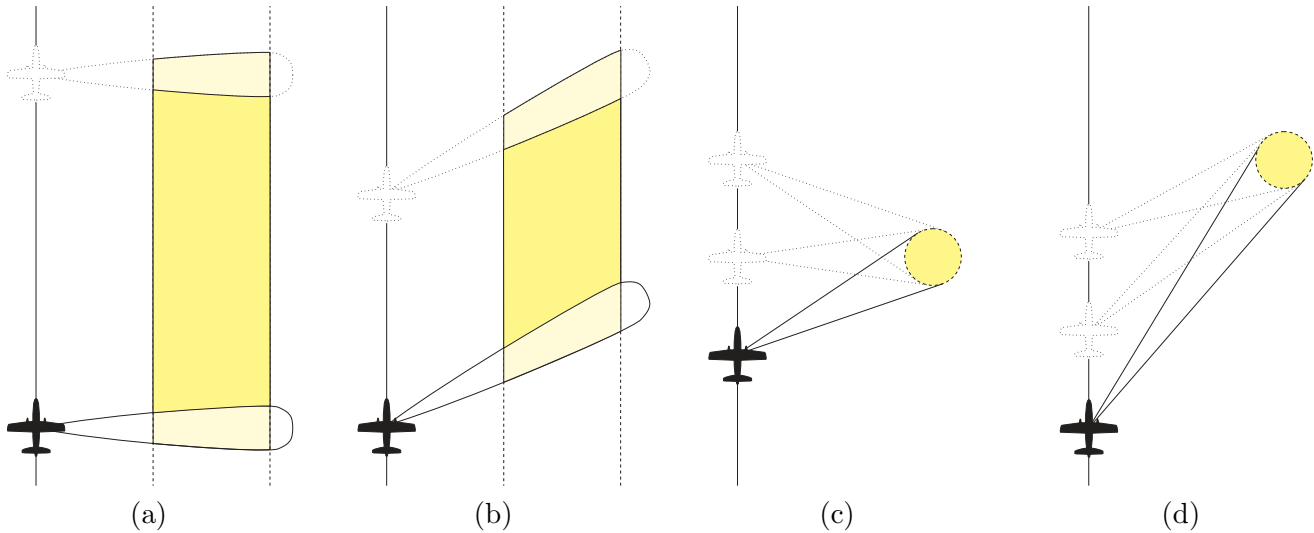
Figure 4.1   Synthetic aperture imaging modes, (a) broadside strip-map, (b) squinted strip-map, (c) broadside spotlight, and (d) squinted spotlight. The light shading is the collected data, the dark shading is the valid data after processing.

4.1   OVERVIEW OF INVERSION

4.1.1   Synthetic aperture imaging modes

The objective of a synthetic aperture imaging system is to obtain an estimate of the acoustic or electromagnetic reflectivity of the area of interest. Figure 4.1 shows the modes commonly employed by synthetic aperture systems. Figures 4.1(a) and 4.1(b) show broadside and squinted (forward looking) strip-map systems. The light shading in the figures corresponds to the area imaged, while the darker shading represents the valid data after processing. When using a strip-map system to image a small target area, a large amount of the recorded data is discarded after processing. Spotlight mode synthetic aperture systems steer the real aperture beam pattern to continuously illuminate the same target area. This beam steering results in ambiguity constraints that are less severe than those for conventional strip-map systems and allows small areas to be imaged without wasting data [82, Chapter 2]. Figures 4.1(c) and 4.1(d) show broadside and squinted spotlight systems. In a squinted strip-map system the squint angle is the same as the steer angle of the beam pattern from boresight. In spotlight mode, the squint angle is defined as the angle from the center of the synthetic aperture to the center of the target area. Squinted synthetic aperture systems are not considered within the scope of this thesis; the background references indicated in Section 1.5 contain information on these systems.

4.1.2 Offset Fourier (holographic) properties of complex images

To obtain an estimate of an object's reflectivity, synthetic aperture systems first obtain an estimate of the objects *Fourier transform* over a limited set of spatial frequencies. The *complex image estimate* is the inverse Fourier transform of this data, and the *image estimate* is the magnitude (or magnitude-squared) of the complex image estimate. The Fourier data obtained by the *coherent* imaging system is termed *offset Fourier data* [115]. This term reflects the fact that Fourier data obtained by the imaging system is not centered about the wavenumber domain origin, but is centered about a wavenumber pair determined by the carrier of the transmitted waveform and the squint angle of the aperture (note that the Fourier transform of a real valued image would normally consist of a dense number of samples centered around the origin in the wavenumber domain).

To determine what is imaged when dealing with offset Fourier data consider the two-dimensional *object reflectivity function*

$$ff(x, y) = |ff(x, y)| \exp\left[j\phi\phi(x, y)\right]. \tag{4.1}$$

If the *phase* of (4.1) is *highly random*, then the Fourier transform of the *phase alone* is extremely broad so that the transform of $ff(x, y)$ (which can be considered to be the convolution of the transforms of the amplitude and phase functions individually) contains magnitude information over a large part of the frequency domain. Thus, the *random phase modulates the magnitude information* over a wide region of Fourier space so that the *magnitude of the reflectivity* may be recovered from offset Fourier data [115]. This is essentially the same idea that is used to generate holograms.

Given this scattering model, it is possible to recover an *image estimate* or the *diffraction limited image*, $|\widehat{ff}(x, y)|$, by obtaining a section of offset Fourier data $\widehat{FF}(k_x, k_y)$. The location of this offset data has no effect on the target surface and shape information contained in the final image [115], however, the final resolution of the image depends on the wavenumber bandwidth of the offset Fourier data [146, p181-183].

The ability of synthetic aperture systems to generate accurate images from a small portion of the image's offset Fourier data is due to the coherent nature of synthetic aperture data [115]. In non-coherent medical imaging applications such as CAT or CT scanners, a full circumference of Fourier data is required before an image can be generated (see Chapter 2 [82]).

The complex valued (real and imaginary) nature of signals in this thesis means that every signal can be referred as an *offset Fourier* signal, eg., the pulse compressed rect spectrum can appear anywhere in Fourier space, yet the magnitude of its inverse transform remains the same. This thesis restricts the use of *offset Fourier* to refer to the windowed wavenumber signal $\widehat{FF}(k_x, k_y)$. The coherent data encountered in synthetic aperture systems is recorded either as a baseband signal or a modulated signal,

	Strip-map Mode	Spotlight mode
Range resolution, $\delta x_{3\mathrm{dB}}$	$\dfrac{c}{2B_c}$	$\dfrac{c}{2B_c}$
Along-track resolution, $\delta y_{3\mathrm{dB}}$	$\dfrac{D}{2}$	$\dfrac{\lambda_0}{2\Delta\theta}$

Table 4.1   'Classic' image resolution in strip-map and spotlight synthetic aperture images.

this raw data then undergoes processing, via interpolations or phase multiplies, to produce the offset Fourier data centered around baseband in the wavenumber domain. This baseband signal is still correctly referred to as offset Fourier data, even though there is no apparent offset from the origin—this baseband data represents a window of the wavenumber domain centered on the carrier wavenumber and squint wavenumber, and *any* block of *this* Fourier data can be extracted, from any location, and inverse Fourier transformed to produce a lower resolution complex image estimate. This last property is the essence of *multi-look processing*, whereby different blocks of offset data known as *looks* are inverse transformed and added incoherently to give a lower resolution image, with improved speckle characteristics.

4.1.3   The implications of digital processing

The 'classic' resolution obtained in the images produced from strip-map and spotlight systems is given in Table 4.1, in the table; c is the speed of the electromagnetic or acoustic waves, B_c is the pulse bandwidth, D is the effective length of the real apertures, λ_0 is the carrier wavelength, and $\Delta\theta$ is the angular diversity of the spotlight system, i.e., the range of slew or steer angles of the real beam pattern. The detailed analysis of each mode in this chapter shows that these 'classic' relationships are a good 'rule of thumb' for determining final image resolution.

The extent of the wavenumber bandwidths in the offset Fourier data is directly related to these image resolution relationships via $B_{k_x} = 2\pi/\delta x_{3\mathrm{dB}}$ and $B_{k_y} = 2\pi/\delta y_{3\mathrm{dB}}$ (in rad/m). The range wavenumber bandwidth in both imaging modes is set by the bandwidth of the transmitted pulse, so the sampling requirements of the range (spatial) signal are identical to the sampling requirements of the transmitted pulse. These sampling requirements are met by the receiver system. The along-track bandwidths of the two imaging systems are different. In strip-map mode, the image resolution depends solely on the effective length of the real apertures, therefore the along-track sampling rate also depends solely on this length. The along-track resolution would seem to indicate an along-track sample spacing of $D/2$ would be adequate to correctly sample the along-track signal. This sample spacing is classically referred to by most strip-map SAR references as being the along-track Nyquist sample spacing. This statement

is a misnomer, resolution of $D/2$ is achieved by processing the along-track Doppler wavenumber signal between the 3dB width of the radiation pattern. The null-to-null width of the radiation pattern in the Doppler wavenumber domain is $B_{k_y} = 8\pi/D$, so a more appropriate 'Nyquist' sample spacing of $D/4$ would seem to be necessary. Chapter 5 shows how a sample spacing of $D/3$ combined with processing only the 3dB width of the radiation pattern yields image estimates with the highest dynamic range. As long as the real apertures are shaded to reduce the sidelobes in the overall aperture beam pattern, minimal aliasing occurs. The choice of a $D/2$ spacing by conventional SAR systems results in alias targets appearing in some images [98, 132]. Airborne systems typically sample the along-track signal so frequently that aliasing is not a problem [132]. Section 4.3.1 discusses these sampling requirements in more detail.

The beam steering used in spotlight mode decouples the real aperture length from the along-track resolution. The along-track resolution is instead determined by the synthetic aperture length and the range from the scene center. The real aperture length then only acts to set the illuminated patch diameter and it still sets the along-track sampling requirements necessary to adequately sample this patch. This along-track sample spacing still needs to be at least $D/4$ to adequately sample the Doppler spectrum of the targets within the spotlight patch (this sampling requirement is not a problem for airborne spotlight SAR systems). The aperture length of a spotlight system can be much longer than that of a strip-map system, often though, the same system is used for both strip-map and spotlight applications.

Several of the effects of a digital implementation of synthetic aperture processing are often ignored or not mentioned in the literature. The first relates to the phase matching that is a key step in all the synthetic aperture inversions. This phase matching is typically applied as a multiplication in a temporal or spatial frequency (wavenumber) domain. In the reciprocal time or space domain, this multiplication acts as a *cyclic* convolution (due to the discrete nature of the sampled signal). The effect of this convolution in terms of image distortion must be clarified. The image distortion is minimal for the pulse compression and azimuth compression operations due to the fact that each phase matching operation acts to compress the signal to a main peak with low sidelobes. In this case it seems unnecessary to account for the cyclic effects of these operations in the image domain. However, the range migration correction causes rotation in the range direction of any block that is being processed. The effects of this rotation can be countered by *overlapping* blocks during processing. Zero padding should not be used to mitigate the effects of cyclic convolution in this case, as zeros get rotated into the image! There is also the final *non-linear* operation of image detection (taking the modulus of the complex image estimate), this non-linear operation increases the wavenumber bandwidth of the image [7, p1047]. To minimize the aliasing effects of this non-linear operation, the offset Fourier data should be padded to at least twice the nominal bandwidths $B_{k_x} = 2\pi/\delta x_{3\text{dB}}$ and $B_{k_y} = 2\pi/\delta y_{3\text{dB}}$ before generation of $|\widehat{ff}(x,y)|$.

4.1.4   Terminology; "inversion" or "matched filtering"

The amount of offset Fourier data obtained in all the system models is limited by one or more of the following; the aperture radiation pattern, the range of steer angles, or the transmitted pulse bandwidth. When developing an inversion scheme for the synthetic aperture models it is possible to use either an inverse filter, such as a Weiner filter, or a matched filter. The matched filter usually gives the most robust response. The output of a Weiner filter usually has a higher resolution than a matched filter due to the fact that the Weiner filter emphasizes the information at the edges of the offset Fourier data. However, it is usually not a problem to extend the bandwidth or viewing angles of radar systems to obtain a desired resolution so that a matched filter is typically used. This thesis considers this property to hold for sonar too.

The inversion schemes developed in this thesis are referred to as *inversion* schemes due to the fact that they cannot *correctly* be referred to as matched filters. In a matched filter, the *amplitude* functions of the system model should also be included. However, these amplitude functions are typically ignored in the SAR literature and the inversion schemes are based on matching the *phase* only. This phase matched data is then weighted with a windowing or weighting function to tailor the spectral response such that the image generated has low sidelobes. The reason for ignoring some of these amplitude functions also has a historical significance; in an optical based processor, pulse compression is achieved in the receiver, while the along-track signal processing is performed with an optical system. Because the signal is pulse compressed in the receiver, any weighting function can be applied on reception before recording the signal on film. In the optical processing of the signal film, along-track amplitude matching is achieved with shaded transparencies, while the phase matching is essentially achieved with a conical lens. Transparencies with the correct light-amplitude transmissivity were difficult to construct and resulted in high insertion losses, so they were replaced with two-level transparencies; opaque and transparent [34]. This replacement is mathematically equivalent to replacing the along-track amplitude function with a range variant rect function representing the along-track extent (synthetic aperture length) of a target in the raw data. Thus, the along-track matched filter is really only a *phase* matched filter. Cutrona [34], Brown [17] and Tomiyasu [153] contain excellent discussions on optical processing.

4.1.5   Doppler wavenumber vs spatial Doppler frequency

The along-track signal exploited in synthetic aperture processing is an amplitude weighted complex exponential that can be parameterized in terms of along-track distance u or along-track slow-time $t_2 = u/v_p$. Many references on SAR use the temporal parameter and approximate the square-root phase function of the along-track signal with a quadratic function. This introduces three terms; the along-track Doppler frequency, Doppler centroid, and Doppler rate. The along-track spatial Doppler frequency, f_d (Hz), is the Fourier domain of the along-track temporal parameter t_2 (s). Spatial Doppler

frequency is related to the Doppler wavenumber (rad/m) used in this thesis by [6]

$$f_d = \frac{2\pi k_u}{v_p}.$$ (4.2)

The Doppler centroid is related to the squint angle of the real aperture and it is essentially the carrier term of the along-track Doppler signal. If the real aperture squint is θ_s, then the offset Fourier data are centered on $(k_x, k_y) = (2k_0 \cos\theta_s, 2k_0 \sin\theta_s)$; the Doppler centroid is then given by $f_{d0} = 4\pi k_0/v \cdot \cos\theta_s$. The Doppler rate is the LFM chirp rate about the Doppler centroid frequency, in a broadside orientation for a target at range x_0 it is given by

$$f_{dr} \approx -\frac{k_0 v_p^2}{\pi x_0} = K_a v_p^2,$$ (4.3)

where f_{dr} is in Hz/s or s^{-1} and the spatial Doppler rate, K_a, is in m^{-2} [116]. When dealing with satellite geometries, the Doppler centroid and Doppler rate must be tracked in order to produce fully focused images, this tracking is discussed in [97].

4.1.6 Figures of merit in synthetic aperture images

The SAR community quantifies image quality via the following figures of merit; range resolution $\delta x_{3\mathrm{dB}}$, along-track resolution $\delta y_{3\mathrm{dB}}$, the peak-to-sidelobe ratio (PSLR), the integrated sidelobe ratio (ISLR), the along-track ambiguity to signal ratio (AASR), and the range ambiguity to signal ratio (RASR) [31]. Sample spacings of the input data (Δ_t, Δ_u) and the output image (Δ_x, Δ_y) are also useful parameters. Many of these parameters are predictable from the system parameters, so comparison between the measured quantities and their theoretical limits is possible. The first four metrics are usually calculated from point-like targets in the final focused image estimate, while the AASR and RASR are dependent on the system geometry and target field and are generally only used during the system design stage. The AASR, RASR, and a further quantity, the peak-to-grating lobe ratio (PGLR), are defined and discussed in Chapter 5.

In SAR systems, radiometric calibration and phase preservation are two image parameters that are gaining increasing importance. Radiometrically calibrated images are important when images obtained from the same area, from different systems or the same system at a different time, are to be compared in a quantitative manner [52,92,121]. Accurate image phase data is important in interferometric applications. In sonar applications, the calibration of target acoustic scattering properties and image phase will also become increasingly important as more mapping and bathymetric (interferometric) synthetic aperture systems are developed.
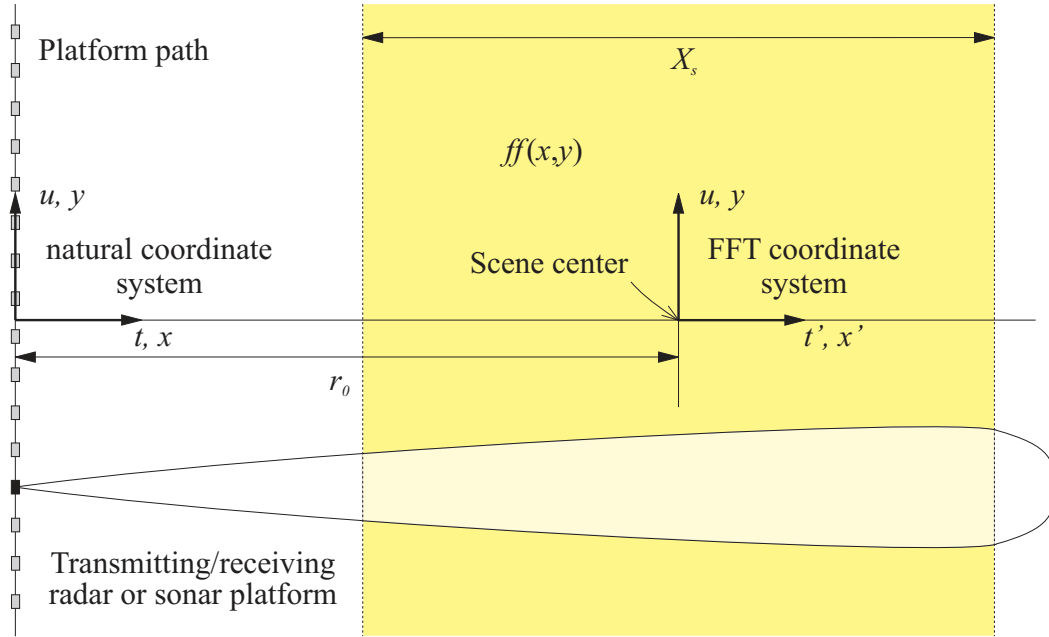
Figure 4.2   Imaging geometry appropriate for a strip-map synthetic aperture system. Note the two coordinate systems; the natural system (x, y) and the translated system necessary for the calculation of spectra via the FFT (see the text for more details).

4.2   SPATIALLY-INDUCED STRIP-MAP SYNTHETIC APERTURE SYSTEM MODEL

Before developing the different inversion algorithms it is necessary to develop a system model. This system model and its inversion are developed in two-dimensions (2-D) with the platform travelling in the same plane as the target area. The extension of the model and its inversion to the actual three dimensional (3-D) geometry is presented after the inversion schemes have been developed.

Figure 4.2 shows the 2-D geometry appropriate for broadside strip-map mode synthetic aperture imaging systems. The target area described by $ff(x, y)$ is considered to consist of a continuous 2-D distribution of omni-directional (aspect independent) and frequency independent reflecting targets. This target area is illuminated by a side-looking radar or sonar system travelling along a straight locus u, with a velocity v_p, flying parallel to the y axis of the target area. The origins of the delay time axis t and the range axis x have been chosen to coincide for the all the algorithms presented in this thesis. In the spatial-temporal and range-Doppler algorithms, the x origin is conventionally referenced from the platform position. However, in the literature for the wavenumber based algorithms the choice of the x origin varies. The choice of the origin for the wavenumber algorithm has an important impact on the processing that is explained shortly. The time axis is sometimes referred to as *fast-time*, while the platform path or aperture u can be divided by the platform velocity v_p to give the *slow-time* axis.

These terms reflect the fact that the signal travels out to the maximum range much faster than the platform traverses the synthetic aperture length. As the platform travels along u, it transmits a wide bandwidth phase modulated (i.e., spread in time) waveform $p_m(t)$ of duration τ_p which is repeated every τ_rep ($\tau_\mathrm{p} \leq \tau_\mathrm{rep}$). On reception, the coherent nature of the transmitter and receiver allows the reflections/echoes that have come from different pulses to be arranged into a 2-D matrix of delay time t versus pulse number. Since the platform ideally travels a constant distance between pulses, the pulse number can be scaled to position along the aperture u in meters. Assuming that the 'stop-start' assumption discussed in Section 2.8 holds, the *strip-map system model* representing the echoes detected at the output of the receiver is approximately described by (3.31)

$$
\begin{aligned}
ee_m(t, u) &\approx \int_x \int_y ff(x, y) \cdot a(t, x, y - u) \odot_t p_m\left(t - \frac{2}{c}\sqrt{x^2 + (y - u)^2}\right) dx\, dy \\
&= \int_x ff(x, u) \odot_u \left\{ a(t, x, u) \odot_t p_m\left(t - \frac{2}{c}\sqrt{x^2 + u^2}\right) \right\} dx,
\end{aligned}
\tag{4.4}
$$

where $a(t, x, y)$ is the spatial-temporal response of the combined transmitter and receiver apertures as developed in Chapter 3, and the symbol $\odot$ represents convolution with respect to the subscripted variable. Representing the output of the system as a convolution in along-track emphasizes the two main problems faced by the inversion schemes; the system response is *range variant* and the dependence of the delay inside $p_m(t)$ on along-track position u can cause the signal response to cover many range resolution cells, an effect known as *range curvature* or *range migration*. Any successful inversion scheme must be able to cope with this range variance and be able to correct for the range curvature.

The echoes detected by the receiver are usually quadrature demodulated to give the baseband version of (4.4):

$$
\begin{aligned}
ee_b(t, u) &= ee_m(t, u)\exp(-j\omega_0 t) \\
&\approx \int_x \int_y ff(x, y) \cdot a(t, x, y - u) \odot_t p_b\left(t - \frac{2}{c}\sqrt{x^2 + (y - u)^2}\right) \\
&\quad \cdot \exp\left(-j2k_0\sqrt{x^2 + (y - u)^2}\right) dx\, dy.
\end{aligned}
\tag{4.5}
$$

The system models presented in the literature typically develop spatial-temporal and range-Doppler processors from quadrature demodulated signals. System models derived from a wavenumber perspective vary between the modulated and demodulated notation. Both notations are presented so that the appropriate amplitude and phase functions required for any particular inversion scheme are clear. Whenever modulated data is recorded, for reasons of efficiency, it is recommended that the data is processed into a complex valued baseband format as soon as possible (see the methods recommended in Sections 2.1.1 to 2.1.3). The processing of baseband data is the most efficient method, as it represents

the smallest data set for both the FFT and any interpolators.

Before developing the inversion schemes it is useful to look at the temporal Fourier transform of the system model. A temporal Fourier transform of (4.4) or (4.5) gives

$$Ee_m(\omega, u) = P_m(\omega) \cdot \int_x \int_y ff(x, y) \cdot A(\omega, x, y - u) \cdot \exp\left(-j2k\sqrt{x^2 + (y - u)^2}\right) dx\, dy \qquad (4.6)$$

and

$$Ee_b(\omega_b, u) = P_b(\omega_b) \cdot \int_x \int_y ff(x, y) \cdot A(\omega_b + \omega_0, x, y - u) \cdot \exp\left(-j2(k_b + k_0)\sqrt{x^2 + (y - u)^2}\right) dx\, dy, \qquad (4.7)$$

where $P_m(\omega)$ and $P_b(\omega_b)$ are the modulated and baseband pulse spectra respectively. Note that the phase function and radiation pattern in the baseband model must be calculated at the modulated radian frequencies and wavenumbers given by $\omega = \omega_b + \omega_0$ and $k = k_b + k_0 = \omega/c$. Throughout the models developed in this thesis, radian frequencies and wavenumbers *with* the subscript '0' refer to the carrier terms and radian frequencies and wavenumbers *without* the subscript 'b' refer to the modulated quantities. Many references dealing with demodulated quantities refer to the baseband quantities as ω or k and define any terms involving the carrier in terms of $\lambda = 2\pi/k_0$.

Many synthetic aperture systems perform pulse compression of the received reflections 'on the fly' in the receiver before storage. The *pulse compressed strip-map system model* is

$$\begin{aligned} ss_m(t, u) &= \int_\tau p_m^*(\tau - t) \cdot ee_m(\tau, u)\, d\tau \\ &= p_m(t) \star_t ee_m(t, u) \end{aligned} \qquad (4.8)$$

or

$$\begin{aligned} ss_b(t, u) &= \int_\tau p_b^*(\tau - t) \cdot ee_b(\tau, u)\, d\tau \\ &= p_b(t) \star_t ee_b(t, u), \end{aligned} \qquad (4.9)$$

where $\star_t$ denotes correlation with respect to time. The temporal Fourier transforms of (4.8) and (4.9) are

$$Ss_m(\omega, u) = |P_m(\omega)|^2 \cdot \int_x \int_y ff(x, y) \cdot A(\omega, x, y - u) \cdot \exp\left(-j2k\sqrt{x^2 + (y - u)^2}\right) dx\, dy \qquad (4.10)$$

and

$$Ss_b(\omega_b, u) = |P_b(\omega_b)|^2 \cdot \int_x \int_y ff(x, y) \cdot A(\omega, x, y - u) \cdot \exp\left(-j2k\sqrt{x^2 + (y-u)^2}\right) dx\, dy, \qquad (4.11)$$

where the baseband model now contains modulated and baseband quantities, and $|P_m(\omega)|^2$ and $|P_b(\omega_b)|^2$ are the pulse compressed spectra. The pulse compressed spectra are envelope functions that control the level of the range sidelobes in the final image. Given that it is sometimes necessary to transmit uniform magnitude pulses (see the discussion in Section 2.6), the pulse compressed spectrum may need a window function applied to it to smooth its spectral response. The point at which this window is applied is discussed during the development of each inversion scheme.

The signals encountered when modelling synthetic aperture systems are characterized by amplitude modulated-phase modulated (AM-PM) functions, eg., the transmitted LFM pulses and as is seen shortly, the along-track signals. The amplitude functions of these signals are slowly varying, so the signal bandwidth is set by the phase modulating function. Thus, it is the phase functions of the synthetic aperture models that are important when developing inversion schemes or when determining sampling and ambiguity constraints.

4.3 STRIP-MAP INVERSION METHODS

4.3.1 Spatial-temporal domain and fast-correlation processing

The along-track convolution in (4.4) causes the target responses in the pulse compressed raw data to spread in the along-track direction. This target spreading can be focused by matched filtering the pulse compressed data with the system range variant point spread function (PSF) or impulse response. The point spread function (or filter function) can be developed as follows; the quadrature demodulated return from a target located at $(x, y) = (x_0, 0)$ for an infinite bandwidth pulse is

$$\begin{aligned}
ss_\delta(t, u) &= ss_b(t, u)|_{ff(x,y)=\delta(x-x_0, y)} \\
&= a(t, x_0, u) \odot_t \delta\left(t - \frac{2}{c}\sqrt{x_0^2 + u^2}\right) \exp\left(-j2k_0\sqrt{x_0^2 + u^2}\right).
\end{aligned} \qquad (4.12)$$

Shifting this function to the origin and ignoring the aperture effects gives the *point spread function* necessary for focusing:

$$\begin{aligned}
pp(t, u; x_0) &= \delta\left[t - \frac{2}{c}\left(\sqrt{x_0^2 + u^2} - x_0\right)\right] \exp\left[-j2k_0\left(\sqrt{x_0^2 + u^2} - x_0\right)\right] \\
&= \delta\left[t - \frac{2}{c}\Delta R(u; x_0)\right] \exp\left[-j2k_0\Delta R(u; x_0)\right]
\end{aligned} \qquad (4.13)$$

The *range migration locus* in the spatial-temporal domain for the point target at range x_0 is

$$\Delta R(u; x_0) = \sqrt{x_0^2 + u^2} - x_0$$
$$\approx \frac{u^2}{2x_0},$$

(4.14)

where the quadratic approximation is used so that the spatial Doppler bandwidth can be determined.

The properties of the PSF are commonly interpreted using the quadratic approximation to the range migration locus,

$$pp(t, u; x_0) \approx \delta\left(t - \frac{u^2}{cx_0}\right) \exp\left(-j\frac{k_0 u^2}{x_0}\right)$$

(4.15)

This PSF determines the beam forming operation required to focus a target along the boresight of the synthetic aperture at range x_0. In this sense, this (quadratic) PSF is similar to the radiation pattern of an aperture as predicted by classical aperture theory. Inversion schemes based on this PSF are then an approximation. The wavenumber inversion scheme in Section 4.3.3 is based on Fourier array theory and it represents a more accurate inversion of the system model.

Interpretation of the PSF can be broken up into two parts; range effects and along-track effects. The variation of the PSF with range is twofold; the overall function depends on x_0 so that the PSF can be referred to as a *range variant* PSF and the second range variation is due to the delay term in the pulse envelope. This delay term causes targets to sweep out range variant curves known as *loci* in the recorded data. This variation is known as *range migration* or *range curvature*. Figure 4.3(a) shows the simulated pulse compressed echoes from four targets imaged using the Kiwi-SAS parameters (see Table 7.4). The range-variant range migration loci of the four targets are clearly shown. The along-track signal is a spatial phase modulated waveform that lies along the range curvature delta-locus. This phase function depends on the carrier and range and depends quadratically on the along-track direction u. This spatial chirp produces an instantaneous Doppler wavenumber of

$$k_{ui}(u) = \frac{d\phi}{du} \approx -\frac{2k_0}{x_0}u = 2\pi K_a u.$$

(4.16)

The spatial chirp rate

$$K_a \approx -\frac{k_0}{\pi x_0}$$

(4.17)

(in m^{-2}) is the spatial analogue of the Doppler rate (1.8). This spatial chirp acts in an analogous way to pulse spreading and disperses the targets in the along-track direction. The main lobe of aperture radiation pattern acts as a window function that limits the majority of the target energy to $u \in$

$[-Y_0/2, Y_0/2]$ where the null-to-null radiation pattern of the aperture has an along-track extent, at range x_0, that is given by

$$Y_0 \approx 2x_0 \frac{\lambda_0}{D} = \frac{4\pi x_0}{k_0 D} \tag{4.18}$$

(The null-to-null width is approximately twice the 3dB width of the radiation pattern). Combining this spatial extent with the spatial chirp rate gives the wavenumber bandwidth of the along-track signal:

$$B_{k_u} \approx 2\pi K_a Y_0 = \frac{8\pi}{D}. \tag{4.19}$$

This wavenumber bandwidth implies that, to avoid spatial aliasing, the sample spacings in along-track must be $\Delta_u \leq D/4$. This sampling requirement is also obtained via the null suppression argument given in Chapter 3, however, Chapter 5 shows that this requirement can be reduced to $D/3$ without loss in image dynamic range. Typically only the 3dB width of the along-track spectrum is processed, $B_p = 4\pi/D$, giving an along-track resolution in the final image of

$$\delta y_{\text{3dB}} = \alpha_w \frac{2\pi}{B_p} = \alpha_w \frac{D}{2}, \tag{4.20}$$

where the constant α_w reflects the weighting effect of the 3dB width of the radiation pattern in along-track.

Initial synthetic aperture processors used spatial-temporal domain processing. Early airborne SAR systems pulse compressed the echoes in the receiver before recording them onto film. The range resolution of these systems was such that the range curvature was so small it did not need to be accounted for. Thus, the image processor only needed to account for the range variant along-track compression. The first airborne SAR processor was developed by Cutrona in 1957 [34]. The conjugate of the PSF given in (4.15) (ignoring range migration) corresponds to an optical system of lenses and transparencies, and the along-track compression operation or correlation operation (multiply and sum, then shift) is performed by *moving* the signal film past this optical SAR processing system (see p116, p120 and p124 of Harger [69] for an example of the lenses, the signal film and an optical processor).

The first spaceborne SAR, Seasat SAR, was launched in 1978. Spaceborne SAR processing typically requires the processor to account for range migration and an effect due to pointing or squint errors known as range walk. To account for these effects, tilted and rotated cylindrical lenses were added to the optical processor (see p149 of Elachi [47]). The main processors for Seasat SAR and the subsequent spaceborne system, SIR-A were optical [47, p127]. Optical processors were seen to have several shortcomings, they had; poor radiometric calibration, long delays due to film development and handling, and poor access to intermediate data such as pixel phase. Advances in digital technology during the 1970's allowed the

processing to become digital.

The operation of a spatial-temporal domain based digital processor can be described as follows; to determine each pixel value in the output image, the corresponding pixel location is found in the raw pulse compressed data. The data along the locus scribed out by the PSF for that point is multiplied by the complex conjugate of the PSF phase and the result is integrated. Since the recorded data is sampled on a rectangular grid, and the locus is a continuously varying function in both t and u, this inversion scheme requires interpolation from the sampled data surrounding the exact position of the locus in t and u. The result of all this processing is the value for a single image pixel and the whole process must be repeated for every image pixel. Clearly this is time consuming. After processing, each point target in the image has a range and along-track resolution given approximately by $\delta x_{3\mathrm{dB}} \times \delta y_{3\mathrm{dB}} \approx c/(2B_c) \times D/2$. See Elachi [47, pp137-138], Curlander and McDonough [31, pp187-189], and Barber [7] for discussions on digital spatial-temporal processors.

The ACID/SAMI SAS system operates with a real-time, spatial-temporal domain, time-delay beamforming algorithm [2]. The exact range migration term in (4.14) is used, so the system focuses using the exact point spread function (4.13) (some literature refers to this method of processing as the *exact* method, however, as this method requires interpolation, it can not be exact). The interpolations required by the time-delay beamforming approach are obtained by oversampling the received signal by 10 times the carrier wavelength of 8 kHz. The real-time image processor requires 16 parallel, high-speed, transputers to handle the high data rates produced by this oversampling. For every pulse transmitted and received, the real-time processor focuses the sampled 2-D block of raw data stored in memory and produces a *single along-track line*. This processing method is extremely numerically intensive. In [2], the ACID system operated with an along-track sample spacing of $D/2$, so the data collected suffered (minor) along-track aliasing. The 'ghost' images discussed in [2] are due to the fact that the crab angles experienced by the towfish shift the main beam of the radiation pattern and place more energy over one of the grating target positions; this is why only one ghost target is seen in the images. Adams *et al* [2] give a similar explanation, however, they refer to a figure that has incorrectly assessed the grating lobe suppression effect of the nulls in the real aperture pattern (i.e., they refer to Tomiyasu's incorrect figure, see Section 3.5).

Alternatively, the correlation processing can be implemented using fast Fourier (or fast polynomial [155]) techniques. This method is normally referred to as *fast-correlation* due to its use of the FFT, however, the name is a slight misnomer as it is not particularly fast! Each pixel in the final image is produced by extracting a 2-D block of data corresponding to the pixels' along-track extent and its range migration locus, this block is multiplied by the Fourier transform of the PSF as calculated for *that pixel* and is then inverse Fourier transformed. The wide bandwidth version of the focusing filter function for

a target at range x_0 (arbitrary y_0) is (see Section 4.3.3)

$$BB(\omega, k_u) \approx \sqrt{\frac{k}{k_0}} \cdot \exp\left[j\left(\sqrt{4k^2 - k_u^2} - 2k\right) \cdot x_0\right]. \qquad (4.21)$$

This focusing filter is not the conjugate of the Fourier transform of the PSF, it is an inverse filter based on the transform. The inverse filter removes wide bandwidth amplitude modulating effects. (For cases where the conjugate of the Fourier transform of the PSF is used, the processor is sometimes termed the *exact transfer function* (ETF) algorithm). The dependence of (4.21) on the range parameter means that a new filter has to be generated for every range-line; hence, the final image is built up on a range-line by range-line basis. By exploiting the depth of focus of the synthetic array (see (3.22)) the updating of the focusing filter can be relaxed. If only one synthetic aperture length of data is stored before processing, the output of each block correlation of *many* pixels is *one* processed pixel. If the processor allows a reasonable along-track distance to be sampled before processing begins, then the convolutional nature of the along-track signal means that each range line is focused for an along-track extent that depends on target range. Targets at the edges of the along-track block are not fully focused as their full synthetic aperture length was not contained in the processor memory during focusing (see the discussion in Section 4.8). The windowing operations described in Section 4.3.3 should also be applied to the Fourier domain before producing the final image. Chapter 7 contains images processed using the fast-correlation algorithm.

Hayes and Gough, 1992 [74], and Hawkins and Gough, 1995 [72] describe SAS processors based on fast-correlation. Rolt's thesis details a time-delay and sum beamforming processor [133]. Beamforming is identical to processing the signals in the spatial-temporal domain, therefore it represents a sub-optimal method of processing the synthetic aperture data. Examples of optical, digital and optical/digital hybrids employing time-domain and fast-correlation techniques are shown on pages 136-150 of Elachi [47].

The spatial-temporal algorithm then, is not particularly efficient, the updating of the PSF for every range index and use of an interpolator make this a very slow algorithm in the time domain, and only marginally faster in the frequency domain. By exploiting two aspects of the system point spread function, the comparatively efficient range-Doppler algorithm is obtained.

4.3.2   The range-Doppler algorithm

The range-Doppler algorithm was originally developed by Wu, at the Jet Propulsion Laboratories, in 1976 [30, 166] for processing spaceborne SAR data. In their 1982 paper Wu *et al* [166] determined that the PSF of the quadrature demodulated uncompressed raw data could be split up into a one dimensional pulse compression operation, followed by the two dimensional correlation of the then pulse

compressed data with the quadratic approximation of the PSF shown in (4.15). They determined that this correlation could be applied most efficiently as a frequency domain fast convolution in the azimuth direction (i.e., in Doppler wavenumber), with a time domain convolver type of operation in the range dimension to handle the range migration, hence the name *range-Doppler* [166, p568]. In 1984, Jin *et al* [84] improved on the original algorithm so that it could be used to process data from squinted systems. This improvement, called secondary range compression (SRC), range compresses an extra (Doppler dependent) LFM range modulation that is induced by the system due to uncompensated effects of the propagating wavefronts. Since its invention, the range-Doppler algorithm has become the world standard for production satellite SAR processors, and by 1992 it had been installed in at least 15 countries [30]. Curlander and McDonough [31], pages 189-209, detail SAR processing using a range-Doppler processor based on the ideas of Wu and Jin in 'classical' range-Doppler notation (i.e. Doppler frequency, Doppler centroid and Doppler rate).

A 1-D Fourier transform of the quadrature demodulated pulse compressed data, $ss_b(t, u)$, in the along-track direction gives the *range-Doppler* data, $sS_b(t, k_u)$. To determine the form of the matched filter in the range-Doppler domain requires the spatial Fourier transform of the PSF:

$$pP(t, k_u; x_0) = \delta \left[t - \frac{2}{c} \Delta R_s(k_u; x_0) \right] \exp \left[-j \left(\sqrt{4k_0^2 - k_u^2} - 2k_0 \right) \cdot x_0 \right]$$
$$\approx \delta \left[t - \frac{x_0}{c} \left(\frac{k_u}{2k_0} \right)^2 \right] \exp \left(j \frac{k_u^2}{4k_0} \cdot x_0 \right)$$

(4.22)

where the *range migration locus* in the *range-Doppler domain* for a target at x_0 is

$$\Delta R_s(k_u; x_0) = x_0 C_s(k_u)$$

(4.23)

and the target at range x_0 scribes out a locus in the time-Doppler matrix given by

$$t(k_u; x_0) = \frac{2}{c} x_0 \left[1 + C_s(k_u) \right]$$

(4.24)

where the *curvature factor* is

$$C_s(k_u) = \frac{1}{\sqrt{1 - \left(\frac{k_u}{2k_0} \right)^2}} - 1$$
$$\approx \frac{1}{2} \left(\frac{k_u}{2k_0} \right)^2$$

(4.25)

The exact versions of (4.22) and (4.25) were derived circa 1992 [6,127], prior to that time the quadratic approximations were used. Bamler [6] discusses the phase aberrations in the range-Doppler algorithm

caused by this quadratic approximation, while Raney's paper derives the exact range-Doppler signal for uncompressed LFM pulses [127]. To keep with the chronological order of this chapter, a discussion on Raney's range-Doppler signal is presented in Section 4.3.4 during the discussion of chirp scaling.

Given the recorded data in the baseband, pulse compressed, range-Doppler domain, the range-Doppler inversion scheme is expressed as two multiplications and a coordinate transformation; that is,

$$\widehat{fF}_b(x, k_y) = W(k_y) \cdot qQ(x, k_y) \cdot \mathcal{T}^{-1}\left\{sS_b(t, k_u)\right\} \tag{4.26}$$

First, $\mathcal{T}^{-1}\{\cdot\}$ defines a coordinate transform that distorts the time axis and *straightens* the range migration loci, effectively decoupling the rows and columns of the range-Doppler matrix [6],

$$x(t, k_u) \equiv \frac{c}{2}\, t\left[1 - C(k_u)\right]$$
$$k_y(t, k_u) \equiv k_u \tag{4.27}$$

This transform has the Jacobian $JJ(x, k_y) \approx 2/c$, which reflects the scaling from a temporal to spatial ordinate.

After the coordinate transformation, along-track compression is performed by the 2-D phase only multiply [6]

$$qQ(x, k_y) \equiv \exp\left[j\left(\sqrt{4k_0^2 - k_y^2} - 2k_0\right)\cdot x\right]$$
$$\approx \exp\left(-j\frac{k_y^2}{4k_0}\cdot x\right), \tag{4.28}$$

followed by a window function $W(k_y)$ that is defined over the along-track processing bandwidth $B_p = 4\pi/D$ rad/m (the 3dB width of the radiation pattern). The range distortion operator in (4.27) is a time domain operation that requires an interpolator with range varying coefficients. Most SAR processors employ 4-8 point interpolation kernels, in an attempt to tradeoff accuracy in favour of processing speed [128]. (Note that the use of the quadratic approximations in (4.25) and (4.28) caused little discernable difference in the broadside Kiwi-SAS simulated images).

The final image estimate obtain via range-Doppler inversion is

$$\left|\widehat{ff}_b(x, y)\right| = \left|\mathcal{F}_{k_y}^{-1}\left\{\widehat{fF}_b(x, k_y)\right\}\right|. \tag{4.29}$$

If the transmitted pulse produces a uniform spectrum, and the along-track window function acts only to remove (deconvolve) the amplitude modulating effect of the radiation pattern across the processed

Doppler bandwidth, then the image estimate is

$$\left| \widehat{ff_b}(x,y) \right| \approx \left| \text{sinc}\left(\frac{2B_c}{c} \cdot x \right) \cdot \text{sinc}\left(\frac{2}{D} \cdot y \right) \odot_{x,y} ff(x,y) \right| \tag{4.30}$$

The final image estimate is a band limited version of the scene reflectivity $ff(x,y)$. The range wavenumber band limit is set by the pulse bandwidth, and the along-track limit is set by the processed Doppler bandwidth, which in turn is set by the overall radiation pattern of the apertures. The range and along-track resolution of this image estimate is $\delta x_{3\text{dB}} \times \delta y_{3\text{dB}} \approx c/(2B_c) \times D/2$. Range and along-track window functions are generally applied to reduce the range and along-track sidelobes in the final image estimate.

In airborne SAR systems that have range migration less than about 30 percent of a range resolution cell, it is not necessary to perform the range migration correction mapping in $\mathcal{T}^{-1}\{\cdot\}$ (i.e., $C(k_u)$ is set to zero). This reduces the range-Doppler algorithm to a simple phase multiply in the range-Doppler domain, making it a very efficient algorithm for processing airborne SAR images. This particular version of range-Doppler is sometimes referred to as *Fresnel approximation-based inversion* [146, p308].

Unfocused synthetic aperture processing is a special case of Fresnel-approximation based inversion. In unfocused processing, the range migration mapping operation is replaced by a window function that restricts the processed along-track extent of any target to $u \in [-L_{\text{uf}}/2, L_{\text{uf}}/2]$, where the unfocused synthetic aperture length, $L_{\text{uf}} = \sqrt{x\lambda_0}$ is set by the criteria that the target migration, ΔR (see 4.14), is less than $\lambda_0/8$. The following range-Doppler window function is applied to remove the energy due the target response that lies outside of the unfocused synthetic aperture length:

$$uU(t,k_u) = \text{rect}\left(\frac{k_u}{4\pi/\sqrt{\lambda_0 x}} \right), \tag{4.31}$$

where $x = ct/2$. Once this window has been applied, the along-track compression phase multiply, (4.28), is applied. The along-track resolution is determined by the Doppler wavenumber bandwidth limiting effect of the windowing function, or equivalently by the space-bandwidth product, as

$$\delta y_{3\text{dB}}^{\text{unfocussed}} = \frac{\sqrt{x\lambda_0}}{2}. \tag{4.32}$$

What are the inherent disadvantages of the range-Doppler algorithm? The mapping operator $\mathcal{T}^{-1}\{\cdot\}$ requires an accurate, and hence slow, interpolator to minimize the phase errors that could be injected during the interpolation process. A second significant disadvantage of the range-Doppler algorithm is that its focusing, phase and registration accuracy deteriorates for wide beam widths and/or high squints [30]. This deterioration is due to incomplete modelling of the synthetic aperture system; the deterministic phase causing this degradation is accounted for by secondary range compression (SRC). SRC has historically only been seen as necessary in squinted SAR systems. Chapter 7 shows that SRC

is also necessary for the broadside (unsquinted) Kiwi-SAS system due the wide spatial bandwidths employed. SRC is usually applied before the coordinate remapping of range time or alternatively it can be applied during the pulse compression step [31, 84]. The details of SRC are discussed when presenting the chirp scaling algorithm in Section 4.3.4. Range-Doppler processors that account for other effects specific to spaceborne and airborne SAR such as, range walk due to earth rotation and antenna squint, and Doppler centroid tracking, are discussed in references [31, 84, 166].

Figure 4.3 demonstrates range/Doppler processing. Figure 4.3(a) shows the simulated pulse compressed echoes from four targets imaged using an imaging system with the same parameters as the Kiwi-SAS (see Table 7.4). Figure 4.3(b) shows the pseudo-Fourier data obtained via a Fourier transform of the pulse compressed data in the along-track direction, i.e., the range-Doppler data. Note how the loci of the two nearest targets overlay perfectly and interfere coherently. Note also the broadening of all the loci at extreme wavenumbers. This broadening is caused by the induced LFM component that is compensated for by SRC. Figure 4.3(c) shows how the mapping $\mathcal{T}^{-1}\{\cdot\}$ has straightened the range dependent loci and SRC has removed the range spreading. Figure 4.3(d) shows the final image estimate. The impulse response for each target in the image estimate is *identical*.

4.3.3   The wavenumber (ω-k_u) algorithm

The original wavenumber algorithm appeared in the radar literature circa 1991 [20] and, because of its origins in geophysics, it was referred to as *seismic migration* [151]. The formulation of this seismic migration algorithm in terms of the wave equation led to initial scepticism of the algorithm as its relationship to traditional focusing methods was not obvious [6]. In 1992, Bamler [6] and Soumekh [144] came up with similar inversions schemes also based on the wave equation. Bamler developed his work to explain and compare the seismic migration algorithm to traditional processing methods. Soumekh appears to have developed his inversion scheme independently using a Fourier decomposition of the spherical wave functions involved. Soumekh's 1994 book, *Fourier Array Imaging* [146], details the application of these wave equation principles to active and passive imaging modalities such as; phased arrays, synthetic apertures, inverse synthetic apertures and interferometric radio astronomy. The wavenumber algorithm is also known as the *range migration algorithm* [21].

The key to the development of the wavenumber processor was the derivation of the two-dimensional Fourier transform of the system model without needing to make a quadratic approximation. By using a Fourier decomposition or by using the principle of stationary phase, the following Fourier transform pair can be derived (see Appendix A) [146, 147]:

$$\mathcal{F}_u\left\{\exp\left(-j2k\sqrt{x^2+(y-u)^2}\right)\right\} \approx \sqrt{\frac{\pi x}{jk}} \cdot \exp\left(-j\sqrt{4k^2-k_u^2}\cdot x - jk_u y\right) \qquad (4.33)$$
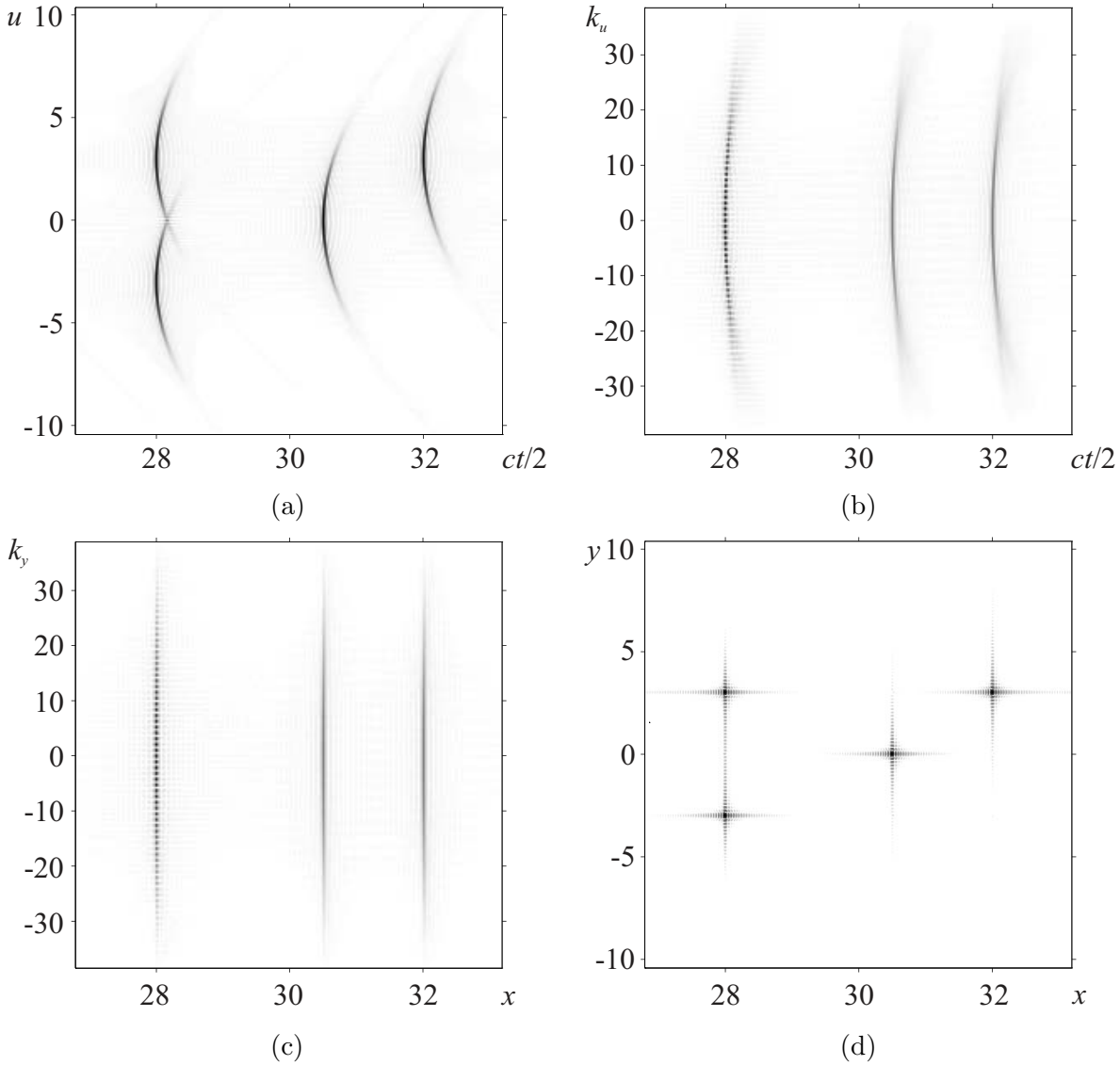
Figure 4.3   Range-Doppler processing. (a) The simulated baseband pulse compressed echo data $|ss_b(t, u)|$, (b) the pulse compressed data in the range-Doppler domain $|sS_b(t, k_u)|$, note the interference in the first locus due to the interference of the two targets at identical range and also note the slight spreading of the locus at extreme wavenumbers. This spreading is due to the induced LFM component that is compensated by SRC. (c) the data after the $\mathcal{T}^{-1}\{\cdot\}$ mapping has straightened the range dependent loci and SRC has removed the range spreading. (d) the final image estimate, the impulse response is identical for all targets. (The spectral data was windowed to extract only the data within the 3dB width of the aperture pattern, no weighting function was applied; this results in an along-track resolution of close to $D/2$. The raw data was sampled at $D/4$ in the along-track direction).

The origin of the amplitude term is interpreted as follows; the spatial phase function inside the Fourier transform operator is equivalent to a spatial chirp of chirp rate $K_a \approx -k/(\pi x)$. The Fourier transform of a chirp produces an amplitude function $\sqrt{j/K_a} \approx \sqrt{\pi x/(jk)}$. The approximation in Appendix A giving this amplitude function is valid for wide bandwidth systems.

The spatial Fourier transform of (4.6) is

$$
\begin{aligned}
EE_m(\omega, k_u) &\approx P_m(\omega) \cdot A(k_u) \cdot \int_x \int_y ff(x, y) \cdot \sqrt{\frac{\pi x}{jk}} \cdot \exp\left(-j\sqrt{4k^2 - k_u^2} \cdot x - jk_u y\right) \, dx \, dy \\
&= \sqrt{\frac{\pi}{jk}} \cdot P_m(\omega) \cdot A(k_u) \cdot \int_x \int_y ff_x(x, y) \cdot \exp\left(-j\sqrt{4k^2 - k_u^2} \cdot x - jk_u y\right) \, dx \, dy,
\end{aligned}
\tag{4.34}
$$

where the range dependent amplitude modulation is absorbed into the modified reflectivity function

$$
ff_x(x, y) = \sqrt{x} \cdot ff(x, y).
\tag{4.35}
$$

The spectrum of this modified reflectivity function is $FF_x(k_x, k_y)$.

The terms in the integral of (4.34) represent a two-dimensional spatial Fourier transform in the wavenumbers

$$
\begin{aligned}
k_x(\omega, k_u) &\equiv \sqrt{4k^2 - k_u^2} \\
k_y(\omega, k_u) &\equiv k_u
\end{aligned}
\tag{4.36}
$$

with the Jacobian $JJ(k_x, k_y) = ck_x/\left(2\sqrt{k_x^2 + k_y^2}\right) \approx c/2$. This non-linear mapping is induced *by the system* during the measurement process. This mapping is referred to as a *Stolt mapping* [151] and is represented by an operator $\mathcal{S}\{\cdot\}$ which gives (4.34) as

$$
EE_m(\omega, k_u) \approx \sqrt{\frac{\pi}{jk}} \cdot P_m(\omega) \cdot A(k_u) \cdot \mathcal{S}\{FF_x(k_x, k_y)\},
\tag{4.37}
$$

where the forward Stolt mapping, $\mathcal{S}\{\cdot\}$, given by

$$
\begin{aligned}
\omega(k_x, k_y) &\equiv \frac{c}{2}\sqrt{k_x^2 + k_y^2} \\
k_u(k_x, k_y) &\equiv k_y,
\end{aligned}
\tag{4.38}
$$

maps the wavenumbers k_x and k_y into the measurement parameters ω and k_u. The Jacobian $JJ(\omega, k_u) = 4k/\left(c\sqrt{4k^2 - k_u^2}\right) \approx 2/c$ reflects the scaling from a wavenumber ordinate to radian frequency. The inverse Stolt mapping necessary to map the measurement parameters back to wavenumbers is given in (4.36).

The offset Fourier data provided by the wavenumber algorithm is neatly summarised by

$$
\begin{aligned}
\widehat{FF}_m(k_x, k_y) &= WW(k_x, k_y) \cdot \mathcal{S}^{-1} \left\{ \sqrt{\frac{k}{k_0}} \cdot P_m^*(\omega) \cdot EE_m(\omega, k_u) \right\} \\
&= WW(k_x, k_y) \cdot \mathcal{S}^{-1} \left\{ \sqrt{\frac{k}{k_0}} \cdot SS_m(\omega, k_u) \right\},
\end{aligned}
\tag{4.39}
$$

where the frequency variation of the range envelope is deconvolved and replaced with its carrier value and, after the pulse compressed data has been inverse Stolt mapped, a window function is applied to extract the range and along-track processing bandwidths and to weight the spectrum to reduce sidelobes in the final image estimate.

At this point it is necessary to recap the discussion on the continuous Fourier transform versus the FFT given in Section 2.3. Due to the dislocation of the x-origin from the center of the swath being imaged, is not possible to *efficiently* implement the inversion scheme presented in (4.39). The correct data format for a digital implementation via the FFT is equivalent to redefining the temporal origin to be centered on range r_0, i.e., $t' = t - 2r_0/c$, and the data obtained from the inverse transform of the spectra produced using the FFT will have an output range origin defined by $x' = x - r_0$. The system model in (4.4) then becomes

$$
\begin{aligned}
ee_m(t', u) \approx \int_{x'} \int_y ff(x', y) \cdot a &\left[\left(t' + \frac{2r_0}{c} \right), (x' + r_0), y - u \right] \\
\odot_t \, p_m &\left[\left(t' + \frac{2r_0}{c} \right) - \frac{2}{c} \sqrt{(x' + r_0)^2 + (y - u)^2} \right] dx' \, dy.
\end{aligned}
\tag{4.40}
$$

and the spectrum of the system model in (4.34) becomes

$$
\begin{aligned}
EE'_m(\omega, k_u) \approx \sqrt{\frac{\pi}{jk}} \cdot P_m(\omega) \cdot A(k_u) &\cdot \exp(j2kr_0) \\
&\cdot \int_{x'} \int_y ff_x(x', y) \cdot \exp\left[-j\sqrt{4k^2 - k_u^2} \cdot (x' + r_0) - jk_u y \right] dx' \, dy \\
&= \sqrt{\frac{\pi}{jk}} \cdot P_m(\omega) \cdot A(k_u) \cdot \exp\left[-j\left(\sqrt{4k^2 - k_u^2} - 2k \right) \cdot r_0 \right] \cdot \mathcal{S}\left\{ FF'_x(k_x, k_y) \right\}
\end{aligned}
\tag{4.41}
$$

where the primed functions, eg., $EE'_m(\omega, k_u)$ and $FF'_x(k_x, k_y)$, represent spectra calculated via the FFT, i.e., phases in these spectra are referenced to $x = r_0$ from the platform location, not the origin $x = 0$.

The wavenumber inversion *realisable* via a digital processor is then

$$\widehat{FF'}_m(k_x, k_y) = WW(k_x, k_y) \cdot \mathcal{S}^{-1} \left\{ \sqrt{\frac{k}{k_0}} \cdot \exp\left[j\left(\sqrt{4k^2 - k_u^2} - 2k \right) \cdot r_0 \right] \cdot P_m^*(\omega) \cdot EE'_m(\omega, k_u) \right\}$$

$$= WW(k_x, k_y) \cdot \mathcal{S}^{-1} \left\{ \sqrt{\frac{k}{k_0}} \cdot \exp\left[j\left(\sqrt{4k^2 - k_u^2} - 2k \right) \cdot r_0 \right] \cdot SS'_m(\omega, k_u) \right\},$$

(4.42)

where the primed functions on the RHS of (4.42) are calculated with respect to the redefined x-axis via the FFT, thus the spectral estimate $\widehat{FF'}_m(k_x, k_y)$ has phases defined relative to r_0. The redefinition of the temporal t-axis and spatial x-axis conveniently removes highly oscillating phase functions prior to the inverse Stolt mapping. It is important that these phase functions are removed *inside* the inverse Stolt operator *before* the (ω, k_u) data are interpolated onto the (k_x, k_y) domain, otherwise this interpolation step has poor performance.

The phase function inside the inverse Stolt operator in (4.42) is the conjugate of the phase of the *exact* 2-D FFT of the system PSF for a target at $x = r_0$, i.e., it is the phase of the system transfer function for a target at $x = r_0$. Thus, the phase factor inside the inverse Stolt operator in (4.42) acts to *compress* all targets about range r_0, and the inverse Stolt mapping removes the distortion in targets at $x \neq r_0$.

Equation (4.42) is still not the most efficient way to implement the wavenumber algorithm digitally. Since the data in (4.42) is in modulated format, the valid input data exists for samples approximately within $\omega \in [\omega_0 - \pi B_c, \omega_0 + \pi B_c]$ and the valid output wavenumbers exist for samples within $k_x \in [2k_0 - 2\pi B_c/c, 2k_0 + 2\pi B_c/c]$. Data outside of these regions is overhead to the inversion scheme. The raw modulated data should be converted to complex valued baseband format as soon as possible (see the methods recommended in Sections 2.1.1 to 2.1.3). The processing of the baseband data is the most efficient method, as it represents the smallest data set for both the FFT and the Stolt interpolator. It is also recommended that the spectral data stored during the conversion from modulated format to baseband is padded with zeros to the next power of 2 to take advantage of fast radix-2 FFT algorithms.

The conversion of the modulated input data to baseband and the process of mapping (interpolating) onto a baseband grid modifies the definition of the Stolt mapping operator to give the baseband inverse Stolt map, $\mathcal{S}_b^{-1}\{\cdot\}$, where

$$k_x(\omega, k_u) \equiv \sqrt{4k^2 - k_u^2} - 2k_0$$
$$k_y(\omega, k_u) \equiv k_u.$$

(4.43)
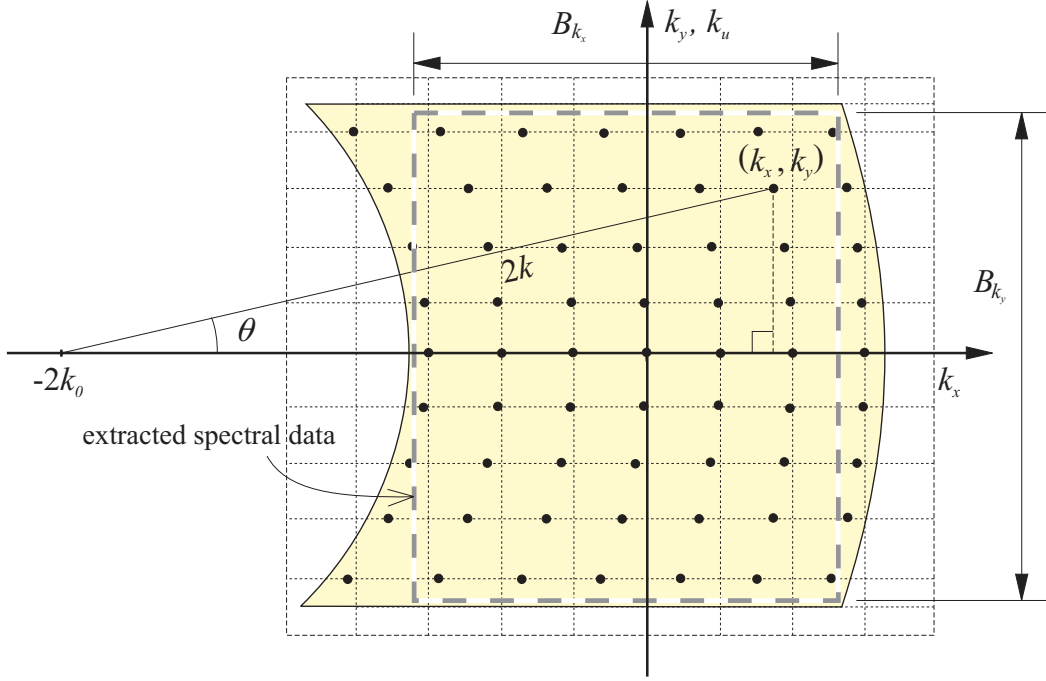
Figure 4.4　The two-dimensional collection surface of the wavenumber data. The heavy dots indicate the locations of the raw data samples along radii $2k$ at height k_u. The underlying rectangular grid shows the format of the samples after mapping (interpolating) to a Cartesian grid on (k_x, k_y). The spatial bandwiths B_{k_x} and B_{k_y} outline the rectangular section of the wavenumber data that is extracted, windowed, and inverse Fourier transformed to produce the image estimate. The slightly offset nature of the mapped baseband data appears as an irrelevant phase factor in the spatial domain. This phase factor does not effect the final image estimate $|\widehat{ff}_b(x,y)|$.

This definition modifies the digital inversion in (4.42) to give the most *efficient* wavenumber inversion

$$\widehat{FF}'_b(k_x, k_y) = WW(k_x, k_y) \cdot \mathcal{S}_b^{-1} \left\{ \sqrt{\frac{k}{k_0}} \cdot \exp\left[j \left(\sqrt{4k^2 - k_u^2} - 2k \right) \cdot r_0 \right] \cdot P_b^*(\omega_b) \cdot EE'_b(\omega_b, k_u) \right\}.$$

$$(4.44)$$

The inverse Stolt mapping of the measured (ω_b, k_u)-domain data onto the (k_x, k_y)-domain is shown in Fig. 4.4. The sampled raw data is seen to lie along radii of length $2k$ in the (k_x, k_y)-wavenumber space. The radial extent of this data is controlled by the bandwidth of the transmitted pulse and the along-track extent is controlled by the overall radiation pattern of the real apertures. The inverse Stolt mapping takes these raw radial samples and re-maps them onto a uniform *baseband* grid in (k_x, k_y) appropriate for inverse Fourier transformation via the inverse FFT. This mapping operation is carried out using an interpolation process such as the one described in [82, pp133-156], other interpolators used for Stolt mapping are analyzed in [95].

Figure 4.4 shows how the Stolt mapping has distorted the spatial frequencies onto a curved shape.

If a pulse with a constant spectrum was transmitted, eg. an unweighted LFM pulse, then the height of the 2-D spectrum is approximately constant with k_x (the frequency variant part of the amplitude term in (4.37) has been deconvolved), and if the real aperture illumination functions are unweighted, the 2-D spectrum has a sinc-like shape in the k_u direction. The windowing operation of $WW(k_x, k_y)$ can be split into two parts; data extraction and data weighting. The windowing operation first *extracts* from the curved spectral data a rectangular area of the wavenumber data. Although other shapes could be extracted, the inverse 2-D Fourier transform of a rectangular spectral area concentrates most of the sidelobe energy along the range and along-track directions. If the same scene is re-imaged using say, a squinted system, and the same rectangular extraction operation is performed, then the two images have the same point spread function. This is extremely important for image comparison, or when multiple images (generated from different passes over a scene) are to be used in interferometric applications. If a rectangular area is not extracted, targets in the scene have a *space variant* point spread function which complicates image interpretation. The choice of the 2-D weighting function to be applied to the extracted data is arbitrary, this thesis extracts a rectangular area and weights it with a 2-D Hamming weighting (1.30 width expansion, -43dB sidelobes, and a coherent gain of 0.54 for each dimension) [71]. Before the k_y weighting is applied across the processed 3dB radiation bandwidth, the amplitude effect of the radiation pattern is deconvolved, i.e.,

$$WW(k_x, k_y) = \operatorname{rect}\left(\frac{k_y}{B_{k_y}}\right) \cdot \frac{1}{A(k_y)} \cdot W_h\left(\frac{k_x}{B_{k_x}}\right) \cdot W_h\left(\frac{k_y}{B_{k_y}}\right), \tag{4.45}$$

where $W_h(\alpha)$ is a 1-D Hamming window defined over $\alpha \in [-1/2, 1/2]$ [71] and the wavenumber bandwidths of the extracted data shown in Fig. 4.4 are

$$
\begin{aligned}
B_{k_x} &= \sqrt{(2k_{\max})^2 - \left(\frac{2\pi}{D}\right)^2} - 2k_{\min} \approx \frac{4\pi B_c}{c} - \frac{2\pi^2}{k_{\max}D^2} \\
B_{k_y} &= \frac{4\pi}{D},
\end{aligned}
\tag{4.46}
$$

where $k_{\min}$ and $k_{\max}$ are the minimum and maximum wavenumbers in the transmitted pulse, B_c is the pulse bandwidth (in Hertz) and D is the effective aperture length. The definition of the x and y axes depends on the number of samples obtained within each wavenumber bandwidth, the sample spacing chosen, and the amount of zero padding used. The resolution in the final image estimate is

$$
\begin{aligned}
\delta x_{3\mathrm{dB}} &= \frac{2\pi}{B_{k_x}} = \alpha_w \cdot \frac{c}{2B_{\mathrm{eff}}} \qquad \text{where} \qquad B_{\mathrm{eff}} \approx B_c - \frac{\pi c}{2k_{\max}D^2} \\
\delta y_{3\mathrm{dB}} &= \frac{2\pi}{B_{k_y}} = \alpha_w \cdot \frac{D}{2},
\end{aligned}
\tag{4.47}
$$

where $\alpha_w = 1.30$ for the Hamming window. If the effect of the radiation pattern is not deconvolved during the windowing operation, the constants representing the width expansion and coherent gain need to be calculated for the overall windowing effect of the radiation pattern and the Hamming window. The strip-map wavenumber processor sampling requirements are identical to those of the spatial-temporal and range-Doppler processors; the range sampling requirements are satisfied in the receiver, while the along-track sampling wavenumber is ideally $k_{us} = 8\pi/D$, i.e. a sample spacing of $\Delta_u = D/4$.

To determine what the wavenumber estimate represents, the baseband version of (4.41) is substituted into (4.44) to give

$$\widehat{FF}'_b(k_x, k_y) = \sqrt{\frac{\pi}{jk_0}} \cdot \text{rect}\left(\frac{k_x}{B_{k_x}}\right) \cdot \text{rect}\left(\frac{k_y}{B_{k_y}}\right) \cdot FF'_x(k_x, k_y), \tag{4.48}$$

where this time the window function $WW(k_x, k_y)$ only rect limits the spatial bandwidths and deconvolves the effect of the radiation pattern, and the transmitted pulse has a uniform spectrum. The image estimate obtained from this wavenumber data is

$$\begin{aligned}
\left|\widehat{ff}_b(x', y)\right| &= \left|\mathcal{F}^{-1}_{k_x, k_y}\left\{\widehat{FF}'_b(k_x, k_y)\right\}\right| \\
&= \left|\sqrt{\frac{\pi}{jk_0}} \cdot \frac{B_{k_x}}{2\pi} \cdot \frac{B_{k_y}}{2\pi} \cdot \text{sinc}\left(\frac{B_{k_x}}{2\pi} \cdot x\right) \cdot \text{sinc}\left(\frac{B_{k_y}}{2\pi} \cdot y\right) \odot_{x,y} ff_x(x', y)\right|.
\end{aligned} \tag{4.49}$$

Redefining the x'-axis back to x and using (4.35) gives,

$$\left|\widehat{ff}_b(x, y)\right| = \left|\sqrt{\frac{\pi x}{jk_0}} \cdot \frac{2B_{\text{eff}}}{c} \cdot \frac{2}{D} \cdot \text{sinc}\left(\frac{2B_{\text{eff}}}{c} \cdot x\right) \cdot \text{sinc}\left(\frac{2}{D} \cdot y\right) \odot_{x,y} ff(x, y)\right| \tag{4.50}$$

When displaying the image estimates, the choice of the x-ordinate is arbitrary. In this thesis, the image estimates are sometimes plotted versus the x-axis and other times against the x'-axis (with the figure caption or text stating r_0).

The systematic development of the inversion scheme has allowed the correct scaling parameters to be derived. The *synthetic aperture improvement factor* is seen to be

$$\text{IF}_{\text{sa}} = 20\log_{10}\left(\sqrt{\frac{\pi x}{k_0}} \cdot \frac{2}{D}\right) = 10\log_{10}\left(\frac{4\pi x}{k_0 D^2}\right). \tag{4.51}$$

This factor can be interpreted as the *space-bandwidth product* of the along-track chirp as

$$K_a L_{\text{sa}}^2 = \frac{k}{\pi x} \cdot \left(\frac{x\lambda}{D}\right)^2 = \frac{4\pi x}{kD^2}, \tag{4.52}$$

or in some references it is misleadingly referred to as the improvement factor due to the integration of

N along-track samples, where

$$N = \frac{L_{\text{sa}}}{\Delta_u} = \frac{4\pi x}{kD^2} \tag{4.53}$$

and the derivation makes the (incorrect) assumption that the along-track samples in the raw data are spaced $\Delta_u = D/2$ apart. The interpretations in (4.52) and (4.53) are both frequency and spatially dependent, by deconvolving the frequency dependence in (4.37), the resulting improvement factor in (4.51) is only spatially dependent.

Section 7.2 and Fig. 7.7 detail the application of the wavenumber algorithm to a simulated target field.

Despite the fact that the Stolt mapping represents an exact inversion of the system model, the wavenumber algorithm still requires an interpolator for a digital implementation. This interpolator operates in the *wavenumber* domain, so it has to be of high accuracy so that minor phase errors in the wavenumber domain, due to the interpolation, do not degrade the final spatial domain image estimate. The requirements of this interpolator reduces the wavenumber algorithm efficiency to about the same level as the range-Doppler algorithm. The replacement of this interpolation step through the use of equivalent phase multiplies in other domains is the essence of the chirp scaling algorithm.

4.3.4   The chirp scaling algorithm

A new wavenumber algorithm that did not require an interpolation step was developed independently by two groups in 1992 [30, 126, 127, 134]. The publication produced by these two groups in 1994 refers to this new inversion scheme as *chirp scaling* [128]. The name refers to the algorithms exploitation of the scaling/shifting properties of LFM or 'chirp' pulses. The chirp scaling algorithm is designed around *curvature equalization*, so that by the time the echo signals are transformed to the two-dimensional wavenumber domain, all of the range migration trajectories have been adjusted to have congruent loci, equivalent to the trajectory at a selected reference range. As all of the resulting range migration trajectories are single valued in the wavenumber domain, then range migration correction may be completed by a phase multiplication that is known, and single valued at each and every point. There is no longer a need for a Stolt mapping to focus targets away from the reference range and the process is inherently phase preserving [128, 134]. Essentially, the chirp scaling algorithm takes advantage of the best features of the range-Doppler and wavenumber algorithms, and removes their deficiencies [30].

Rather than start with the range compressed data as most of the previous inversion schemes have done, the chirp scaling algorithm starts with the quadrature demodulated raw echo data. *Direct* calculation of the spatial Fourier transform of the system model shown in (4.5) to yield the range-Doppler signal of the system model *cannot* be done when the range curvature term is included in the envelope

of the transmitted pulse (the range-Doppler algorithm used a quadratic approximation). By assuming that a large bandwidth LFM pulse is transmitted, Raney [126,127] showed that it *was possible* to obtain the required range-Doppler signal *without* using a Stolt mapping, thus removing the interpolator from the inversion scheme. This range-Doppler signal is obtained by first performing a 2-D Fourier transform on the system model to give the wavenumber data, where the phase of the LFM pulse adds an extra term quadratic in ω. This extra phase factor allows a minor approximation to be made which avoids the Stolt mapping, and allows the inverse temporal Fourier transform to be obtained, yielding the required range-Doppler signal. The range-Doppler signal produced in this manner is applicable to wide beam width, wide swath-width and moderately squinted systems, unlike the quadratic approximation used in the range-Doppler algorithm [30,126,127,134]. The range-Doppler signal of the PSF still containing the LFM transmitted pulse is then (Appendix A)

$$
\begin{aligned}
pP(t, k_u; x) = {} & \sqrt{\frac{\pi x}{j k_0}} \cdot A(k_u) \cdot \operatorname{rect}\left(\frac{t - \frac{2}{c}\Delta R_s(k_u; x)}{\tau_c}\right) \\
& \cdot \exp\left\{ j\pi K_s(k_u; x) \cdot \left[t - \frac{2}{c}\Delta R_s(k_u; x) \right]^2 \right\} \\
& \cdot \exp\left\{ -j\left(\sqrt{4k_0^2 - k_u^2} - 2k_0 \right) \cdot x \right\},
\end{aligned}
\tag{4.54}
$$

where the range migration locus in range-Doppler space and the along-track phase functions are recognizable from the discussion of the range-Doppler processor [see (4.22), (4.23) and (4.25)]. The range-Doppler signal (4.54) is also seen to contain a series of delayed, range and Doppler wavenumber dependent, chirps with a modified or *scaled chirp rate* given by

$$
K_s(k_u; x) = \frac{1}{1/K_c - K_{src}(k_u; x)},
\tag{4.55}
$$

where the additional range dependent chirp term,

$$
K_{src}(k_u; x) = \frac{8\pi x}{c^2} \cdot \frac{k_u^2}{(4k_0^2 - k_u^2)^{3/2}},
\tag{4.56}
$$

is compensated during processing by *secondary range compression* (SRC). This range distortion is *not* a function of the transmitted chirp rate K_c: rather, it is a function of the geometry. Range distortion is a direct consequence of the lack of orthogonality between 'range' and 'along-track' for signal components away from zero Doppler wavenumber, and it applies to any form of transmitted pulse, not just to LFM [128].

The first step in the chirp scaling algorithm is to spatial Fourier transform the raw data into the range-Doppler domain. The range-Doppler data is then perturbed by a phase multiply that causes all

the *phase centers* of the reflected chirps to have a range migration locus that is *identical* to the reference range, r_0. Mathematically this operation is

$$mM_b(t, k_u) = eE_b(t, k_u) \cdot \phi\Phi_1(t, k_u), \tag{4.57}$$

where the *chirp scaling multiplier* is

$$\phi\Phi_1(t, k_u) = \exp\left\{j\pi K_s(k_u; r_0) \cdot C_s(k_u) \cdot [t - t_0(k_u)]^2\right\} \tag{4.58}$$

and the time locus of the reference range in the range-Doppler domain is

$$t_0(k_u) = \frac{2}{c} \cdot r_0 \cdot [1 + C_s(k_u)]. \tag{4.59}$$

The application of the *range variant* chirp scaling multiplier is easily applied in the range-Doppler domain where range *is* a parameter. The wavenumber algorithm removes the range variant nature of the system equations in the 2-D wavenumber domain where range *is not* available as a parameter. The chirp scaled data is then temporal Fourier transformed to the 2-D wavenumber domain. The range migration loci of all targets now follow the same reference curvature; however, the chirp scaling operation induces a further Doppler dependence in the range chirp. Pulse compression, including secondary range compression, bulk range migration correction, deconvolution of the frequency dependent amplitude term, and windowing is performed by

$$NN_b(k_x, k_y) = WW(k_x, k_y) \cdot \mathcal{C}^{-1}\left\{MM_b(\omega_b, k_u) \cdot \Theta\Theta_2(\omega_b, k_u)\right\}, \tag{4.60}$$

where

$$\Theta\Theta_2(\omega_b, k_u) = \sqrt{\frac{k}{k_0}} \cdot \exp\left\{j\frac{\omega_b^2}{4\pi K_s(k_u; r_0)[1 + C_s(k_u)]}\right\} \cdot \exp\left\{j2k_b r_0 C_s(k_u)\right\}, \tag{4.61}$$

the transform $\mathcal{C}^{-1}\{\cdot\}$ given by

$$k_x(\omega_b, k_u) \equiv \frac{2\omega_b}{c}$$
$$k_y(\omega_b, k_u) \equiv k_u \tag{4.62}$$

has the Jacobian $c/2$, and the window function is identical to that used by the wavenumber algorithm, (4.45). The point at which the measurement parameters, (t, u) or (ω_b, k_u), are scaled to the image parameters, (x, y) or (k_x, k_y), is entirely arbitrary, and is done at this stage to allow the use of the window function, (4.45).

The range-Doppler estimate of the image is obtained via

$$\widehat{fF}_b(x, k_y) = nN_b(x, k_y) \cdot \psi\Psi_3(x, k_y), \tag{4.63}$$

where along-track compression and residual phase compensation is performed by

$$\psi\Psi_3(x, k_y) = \exp\left\{ j \left(\sqrt{4k_0^2 - k_y^2} - 2k_0 \right) \cdot x \right\} \cdot \exp\left\{ -j\frac{4\pi}{c^2} K_s(k_y; r_0) C_s(k_y) \left[1 + C_s(k_y) \right] (x - r_0)^2 \right\}. \tag{4.64}$$

Inverse spatial Fourier transformation of (4.63) then leads to the final complex image estimate [126]. The complex image estimate obtained via the chirp scaling algorithm is neatly summarised as

$$\widehat{ff}_b(x, y) = \mathcal{F}_{k_y}^{-1} \left\{ \mathcal{F}_{k_x}^{-1} \left\{ WW(k_x, k_y) \cdot \mathcal{C}^{-1} \left\{ \mathcal{F}_t \left\{ eE_b(t, k_u) \cdot \phi\Phi_1(t, k_u) \right\} \cdot \Theta\Theta_2(\omega_b, k_u) \right\} \right\} \cdot \psi\Psi_3(x, k_y) \right\}. \tag{4.65}$$

Appendix A details the derivation of the range-Doppler signal and the phase functions of the chirp scaling algorithm.

If a waveform other than LFM has been employed for the transmitted pulse, the raw data first needs to be range compressed with the appropriate deconvolving function, then respread by convolution with a LFM pulse. In cases where highly squinted real apertures are employed, a non-linear FM chirp scaling is required instead of the original linear chirp scaling phase perturbation [30,36,128]. In airborne systems further modifications can be made to incorporate compensation for motion errors and Doppler centroid variations [110].

At no time throughout the chirp scaling algorithm is it necessary to use an interpolator. All operations are achieved with phase multiplies and Fourier transforms, both relatively efficient processes. Thus, the chirp scaling algorithm is the *most efficient* generalized processor to date produced from the SAR community. An interesting additional advantage of the chirp scaling algorithm is that it can be run backwards [126] over a perfectly focused scene to produce the raw data necessary for testing any of the algorithms in this thesis. If a 'perfect' interpolator was available, then the same could be said for the other wavenumber algorithms, but this is seldom true. This defocusing technique is useful when a test field containing a large number of reflectors needs to be generated, or alternatively, it can take perfectly focused *actual* imagery and defocus it for the other algorithms to process.

4.3.5   Accelerated chirp scaling

The efficiency of the chirp scaling algorithm can be improved *dramatically* with the inclusion of a *preliminary* range compression step. This preliminary pulse compression operation acts to reduce the

temporal extent of the transmitted chirp to the order of the maximum range migration expected in the scene. The full temporal support of the transmitted range chirp is *not* required for the implementation of chirp scaling. The chirp length required by chirp scaling multiplier only needs to be long enough to support the shift of the chirp's phase center to that of the reference range. Generally the scene center is chosen as the reference range so that the shift of the phase center is never particularly large. Bulk range curvature is performed during a later phase multiply.

In sonar applications, the chirp length of the transmitted LFM waveform is often a significant fraction of, or indeed in CTFM, equivalent to, the repetition period. These long chirps represent a significant and unnecessary overhead. Consider for example the processing of Fig. 7.11 in the results chapter. The transmitted pulse consisted of a 50ms LFM pulse chirped from 40kHz to 20kHz. The extracted area of the image covers about 5m in range or about 200 samples, but the LFM pulse exists for another 37.5m, which is 1500 more samples. To operate a chirp scaling algorithm over this data set requires operations on a matrix containing 1700 elements in range, but the final focused image contains only 200 valid samples! The range migration expected for the target shown in Fig. 7.11 is less than 40cm, or about 16 samples. The chirp rate of the pulse can be *increased* so that its temporal extent is of this order, say 1m. The matrix required for processing via the chirp scaling algorithm is then 240 samples in range and it contains 200 valid samples after focusing. This reduction in matrix size results in a dramatic improvement in processing efficiency.

The preliminary pulse compression step proceeds as follows; after the data has been packed into the 2-D format required to represent $ee_b(t, u)$, the data is temporally Fourier transformed to give $Ee_b(\omega_b, u)$. This data is then pulse compressed using the phase function $\exp[j\omega_b^2/(4\pi K_c)]$ and respread (i.e., re-chirped) with $\exp[-j\omega_b^2/(4\pi K_n)]$, where the new (higher) chirp rate $K_n = B_c/\tau_n = K_c\tau_c/\tau_n$ and where τ_n is the required reduced temporal extent of the order of the range migration. Once these phase multiplies have been performed, the ω_b domain is decimated to reduce the scene size in the temporal domain. The equivalent operation would be to return the data to the temporal domain and to discard the excess data points; however, that involves the inverse Fourier transformation of a large matrix— decimation followed by the inverse Fourier transform is the more efficient and equivalent operation. The new chirp rate K_n is used in all subsequent chirp scaling processing of the data.

Another equally useful application of this technique is the re-insertion of a chirp to data that has been stored in pulse compressed format. Re-insertion of the chirp corresponds to a convolution in range, so the preliminary step must this time, increase the size of the pulse compressed matrix $ss_b(t, u)$, and then generate the chirped matrix $ee_b(t, u)$. Increasing the matrix size avoids the corrupting effects of cyclic convolution. The re-chirped data is then passed through the chirp scaling processor.

4.4   SPOTLIGHT SYNTHETIC APERTURE SYSTEM MODEL

Figure 4.5 shows the 2-D geometry appropriate for broadside spotlight mode synthetic aperture imaging systems. The target area described by $ff(x, y)$ is considered to consist of a continuous 2-D distribution of omni-directional (aspect independent) reflecting targets. This target area is much smaller than that imaged by a strip-map system and it is *continuously* illuminated by steering or slewing the antenna of a side-looking radar or sonar system. The platform travels along a straight locus u, with a velocity v_p, flying parallel to the y axis of the target area. The origin of the x-axis is the same as in the strip-map algorithms, however, the system model is developed around the offset x'-axis, where $x' = x - r_0$. The distance r_0 is known as the *standoff* distance and it corresponds to the closest approach of the platform to the scene center. In the development of the tomographic formulation, the origin of the time axis is *not* the same as in the strip-map algorithms; the spotlight temporal origin t_s tracks the center of the spotlight scene, i.e., $t_s = 0$ varies with along-track location and corresponds to $t = 2/c \cdot \sqrt{r_0^2 + u^2}$ or $t' = 2/c \cdot \left(\sqrt{r_0^2 + u^2} - r_0 \right)$ (the derivation of a spotlight model with the original t and t' axes is also made for comparative purposes). The antenna steering causes the beam pattern of the real aperture to limit the spatial information imaged by the spotlight system to a region of radius X_0 centered on the scene origin. The transmitter and receiver systems operate in a similar manner to strip-map mode, but due to the small size of the target area certain approximations can be made that lead to an extremely efficient inversion scheme. This model and its inversion scheme is known as the *tomographic* formulation due to its striking similarity to computerized tomography (CT) [82, 114].

The following sections develop aspects of the spotlight mode; the plane wave assumption that relates the tomographic model to the strip-map model, and the generalized method for processing spotlight data collected with arbitrary system parameters.

4.4.1   System model for tomographic spotlight synthetic aperture imaging

The interpretation of spotlight mode as a tomographic inversion was first presented for a two-dimensional imaging geometry by Munson in 1983 [114], this was then extended to three-dimensions by Jakowatz *et al* in 1995 [82, 83]. A review of computerized tomography for comparison to spotlight inversion is given in Chapter 2 of Jakowatz *et al* [82].

Figure 4.5 shows the concept of the tomographic spotlight geometry. If a plane-wave assumption is made, i.e., the wavefronts across the spotlight scene are assumed to have such a small amount of curvature that they are effectively straight lines, then the reflectivity function measured at point u along the aperture is

$$f_\theta(x_\theta) = \int_{y_\theta} ff_\theta(x_\theta, y_\theta) dy_\theta, \tag{4.66}$$
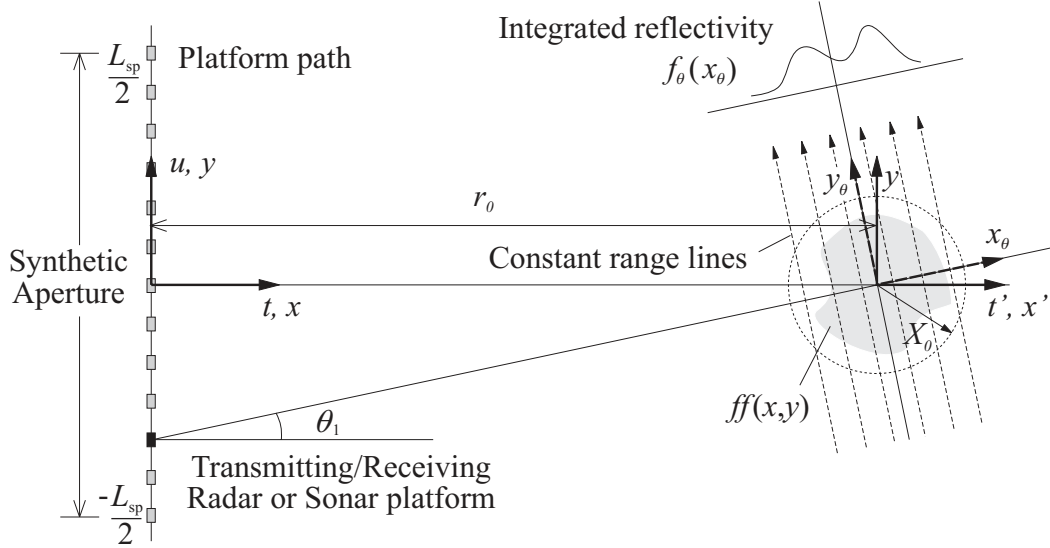
Figure 4.5 Imaging geometry appropriate for the tomographic formulation of a spotlight system. Constant range lines exist if wavefront curvature can be ignored. Note the natural coordinate system (x, y), the translated coordinate system necessary for FFT processing (x', y), and the rotated coordinate system (x_θ, y_θ) (see the text for more details).

where $ff_\theta(x_\theta, y_\theta)$ is the reflectivity function in the rotated domain (x_θ, y_θ), which is the (x', y) domain rotated through angle θ_1. If the modulated transmitted signal $p_m(t)$ is a LFM chirp of rate K_c then the signal received along the aperture is

$$ee_m(t_s, u) = \mathcal{R}^{-1} \left\{ \int_{x_\theta} f_\theta(x_\theta) \cdot p_m \left(t_s - \frac{2}{c} x_\theta \right) dx_\theta \right\}, \tag{4.67}$$

where t_s is the range gated time which has its origin at the *spotlight scene center* and $\mathcal{R}^{-1}\{\cdot\}$ is an inverse rotation operator that is defined shortly. Because the scene extent in the range direction is small compared to the standoff range, pulse compression using deramp processing in the receiver is more efficient than matched filtering (see Section 2.4.3). The output of the demodulator at point u in the aperture after deramp processing is the baseband pulse compressed signal

$$\begin{aligned} Ss'_b(\omega_b, u) &= \mathcal{D} \left\{ p_m^*(t_s) \cdot ee_m(t_s, u) \right\} \\ &= \mathcal{D} \left\{ \mathcal{R}^{-1} \left\{ \int_{x_\theta} f_\theta(x_\theta) \cdot \exp \left[-j\frac{2}{c}(\omega_0 + 2\pi K_c t_s) x_\theta \right] dx_\theta \right\} \right\} \\ &= \frac{1}{2\pi K_c} \cdot \mathcal{R}^{-1} \left\{ \int_{x_\theta} \int_{y_\theta} ff_\theta(x_\theta, y_\theta) \exp(-j2kx_\theta) dx_\theta dy_\theta \right\}, \end{aligned} \tag{4.68}$$

where the Deramp operator $\mathcal{D}\{\cdot\}$ has mapped the range gated time t_s to the radian frequencies within

the transmitted pulse via

$$\omega_b(t_s, u) \equiv 2\pi K_c t_s, \tag{4.69}$$

which has the Jacobian, $1/(2\pi K_c)$. The inverse rotation mapping operator $\mathcal{R}^{-1}\{\cdot\}$ performs the coordinate transformation from (x_θ, y_θ) back to (x', y) as defined by

$$\begin{bmatrix} x' \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{bmatrix} \begin{bmatrix} x_\theta \\ y_\theta \end{bmatrix}, \tag{4.70}$$

where the angle from the current platform location u to the scene center is $\theta_1 = \tan^{-1}(-u/r_0)$, the Jacobian of the rotation is unity, and $x' = x - r_0$ is the modified range ordinate defined from the scene center as required for a digital implementation via the FFT.

Finally, the deramped data collected by the spotlight system is given by the *tomographic spotlight system model*:

$$Ss_b'(\omega_b, u) = \frac{1}{2\pi K_c} \cdot \int_{x'} \int_y ff(x', y) \exp(-j2k\cos\theta_1 x' - j2k\sin\theta_1 y) dx' dy. \tag{4.71}$$

4.4.2   Relating the tomographic model to the strip-map model

If the effect of the radiation pattern is ignored, a temporal Fourier transform of the system model in the modified coordinate system given in (4.40) gives

$$Ee_m'(\omega, u) = P_m(\omega) \cdot \exp(j2kr_0) \cdot \int_{x'} \int_y ff(x', y) \cdot \exp\left[-j2k\sqrt{(x' + r_0)^2 + (y - u)^2}\right] dx' dy. \tag{4.72}$$

Making the variable transformation from (x', y) to the rotated domain (x_θ, y_θ) gives

$$Ee_m'(\omega, u) = P_m(\omega) \cdot \exp(j2kr_0)$$
$$\cdot \mathcal{R}^{-1}\left\{ \int_{x_\theta} \int_{y_\theta} ff_\theta(x_\theta, y_\theta) \cdot \exp\left(-j2k\sqrt{\left(\sqrt{r_0^2 + u^2} + x_\theta\right)^2 + y_\theta{}^2}\right) dx_\theta dy_\theta \right\}. \tag{4.73}$$

The following approximation is valid as we are assuming minimum wavefront curvature, i.e., all targets along the line y_θ are effectively the same range from the measurement platform:

$$\sqrt{(\sqrt{r_0^2 + u^2} + x_\theta)^2 + y_\theta{}^2} \approx \sqrt{r_0^2 + u^2} + x_\theta. \tag{4.74}$$

Substituting this into (4.73) gives the *plane wave spotlight system model*:

$$
\begin{aligned}
Ee'_m(\omega, u) &\approx P_m(\omega) \cdot \exp(j2kr_0) \cdot \mathcal{R}^{-1} \left\{ \int_{x_\theta} \int_{y_\theta} ff_\theta(x_\theta, y_\theta) \cdot \exp \left\{ -j2k \left(\sqrt{r_0^2 + u^2} + x_\theta \right) \right\} dx_\theta dy_\theta \right\} \\
&= P_m(\omega) \cdot \exp(j2kr_0) \\
&\quad \cdot \mathcal{R}^{-1} \left\{ \exp \left(-j2k \sqrt{r_0^2 + u^2} \right) \cdot \int_{x_\theta} \int_{y_\theta} ff_\theta(x_\theta, y_\theta) \cdot \exp \left(-j2kx_\theta \right) dx_\theta dy_\theta \right\} \\
&= P_m(\omega) \cdot \exp \left[-j2k \left(\sqrt{r_0^2 + u^2} - r_0 \right) \right] \\
&\quad \cdot \int_{x'} \int_y ff(x', y) \cdot \exp \left(-j2k \cos \theta_1 x' - j2k \sin \theta_1 y \right) dx' dy,
\end{aligned}
$$

$$(4.75)$$

where we have made use of the inverse rotation mapping operator in (4.70). If the reflections are pulse compressed, then (4.75) becomes

$$
\begin{aligned}
Ss'_m(\omega, u) = |P_m(\omega)|^2 \cdot \exp \left[-j2k \left(\sqrt{r_0^2 + u^2} - r_0 \right) \right] \\
\cdot \int_{x'} \int_y ff(x', y) \cdot \exp \left(-j2k \cos \theta_1 x' - j2k \sin \theta_1 y \right) dx' dy,
\end{aligned}
$$

$$(4.76)$$

which in demodulated notation becomes

$$
\begin{aligned}
Ss'_b(\omega_b, u) = |P_b(\omega_b)|^2 \cdot \exp \left[-j2k \left(\sqrt{r_0^2 + u^2} - r_0 \right) \right] \\
\cdot \int_{x'} \int_y ff(x', y) \cdot \exp \left(-j2k \cos \theta_1 x' - j2k \sin \theta_1 y \right) dx' dy.
\end{aligned}
$$

$$(4.77)$$

The differences between plane wave and tomographic models can now be investigated. The plane wave model shown in (4.77) differs from the tomographic formulation shown in (4.71) by two important factors. The first factor $|P_b(\omega)|^2$ is the baseband pulse compressed spectrum which drops out of the tomographic formulation due to the use of deramp processing (it is replaced by a constant). The inversion scheme developed here did not make this assumption, so (4.75) or (4.77) are appropriate for use in developing an inversion scheme for any form of transmitted pulse. The phase factor in (4.75) and (4.77) drops out of the tomographic formulation due to the definition of time from the scene center, i.e., in a tomographic spotlight system, data is gated so that only echoes from around the scene center are stored. This phase factor also drops out of the plane wave model if its derivation uses the same time gating scheme as used to derive the tomographic model.

4.4.3   Tomographic spotlight inversion

The data obtained in the tomographic system model can be interpreted as a polar region of offset Fourier data. Equation (4.71) can be re-written as (ignoring the constant)

$$
\begin{aligned}
Ss'_b(\omega_b, u) &\approx \int_{x'} \int_y ff(x', y) \exp(-j2k\cos\theta_1 x' - j2k\sin\theta_1 y) dx dy \\
&= \mathcal{P}_b \left\{ FF'(k_x, k_y) \right\},
\end{aligned}
\tag{4.78}
$$

where the baseband polar transformation induced by the measurement system, $\mathcal{P}_b\{\cdot\}$, transforms the baseband data from the Cartesian (k_x, k_y)-domain to the polar (ω_b, u)-domain via

$$
\begin{aligned}
\omega_b(k_x, k_y) &= \frac{c}{2}\sqrt{k_x^2 + k_y^2} - \omega_0 \\
u(k_x, k_y) &= -\frac{k_y}{k_x} r_0.
\end{aligned}
\tag{4.79}
$$

The relevant Jacobian for the transformation,

$$
JJ(\omega, u) = -\frac{4\omega}{c} \cdot \frac{r_0}{(u^2 + r_0^2)},
\tag{4.80}
$$

reflects the distortion induced by the measurement system that transforms polar samples of the wavenumber domain onto rectangular spatial-temporal frequency coordinates.

Actually, it is more natural to consider that the data is mapped into the modulated signal and then the signal is demodulated by the receiver (or after reception). This gives the model,

$$
\begin{aligned}
Ss'_m(\omega, u) &\approx \int_{x'} \int_y ff(x', y) \exp(-j2k\cos\theta_1 x' - j2k\sin\theta_1 y) dx dy \\
&= \mathcal{P} \left\{ FF'(k_x, k_y) \right\},
\end{aligned}
\tag{4.81}
$$

where the polar transformation induced by the measurement system, $\mathcal{P}\{\cdot\}$, transforms the data from the Cartesian (k_x, k_y)-domain to the polar (ω, u)-domain via

$$
\begin{aligned}
\omega(k_x, k_y) &= \frac{c}{2}\sqrt{k_x^2 + k_y^2} \\
u(k_x, k_y) &= -\frac{k_y}{k_x} r_0.
\end{aligned}
\tag{4.82}
$$

then the receiver imposes the $-\omega_0$ shift seen in (4.79). As with the previous imaging algorithms, it is always more efficient to process the complex valued baseband data.

There are two interesting points to note about the tomographic model; the measurement system

produces *polar* samples of the 2-D Fourier transform of the imaged scene and the beam patterns of the real apertures do not appear in the system model. The range Fourier transform has been avoided by using deramp processing to produce the Fourier samples directly, and the plane wave assumption in conjunction with having the origin of time at the scene center has allowed the along-track Fourier transform to be avoided. Time gating about the scene center removes the along-track chirp component of the signal model. The beam patterns do not enter the system model as they are essentially only used to 'window' the limited target area.

From (4.78) the inversion scheme to obtain an estimate of the image Fourier transform is

$$\widehat{FF}'_b(k_x, k_y) = WW(k_x, k_y) \cdot \mathcal{P}_b^{-1}\left\{Ss'_b(\omega_b, u)\right\}, \tag{4.83}$$

where $\mathcal{P}_b^{-1}\{\cdot\}$ represents the inverse polar transform from the polar version of the (ω_b, u) data on to a baseband rectangular (k_x, k_y) grid as defined by

$$
\begin{aligned}
k_x(\omega_b, u) &= 2k\cos\theta_1 - 2k_0 = \frac{2kr_0}{\sqrt{u^2 + r_0^2}} - 2k_0 \\
k_y(\omega_b, u) &= 2k\sin\theta_1 = -\frac{2ku}{\sqrt{u^2 + r_0^2}},
\end{aligned}
\tag{4.84}
$$

which has the Jacobian

$$JJ(k_x, k_y) = -\frac{cr_0}{2} \cdot \frac{\sqrt{k_x^2 + k_y^2}}{k_x^2}. \tag{4.85}$$

This reformatted spectral data is then appropriate for windowing, padding, and finally inverse Fourier transforming to produce the image estimate $|\widehat{ff}_b(x, y)|$.

To obtain a clearer understanding of the processing requirements, the locations of the raw data in terms of (ω_b, u) in the (k_x, k_y)-domain are shown in Fig. 4.6. The data is collected along polar radii in the wavenumber domain and is valid only for wavenumbers covered by the gated transmitted signal. The inverse polar remapping takes the polar samples and re-maps them onto a uniform (baseband) grid in (k_x, k_y) appropriate for inverse Fourier transformation via the inverse FFT. This remapping operation is carried out using an interpolation process such as the one described in [82, pp133-156]. Following the remapping operation, an extraction and weighting operation similar to the process used when generating strip-map images is performed. Because the radiation pattern does not enter as an amplitude modulation term, $WW(k_x, k_y)$ is a 2-D Hamming window across the spatial bandwidths of
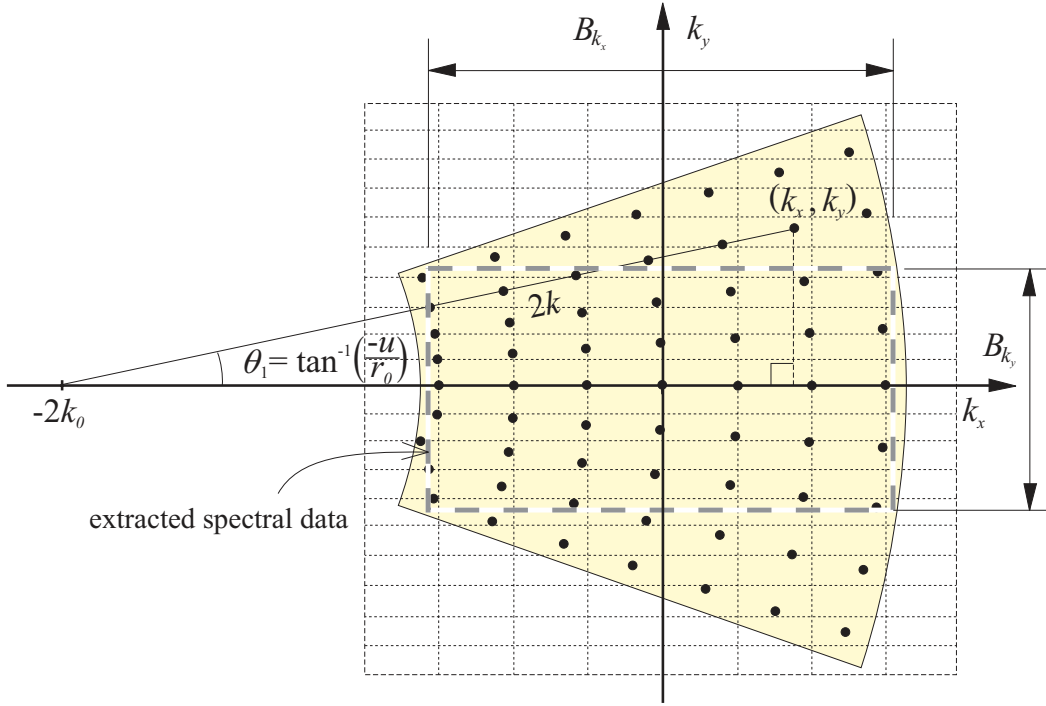
Figure 4.6   The two-dimensional collection surface of the wavenumber data. The heavy dots indicate the locations of the raw polar samples along radii $2k$ at angle θ_1. The underlying rectangular grid shows the format of the samples after the polar to Cartesian mapping (interpolation). The spatial bandwidths B_{k_x} and B_{k_y} outline the rectangular section of the wavenumber data that is extracted, windowed, and inverse Fourier transformed to produce the image estimate. Note that this extracted rectangular area is a much larger proportion of the polar data in a realistic system.

the extracted data shown in Fig. 4.6, where

$$B_{k_x} \approx \sqrt{(2k_{\max})^2 - (k_{\min}\Delta\theta)^2} - 2k_{\min} \approx \frac{4\pi B_c}{c} - \frac{(k_{\min}\Delta\theta)^2}{4k_{\max}}$$

$$B_{k_y} \approx 4k_{\min}\sin\left(\frac{\Delta\theta}{2}\right) \approx 2k_{\min}\Delta\theta,$$

(4.86)

where $k_{\min}$ and $k_{\max}$ are the minimum and maximum wavenumbers in the transmitted pulse, B_c is the deramped pulse bandwidth (in Hertz), D is the effective aperture length, $\Delta\theta \approx L_{\mathrm{sp}}/r_0$ is the angular diversity of the aperture, i.e. it is the difference between the final squint angle and the initial squint angle and L_{sp} is the spotlight synthetic aperture length shown in Fig. 4.5. In typical spotlight SAR systems the range of slew angles is small and the polar data is close to a rectangular shape in the wavenumber domain.

The definition of the x and y ordinates in the final image estimate depends on the number of samples obtained during the interpolation onto the wavenumber domain, the sample spacing chosen, and the

amount of zero padding used. The resolution in the final image estimate is

$$\delta x_{\text{3dB}} = \frac{2\pi}{B_{k_x}} \approx \frac{c}{2B_{\text{eff}}} \qquad \text{where} \qquad B_{\text{eff}} \approx B_c - \frac{c(k_{\min}\Delta\theta)^2}{16\pi k_{\max}}$$
$$\delta y_{\text{3dB}} = \frac{2\pi}{B_{k_y}} \approx \frac{\pi}{k_{\min}\Delta\theta} \approx \frac{\pi r_0}{k_{\min}L_{\text{sp}}}. \tag{4.87}$$

The deramp processing assumed in the receiver of the tomographic model effectively transduces the raw signal directly into a demodulated signal in Fourier space. This property allows the sampling rate of the demodulator to be substantially less than the signal bandwidth (see Section 2.4.3). Therefore the range sampling requirement are again met by the receiver, but at a much lower rate than any of the other models. The along-track sampling requirement can be determined as follows; with reference to Fig. 4.6, the sample spacing along k_y at $k_x = 2k_{\min}$ must be such that it can reconstruct a patch of diameter $2X_0$, i.e., $\Delta_{k_y} = 2\pi/(2X_0)$. The sample spacing along k_y is related to the along-track spacing via $\Delta_{k_y} \approx 2k_{\min}\Delta_u/r_0$. Equating these two expressions gives the along-track sample spacing

$$\Delta_u \approx \frac{\pi r_0}{2k_{\min}X_0} = \frac{D}{4}, \tag{4.88}$$

where the patch diameter $2X_0 = 4\pi r_0/(k_{\min}D)$ is the null-to-null width of the aperture radiation pattern at the lowest frequency. So for spotlight mode systems, the real aperture length and the standoff range set the patch size and the along-track sample spacing, while the synthetic aperture length and the standoff range set the along-track resolution.

The required spotlight synthetic aperture length L_{sp} is determined by giving an initial estimate of the required range and along-track resolution in the final image. If equal resolution in each direction is required then setting $\delta y_{\text{3dB}} = \delta x_{\text{3dB}} \approx c/(2B_c)$, and using (4.87) to determine L_{sp} gives a reasonable estimate of the minimum synthetic aperture length. Then the effective bandwidth B_{eff} of the transmitted pulse can be determined. If both of the resolution estimates based on these parameters are considered conservative and slightly larger values are used, then the final image estimate has the desired resolution properties. Note that deramp processing also results in a loss of bandwidth that must also be accounted for (see Appendix D of Jakowatz *et al* [82]).

4.4.4 Plane wave spotlight inversion

A spotlight inversion scheme based on the plane wave assumption operates slower than the tomographic formulation due to the extra operations of pulse compression (a 1-D Fourier transform and a multiply) and the phase multiply to align the reflections about the scene center. However, it is a useful inversion scheme if a waveform other than a LFM waveform has been transmitted, and the system parameters

Wavefront curvature	Deramp error	Reconstruction without polar remapping	
		Quadratic error	Hyperbolic error
$2\,\delta y_{3\mathrm{dB}}\sqrt{\dfrac{2r_0}{\lambda_0}}$	$2\,\delta y_{3\mathrm{dB}}\dfrac{f_0}{\sqrt{k_c}}$	$\dfrac{4\,\delta y_{3\mathrm{dB}}^2}{\lambda_0}$	$\dfrac{4\,\delta x_{3\mathrm{dB}}\,\delta y_{3\mathrm{dB}}}{\lambda_0}$

Table 4.2 Limitations on the spotlight patch diameter due to the approximations made developing the tomographic spotlight model (see pp95-97 and Appendix B of Jakowatz *et al*, 1996 [82]).

still satisfy the plane wave assumption. The inversion of (4.77) is

$$
\begin{aligned}
\widehat{FF}'_b(k_x, k_y) &= WW(k_x, k_y) \cdot \mathcal{P}_b^{-1}\left\{ \exp\left[j2k\left(\sqrt{r_0^2 + u^2} - r_0 \right) \right] \cdot P_b^*(\omega_b) \cdot Ee'_b(\omega_b, u) \right\} \\
&= WW(k_x, k_y) \cdot \mathcal{P}_b^{-1}\left\{ \exp\left[j2k\left(\sqrt{r_0^2 + u^2} - r_0 \right) \right] \cdot Ss'_b(\omega_b, u) \right\},
\end{aligned}
\tag{4.89}
$$

where the polar-to-Cartesian mapping is the same as for the tomographic formulation and where the phase multiply (along-track dechirp operation) can be removed by using the same time gating scheme as used in the tomographic formulation.

4.4.4.1   Limitations of the tomographic/plane wave inversion schemes

During the development of the spotlight model, two important assumptions were made; the first was that deramp processing could be used, and the second was that the wavefronts could be considered to be plane over the patch being imaged. As these assumptions break down, linear and quadratic errors are introduced in the wavenumber domain [82, p362,p364]. These phase errors cause imaging blurring and distortion, ultimately limiting the patch size and resolution achievable by a tomographic spotlight processor. Table 4.2 displays the limitations of the assumptions made throughout the derivation of the tomographic model (see pp95-97 and Appendix B of Jakowatz *et al*, 1996 [82]), it also contains the limitation on whether polar reformatting has to be performed or not. Images containing examples of these errors are given in Appendix E of Jakowatz *et al*, 1996 [82].

To examine the effects of the tomographic assumptions on a radar system consider a system with $\lambda_0 = 3$cm, $r_0 = 10$km, $K_c = 7 \times 10^{12}$Hz/s, and $\delta x_{3\mathrm{dB}} = \delta y_{3\mathrm{dB}} = 1$m. The maximum processable patch size before the plane wave assumption breaks down is about 1.6km and before the deramp assumption breaks down is about 7.5km. If the polar remapping step is missed out then the patch size is limited to about 130m. Given then, that polar remapping is performed, this radar system is able to produce high-resolution images of a reasonable area of terrain. Consider the same errors now for the Kiwi-SAS

with $\lambda_0 = 5$cm, $r_0 = 100$m, $K_c = 4 \times 10^5$Hz/s, and $\delta x_{3\text{dB}} = \delta y_{3\text{dB}} = 5$cm. The plane wave and deramp errors limit the patch to 6.3m and 4.7m respectively, and if polar remapping is not performed this drops to 20cm! Obviously, for the Kiwi-SAS system, producing spotlight images in this way is not an option. However, the following generalized spotlighting algorithm does not suffer from any of the limitations of the tomographic/plane wave model.

4.4.5 General spotlight inversion

When the plane wave assumption of the tomographic model no longer holds, the wavenumber inversion methods developed for strip-map models can still be used. The offset Fourier data for spotlight mode is produced in *exactly* the same way as in the wavenumber inversions developed for strip-map systems, however the beam steering employed by the spotlight mode produces a larger Doppler bandwidth which allows higher along-track resolution in the final image. This higher resolution comes at the expense of an image of limited spatial extent. The overall path travelled by the platform, i.e., the spotlight synthetic aperture length, enters as the function that limits the AM-PM waveform which sets the along-track Doppler bandwidth producing the along-track resolution. The baseband version of (4.72) is

$$Ee_b'(\omega_b, u) = P_b(\omega_b) \cdot \exp(j2kr_0) \cdot \int_{x'} \int_y ff(x', y) \cdot \exp\left[-j2k\sqrt{(x' + r_0)^2 + (y - u)^2}\right] dx'dy, \quad (4.90)$$

where the spotlight aperture is limited in along-track to $u \in [-L_{\text{sp}}/2, L_{\text{sp}}/2]$.

The along-track sampling requirement is based on the Doppler bandwidth produced by targets in the spotlight patch. The instantaneous Doppler of the phase function in (4.90) is

$$k_{ui}(\omega_b, u) = \frac{2k(y - u)}{\sqrt{(x' + r_0)^2 + (y - u)^2}}. \quad (4.91)$$

As the synthetic aperture is limited to $u \in [-L_{\text{sp}}/2, L_{\text{sp}}/2]$, the target dependent band of Doppler wavenumbers transduced by the system is

$$k_u(x, y) \in \left[\frac{2k(y - L_{\text{sp}}/2)}{\sqrt{(x' + r_0)^2 + (y - L_{\text{sp}}/2)^2}}, \frac{2k(y + L_{\text{sp}}/2)}{\sqrt{(x' + r_0)^2 + (y + L_{\text{sp}}/2)^2}}\right]. \quad (4.92)$$

Maximum Doppler signals are produced by targets at $x' = 0$, $|y| \leq X_0$, for the highest transmitted wavenumber, k_{max}, so the maximum Doppler bandwidth is

$$B_{k_u} = \frac{4k_{\text{max}}(X_0 + L_{\text{sp}}/2)}{\sqrt{r_0^2 + (X_0 + L_{\text{sp}}/2)^2}} \quad (4.93)$$

$$= 2 \cdot 2k_{\text{max}} \sin \theta_2.$$

This bandwidth can be interpreted as twice the maximum Doppler signal produced by a target lying on the opposite side of the patch from one end of the aperture, where θ_2 is the angle between these two locations. The along-track sample spacing adequate to sample this Doppler bandwidth is

$$\Delta_u = \frac{2\pi}{B_{k_u}} = \frac{\pi\sqrt{r_0^2 + (X_0 + L_{\text{sp}}/2)^2}}{2k_{\max}(X_0 + L_{\text{sp}}/2)}. \tag{4.94}$$

Unnervingly this along-track sample spacing depends on the synthetic aperture length! This restriction can be removed if it is noted that the along-track signal is composed of an ensemble of chirp-like signals with different origins corresponding to target locations. This observation, if combined with the fact that the patch diameter $2X_0$ is much less than the along-track extent of the aperture L_{sp}, allows a deramping-type operation to be performed on the along-track signal to *compress* the along-track signal. Consider the *phase* of the return from a point target at $(x', y) = (0, 0)$; that is,

$$Ee_0'(\omega_b, u) \equiv \exp\left[-j2k\left(\sqrt{r_0^2 + u^2} - r_0\right)\right]. \tag{4.95}$$

The conjugate of this signal can be used to deramp, dechirp or *compress* the along-track signal:

$$
\begin{aligned}
Ee_c'(\omega_b, u) &\equiv Ee_b'(\omega_b, u) \cdot Ee_0'^{*}(\omega_b, u) \\
&= P_b(\omega_b) \cdot \int_{x'} \int_y ff(x', y) \cdot \exp\left\{-j2k\left[\sqrt{(x' + r_0)^2 + (y - u)^2} - \sqrt{r_0^2 + u^2}\right]\right\} dx'dy.
\end{aligned} \tag{4.96}
$$

The Taylor series expansion of the square root function is (with $L_{\text{sp}} \ll r_0$) [146, p289]

$$\sqrt{(x' + r_0)^2 + (y - u)^2} - \sqrt{r_0^2 + u^2} \approx \frac{(x'^2 + y^2) + 2xr_0}{2r_0} - \frac{uy}{r_0} + \ldots \tag{4.97}$$

The higher order terms of (4.97) are negligible [146, p289]. Thus, the only significant term that depends on u on the right side of (4.97) is $-yu/r_0$. In this case, the instantaneous Doppler wavenumbers generated from the along-track phase function in (4.96) are

$$k_u(y) = \frac{\partial\phi(u)}{\partial u} \approx -2k\frac{y}{r_0}. \tag{4.98}$$

Targets within the patch of diameter $2X_0$ generate a Doppler bandwidth of

$$B_{k_u} = 2k_{\max}\frac{2X_0}{r_0}, \tag{4.99}$$

implying an along-track sample spacing of

$$\Delta_{u_c} = \frac{2\pi}{B_{k_u}} = \frac{\pi r_0}{2k_{\max}X_0} = \frac{D_s}{4},$$
(4.100)

where the subscript 'c' refers to the sample spacing of the compressed signal and D_s is the spotlight aperture length (which is generally larger than the aperture length used in strip-map applications). Thus, by compressing the along-track signal we arrive at the same along-track sampling rate as all the other synthetic aperture models. This analysis shows that the raw spotlight data can be recorded at sample spacings of $D_s/4$, however, once collected, it must be compressed in the along-track direction and upsampled to the along-track sample spacing specified in (4.94).

Practical implementation of this along-track compression concept corresponds to a pre-processing step. If the raw data $ee_b(t, u)$ is collected at a sample spacing $\Delta_{u_c} = D_s/4$ then it is aliased in the along-track direction and it needs to be upsampled to the along-track spacing required by (4.94). The upsampling operation is achieved by first performing a temporal Fourier transform (via the FFT) and compressing the along-track signal:

$$Ee'_c(\omega_b, u) = Ee'_b(\omega_b, u) \cdot \exp\left[j2k\left(\sqrt{r_0^2 + u^2} - r_0\right)\right],$$
(4.101)

where it is important to note that all functions are sampled at the same rate Δ_{u_c}. Now since $Ee'_c(\omega_b, u)$ has lost most of its high spatial frequencies, it is usually well sampled at Δ_{u_c} and so its along-track Fourier transform $EE'_c(\omega_b, k_u)$ is not aliased. Thus, we can pad the Doppler wavenumber dimension to the sampling rate required via (4.94)

$$EE'_{cd}(\omega_b, k_u) \equiv \begin{cases} 0 & -\frac{\pi}{\Delta_u} \leq k_u \leq -\frac{\pi}{\Delta_{u_c}} \\ EE'_c(\omega_b, k_u) & \text{for} \quad |k_u| \leq \frac{\pi}{\Delta_{u_c}} \\ 0 & \frac{\pi}{\Delta_{u_c}} \leq k_u \leq \frac{\pi}{\Delta_u} \end{cases}$$
(4.102)

This correctly sampled signal can then be inverse spatial Fourier transformed and have the along-track chirp reinserted, i.e. it can be *decompressed*,

$$Ee'_d(\omega_b, u) = Ee'_{cd}(\omega_b, u) \cdot \exp\left[-j2k\left(\sqrt{r_0^2 + u^2} - r_0\right)\right],$$
(4.103)

where the phase function is this time sampled at Δ_u given in (4.94). This data is now adequately sampled for inversion.

The generalized spotlight inversion can proceed via any of the strip-map algorithms, for example,

the wavenumber inversion would be as follows; the image wavenumber estimate is

$$\widehat{FF}'_b(k_x, k_y) = WW(k_x, k_y) \cdot \mathcal{S}_b^{-1} \left\{ \exp\left[j\left(\sqrt{4k^2 - k_u^2} - 2k\right) \cdot r_0 \right] \cdot P_b^*(\omega) \cdot EE'_d(\omega_b, k_u) \right\} \qquad (4.104)$$

where the spectrum of the decompressed spotlight data, $EE'_d(\omega, k_u)$, is now adequately sampled.

Figure 4.7 shows the data produced by a hypothetical spotlight SAS system using the parameters in Table 7.4 with the requirement that $\delta y_{3\text{dB}} = 2\delta x_{3\text{dB}}$ (so that the bitmaps produced were not too large). Figure 4.7(a) shows the pulse compressed raw data from 4 targets within a patch of diameter $X_0 = 3\text{m}$ at a range of $r_0 = 30\text{m}$. Due to the beam steering employed in spotlight mode, the targets always appear in the data and no along-track amplitude modulation by the radiation pattern occurs. Figure 4.7(b) shows the along-track Fourier transform of the data in Fig. 4.7(a) after is has been resampled to satisfy the sampling requirement in (4.94). Each target produces a different Doppler bandwidth, contrast this with the strip-map data in Fig. 4.3(b). Also note that the induced LFM component that is corrected by SRC is more obvious in Fig. 4.7(b) where there is no along-track amplitude modulating effects to lower the signal amplitude. Figure 4.7(c) shows the spectrum of the spotlight data. Note how the highest frequency in the transmitted pulse produces the largest Doppler wavenumbers, thus the sampling requirement in (4.94) depends on $k_{\max}$. The mapping of this spectral data into the wavenumber space of the image estimate does not correspond to a polar mapping as it does in the tomographic formulation, it is a Stolt mapping (given that a wavenumber algorithm is being used in this example, however, the shape remains the same when a range-Doppler or chirp scaling algorithm is used). The phase matched and Stolt mapped wavenumber data is shown in Fig. 4.7(d). Because of its unusual shape, an image produced from this curved spectral data produces undesirable sidelobe responses in the final image estimate. Thus, it is normal to extract a rectangular region (to force the sidelobes into the along-track and range directions) and to weight the data with a suitable weighting function. Figures 4.7(e) and (f) show the extracted wavenumber data and the final image estimate (no weighting was used to reduce the sidelobes in the image). It is also common practice to extract partially overlapping segments of the wavenumber data in Fig. 4.7(d) and use that to produce different "looks" of the scene. Non-coherent addition of these looks reduces speckle in the final image estimate (for more details on multilook processing see Section 4.7).

In spotlight mode, the Doppler bandwidth is not determined by the radiation pattern, and hence the *real* aperture length, it is determined by the length of the *synthetic* aperture and the target location. To ensure each target in the final image estimate has a space invariant point spread function, the window function (applied to Fig. 4.7(d) to produce Fig. 4.7(e)) must window the along-track wavenumber bandwidth to the Doppler bandwidth produced by a hypothetical target at the scene center, i.e., the
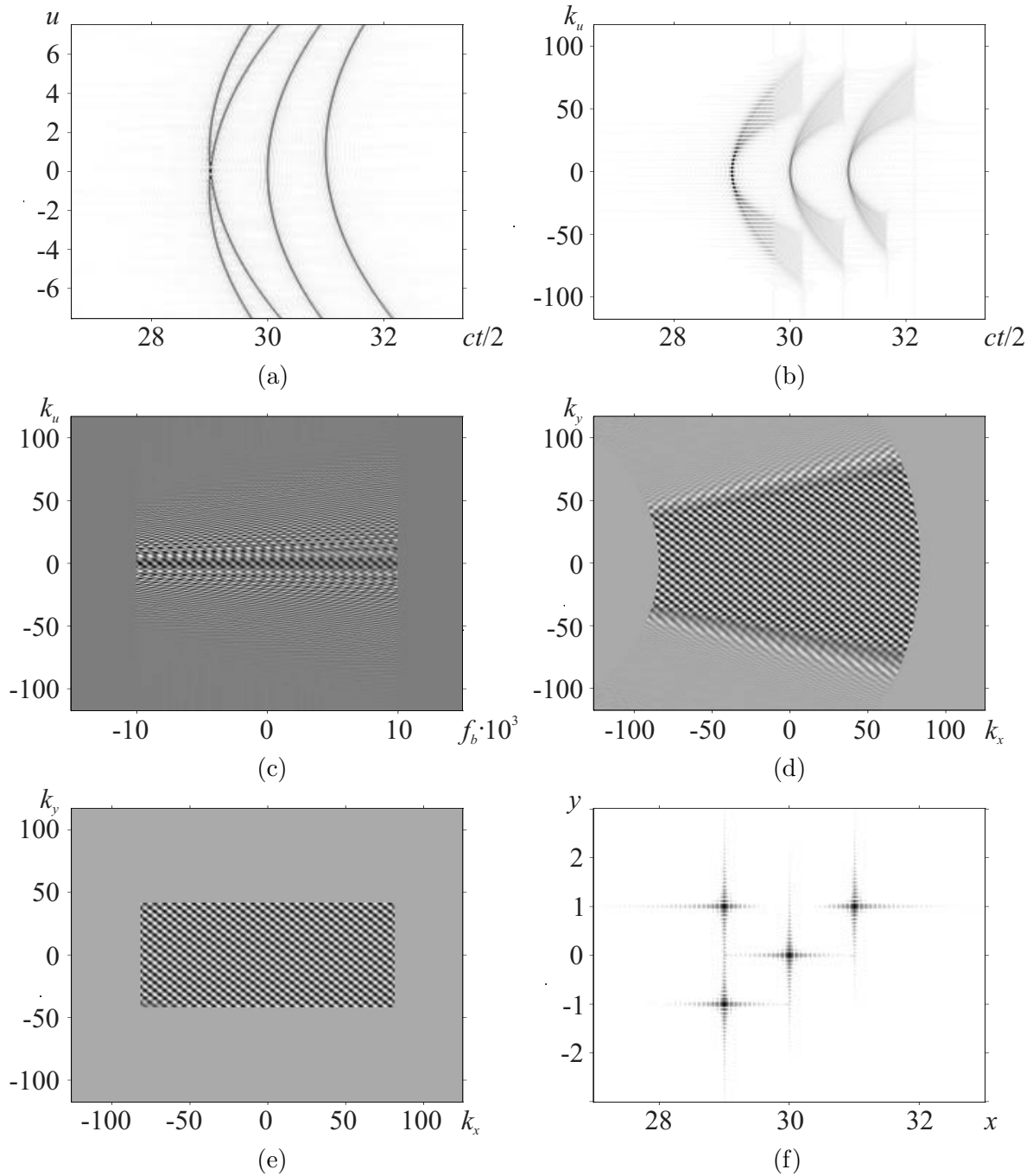
Figure 4.7   Generalized spotlight imaging. (a) raw pulse compressed spotlight data $|ss_b(t,u)|$. (b) range-Doppler data $|sS_b(t,k_u)|$, note the residual LFM components at extreme wavenumbers (these are compensated by SRC in the range-Doppler algorithm). (c) the data spectrum $\mathrm{real}\{SS'_b(\omega_b,k_u)\}$. (d) the spectral data after phase matching and Stolt mapping. (e) the spectral estimate, $\mathrm{real}\{\widehat{FF}'_b(k_x,k_y)\}$, after windowing to ensure reasonable sidelobe response in the image estimate (a weighting function would also normally be applied). (f) the image estimate $|\widehat{ff}_b(x,y)|$ (note that the final image extent or patch size is smaller than the extent of the raw data).

along-track wavenumber signal is limited to

$$B_{k_y} = \frac{2k_{\min}L_{\mathrm{sp}}}{\sqrt{r_0^2 + (L_{\mathrm{sp}}/2)^2}} \approx 2k_{\min}\Delta\theta, \tag{4.105}$$

where $\Delta\theta \approx L_{\mathrm{sp}}/r_0$ is the angular range of slew angles. The range wavenumber bandwidth available for producing the final image estimate is then

$$B_{k_x} \approx \sqrt{(2k_{\max})^2 - (k_{\min}\Delta\theta)^2} - 2k_{\min} \approx \frac{4\pi B_c}{c} - \frac{(k_{\min}\Delta\theta)^2}{4k_{\max}}. \tag{4.106}$$

The final image has the following resolution

$$\begin{aligned}
\delta x_{3\mathrm{dB}} &= \frac{2\pi}{B_{k_x}} \approx \frac{c}{2B_{\mathrm{eff}}} \qquad \text{where} \qquad B_{\mathrm{eff}} \approx B_c - \frac{c(k_{\min}\Delta\theta)^2}{16\pi k_{\max}} \\
\delta y_{3\mathrm{dB}} &= \frac{2\pi}{B_{k_y}} \approx \frac{\pi}{k_{\min}\Delta\theta} \approx \frac{\pi r_0}{k_{\min}L_{\mathrm{sp}}}.
\end{aligned} \tag{4.107}$$

Because the along-track bandwidth of the targets produced via a generalized spotlight processor have to be windowed to ensure that the impulse response of the targets in the final image estimate are spatially invariant, images produced using generalized spotlighting have the same resolution as the tomographic formulation, i.e., targets have the same along-track bandwidth after processing. Generalized spotlight inversion is not restricted by the plane wave approximation of the tomographic model or the quadratic error function arising from the deramp processing assumption and as such can process images from any spotlight system.

4.5   VELOCITY-INDUCED SYNTHETIC APERTURE IMAGING

This section develops the *velocity-induced* synthetic aperture model, i.e., the generalized form of Doppler beam sharpening (DBS), and relates the signal model to that of *spatially-induced* synthetic apertures.

4.5.1   FM-CW signal model

Consider the transmitted signal in modulated complex form [146, pp64-100]

$$p_{mr}(t) = p_{br}(t)\exp(j\omega_0 t), \tag{4.108}$$

where the baseband *repeated* signal is

$$
\begin{aligned}
p_{br}(t) &= p_b(t) \odot_t \sum_n \delta(t - n\tau_{\mathrm{rep}}) \\
&= \sum_n p_b(t - n\tau_{\mathrm{rep}})
\end{aligned}
\tag{4.109}
$$

and the repeated pulse is

$$
p_b(t) = \mathrm{rect}\left(\frac{t}{\tau_{\mathrm{p}}}\right) \cdot \exp(j\pi K_c t^2).
\tag{4.110}
$$

Because it is impossible to transmit and receive for an infinite period of time, the model needs a time limiting function analogous to the L_{sa} or L_{sp} used in spatially-induced synthetic aperture systems, i.e. $u \in [-L_{\mathrm{sa}}/2, L_{\mathrm{sa}}/2]$ or $u \in [-L_{\mathrm{sp}}/2, L_{\mathrm{sp}}/2]$. The temporal analogue is $t \in [-T/2, T/2]$, so that $p_{br}(t)$ becomes

$$
\begin{aligned}
p_{br}(t) &= w(t) \cdot \left[p_b(t) \odot_t \sum_n \delta(t - n\tau_{\mathrm{rep}}) \right] \\
&= w(t) \cdot \sum_n p_b(t - n\tau_{\mathrm{rep}}),
\end{aligned}
\tag{4.111}
$$

where $w(t)$ is a standard window function typically taken as $\mathrm{rect}(t/T)$. The temporal Fourier transform of this signal model gives

$$
\begin{aligned}
P_{br}(\omega_b) &= W(\omega_b) \odot_{\omega_b} \left[P_b(\omega_b) \cdot \frac{2\pi}{\tau_{\mathrm{rep}}} \cdot \sum_n \delta\left(\omega_b - \frac{2\pi n}{\tau_{\mathrm{rep}}} \right) \right] \\
&= \frac{2\pi}{\tau_{\mathrm{rep}}} \cdot \sum_n P_b(n\omega_{\mathrm{rep}}) W(\omega_b - n\omega_{\mathrm{rep}}).
\end{aligned}
\tag{4.112}
$$

The repeated pulse spectrum consists of weighted, repeated copies of the narrow function $W(\omega_b)$ located at harmonics $n\omega_{\mathrm{rep}} = 2\pi n/\tau_{\mathrm{rep}}$, windowed in frequency by the pulse spectrum $P_b(\omega_b)$ which at this stage still has a phase function associated with it.

The transmitted FM-CW waveform is equivalently expressed as a linear combination of the (modulated) harmonics at $\omega_n = \omega_0 + n\omega_{\mathrm{rep}}$; that is,

$$
p_{mr}(t) = \sum_n p_{mn}(t),
\tag{4.113}
$$

where the complex modulated harmonic is

$$p_{mn}(t) = P_b(n\omega_{\text{rep}})\exp[j(\omega_0 + n\omega_{\text{rep}})t] \tag{4.114}$$

for $t \in [-T/2, T/2]$, where $P_b(n\omega_{\text{rep}})$ is a *complex constant* for each harmonic.

4.5.2   FM-CW synthetic aperture model

The system model is developed by considering a single harmonic of the FM-CW signal $\omega_n = \omega_0 + n\omega_{\text{rep}}$, i.e., the transmitted pulse is considered to be

$$p_{mn}(t) = \exp(j\omega_n t). \tag{4.115}$$

The effects of limiting time, i.e. limiting the synthetic aperture, the pulse spectrum constant for each harmonic $P_b(n\omega_{\text{rep}})$, and the radiation patterns of the real apertures is considered later. The system model for the nth harmonic is

$$
\begin{aligned}
vv_n(t_1, t_2) &= \int_x \int_y ff(x,y) \cdot p\left(t_1 - \frac{2}{c}\sqrt{x^2 + (y - v_p t_2)^2}\right) dxdy \\
&= \int_x \int_y ff(x,y)\exp\left[j\omega_n\left(t_1 - \frac{2}{c}\sqrt{x^2 + (y - v_p t_2)^2}\right)\right] dxdy \\
&= \exp(j\omega_n t_1) \cdot \int_x \int_y ff(x,y)\exp\left(-j2k_n\sqrt{x^2 + (y - v_p t_2)^2}\right) dxdy,
\end{aligned}
\tag{4.116}
$$

where time t has been split into range or fast-time t_1 and along-track or slow-time $t_2 = u/v_p$, where u is the along-track axis and v_p is the platform velocity, and $k_n = \omega_n/c$. Fourier transforming with respect to fast-time t_1 gives

$$Vv_n(\omega_1, t_2) = \delta(\omega_1 - \omega_n) \cdot \int_x \int_y ff(x,y)\exp\left(-j2k_n\sqrt{x^2 + (y - v_p t_2)^2}\right) dxdy. \tag{4.117}$$

Fourier transforming with respect to t_2 using the principle of stationary phase gives

$$VV_n(\omega_1, \omega_2) = \delta(\omega_1 - \omega_n) \cdot \sqrt{\frac{\pi x}{jk_n}} \cdot \int_x \int_y ff(x,y)\exp\left(-j\sqrt{4k_n^2 - \left(\frac{\omega_2}{v_p}\right)^2} \cdot x - j\left(\frac{\omega_2}{v_p}\right) \cdot y\right) dxdy \tag{4.118}$$

which is analogous to the spatially-induced synthetic aperture recording in 2-D Fourier space shown in (4.34).

In reality time t is treated as a *continuous* variable and it is *not* sectioned into slow- and fast-time

components before Fourier transformation to the Fourier domain of t, i.e. Ω. To develop the relationship between the *required* 2-D signal shown in (4.118) and the available FM-CW signal in 1-D Fourier space, consider the signal:

$$
\begin{aligned}
Vv(\omega_n, t) &\equiv vv_n(t_1, t_2)|_{t_1=t, t_2=t} \\
&= \int_{\omega_1} \int_{\omega_2} VV_n(\omega_1, \omega_2) \exp\left[j(\omega_1 + \omega_2)t\right] d\omega_1 d\omega_2.
\end{aligned}
\tag{4.119}
$$

Making the two-dimensional transform given by

$$
\begin{aligned}
\Omega &\equiv \omega_1 + \omega_2 \\
\beta &\equiv \omega_1 - \omega_2,
\end{aligned}
\tag{4.120}
$$

gives

$$
Vv(\omega_n, t) = \int_{\beta} \int_{\Omega} VV_n\left(\frac{\Omega + \beta}{2}, \frac{\Omega - \beta}{2}\right) \exp(j\Omega t) d\Omega d\beta.
\tag{4.121}
$$

Taking the Fourier transform with respect to time t gives

$$
VV(\omega_n, \Omega) = \int_{\beta} SS_n\left(\frac{\Omega + \beta}{2}, \frac{\Omega - \beta}{2}\right) d\beta
\tag{4.122}
$$

Substituting (4.118) into (4.122) and carrying out the integration over the delta function gives

$$
\begin{aligned}
VV(\omega_n, \Omega) &= \sqrt{\frac{\pi}{jk_n}} \cdot \int_x \int_y ff_x(x, y) \exp\left(-j\sqrt{4k_n^2 - \left(\frac{\Omega - \omega_n}{v_p}\right)^2} \cdot x - j\left(\frac{\Omega - \omega_n}{v_p}\right) \cdot y\right) dx dy \\
&= \sqrt{\frac{\pi}{jk_n}} \cdot \mathcal{H}\left\{FF_x(k_x, k_y)\right\}.
\end{aligned}
\tag{4.123}
$$

The coordinate remapping $\mathcal{H}^{-1}\{\cdot\}$ is given by

$$
\begin{aligned}
k_x(\omega_n, \Omega) &\equiv \sqrt{4k_n^2 - \left(\frac{\Omega - \omega_n}{v_p}\right)^2} \\
k_y(\omega_n, \Omega) &\equiv \left(\frac{\Omega - \omega_n}{v_p}\right).
\end{aligned}
\tag{4.124}
$$

Including the effects of the transmitted pulse and real apertures gives

$$VV(\omega_n, \Omega) = \sqrt{\frac{\pi}{jk_n}} \cdot P_b(n\omega_{\text{rep}}) \cdot A\left(\frac{\Omega - \omega_n}{v}\right) \cdot \mathcal{H}\left\{FF_x(k_x, k_y)\right\}, \tag{4.125}$$

where $P_b(n\omega_{\text{rep}})$ is a complex constant for each harmonic n, and the amplitude function $A(\cdot)$ is identical to the radiation function used in spatially-induced synthetic aperture model with $k_u = (\Omega - \omega_n)/v_p$ and $|k_u| \leq 2\pi/v_p \cdot \text{PRF}$.

4.5.3   Relationship to the spatially-induced synthetic aperture model

The relationship to the spatially-induced model in (4.37) is

$$EE_m(\omega, k_u) = \mathcal{V}\left\{VV(\omega_n, \Omega)\right\} \tag{4.126}$$

where the mapping operator $\mathcal{V}\{\cdot\}$ is

$$\begin{aligned} \omega(\omega_n, \Omega) &\equiv \omega_n \\ k_u(\omega_n, \Omega) &\equiv \left(\frac{\Omega - \omega_n}{v_p}\right) \end{aligned} \tag{4.127}$$

i.e., the baseband data, $|(\Omega - \omega_n)| \leq \pi \cdot \text{PRF}$, around each harmonic ω_n is the same as the Doppler wavenumber data, and the ω_n components are the same as the spectral components of the transmitted pulse $p_b(t)$ sampled at ω_s that are normally denoted as the range temporal frequencies ω.

Velocity-induced and spatially-induced synthetic aperture systems both exploit the Doppler effects produced between pulses that are induced by relative-platform target motion. The spatially-induced model formats the raw 1-D time sequence of echoes into a 2-D array and then performs a 2-D FFT to yield the frequency-wavenumber data necessary for producing the image estimate. Conversely, the velocity-induced model performs a longer 1-D Fourier data of the raw data, and then segments the data into the 2-D matrix required for inversions. Both schemes ultimately produce the same wavenumber data.

4.6   GENERALIZATION OF THE INVERSION SCHEMES TO THREE-DIMENSIONS

The system models developed in this chapter have assumed that the platform path was in the same plane as the scene being imaged. Obviously this is not possible in a realistic situation. The general

form of the (x, y) slant-plane to (x_g, y_g) ground-plane remap is

$$\widehat{ff}(x_g, y_g) = \mathcal{G}\left\{\widehat{ff}(x, y)\right\} \tag{4.128}$$

In the simplest conversion from the collected range parameter x to the ground range parameter x_g, the along-track variable is not altered and a fixed depression or incidence angle from the platform to the imaged scene is assumed. The required mapping, $\mathcal{G}\{\cdot\}$, is simply

$$
\begin{aligned}
x_g(x, y) &\equiv \frac{x}{\cos \phi_g} = \frac{x}{\sin \phi_i} \\
y_g(x, y) &\equiv y,
\end{aligned}
\tag{4.129}
$$

where the depression or *grazing angle*, ϕ_g, is the angle subtended from horizontal to the slant range plane defined by the boresight to the scene center. The complementary angle to the grazing angle, $\phi_i = 90 - \phi_g$, is the *incidence angle*, i.e., the angle from the nadir to boresight. For this simple model then, conversion from slant-range to ground range is a simple pixel scaling. This scaling alters the range resolution in the ground plane to

$$\delta x_{g3\text{dB}} = \frac{c}{2B \cos \phi_g}, \tag{4.130}$$

the along-track resolution is not affected.

In practical SAR systems, images generally require interpolation onto the ground plane [30]. This interpolation is necessary in spaceborne imaging applications where the final image is *geocoded*, i.e., mapped onto a geoid representing the earth's surface. Interpolation is also required in situations where the final image is registered to a map grid. Key locations, such as cross-roads or known reflectors, are located in the SAR image and are used to determine the distortion necessary to overlay the SAR image to a topographic map, or aerial photograph of the same scene. The interpolators used in these slant-to-ground plane conversions do not effect the image quality as much as the interpolators used during processing [30]. Chapter 8 of Curlander and McDonough [31] details the geometric calibration of SAR data.

In spotlight SAR, the collected spectral data can be interpreted as a 2-D ribbon through the 3-D Fourier (wavenumber) space of the imaged scene [82]. During the polar-to-rectangular conversion of the raw data, or during the Stolt mapping, this wavenumber data can be interpolated onto a 2-D wavenumber surface that is the wavenumber space of the ground-plane. Inverse Fourier transformation then produces the ground-plane image estimate.

Slant-to-ground plane conversion is necessary if the final image is to be compared with another medium such as a map, or photograph. In sonar applications, the only possible medium for comparison

is generally *another* sonar image. If both sonar images have been collected using the same geometry, then it is unnecessary to perform the conversion. In many cases, the grazing angles employed by sonar systems are shallow enough that the slant-to-ground plane conversion can be ignored.

Another method of producing ground-plane rectified images is presented by Shippey and Nordkvist, 1996 [140]. In their time-domain beamforming algorithm, the time delay and sum beamforming operations are calculated for an arbitrarily chosen grid in the ground plane. In this way, ground plane images are produced directly.

4.7   MULTI-LOOK PROCESSING

The final magnitude images produced by processing the full extent of the wavenumber domain estimate (i.e., the offset Fourier data) have a grainy appearance. This grainy appearance is due to random intensity variations in the image caused by the coherent interference of multiple reflections originating from targets lying within a single resolution cell [96,102]. This grainy appearance is referred to as *speckle* and the images are referred to as *speckle images*. For purposes of visual interpretation it is generally desirable to reduce the random pixel fluctuations in speckle images so that the pixel fluctuations are close to some local average value. *Multi-look processing* is the generic term given to methods that modify the speckle statistics in synthetic aperture images, thus improving the *radiometric resolution* of the image. The radiometric resolution of an image is defined to be the accuracy to which an image pixel can be associated with a targets scattering characteristic. Improvements in radiometric resolution usually come at the expense of reduced geometric resolution; geometric resolution being the accuracy to which a targets location, size and shape can be determined (for more detail see p3, Chaps. 7 and 8 of [31]).

Many references refer to multi-look processing as a method of filtering the along-track Doppler bandwidth into independent (non-overlapped) sections [96] that are used to produce multiple images, known as *looks*. Because each look has been obtained from a different section of the Doppler bandwidth and the Doppler frequency of a target is determined by a targets aspect angle from the *real aperture*, each look has different speckle characteristics. Thus, each look when combined incoherently produces an image with improved speckle characteristics, albeit with reduced along-track resolution. Multi-look processing works due to the offset Fourier nature of the coherent data. No matter where in the range-Doppler spectrum the data lies, an estimate of the image is formed. The complex exponential representing the offset of the Doppler section from zero Doppler is an irrelevant carrier back in the spatial domain, the incoherent summing operation contains a modulus operation that removes this phase function from each look before they are summed.

Multi-look processing is often applied to spaceborne SAR data as it acts to suppress the along-track ambiguities caused by inadequate along-track sample spacings, i.e., Δ_u close to $D/2$, instead of less

than $D/3$. Each Doppler segment has a different level of ambiguous energy; those close to the Doppler folding wavenumber have the highest levels. The ambiguous targets produced in each of these looks are suppressed when the looks are incoherently combined due the varying statistics and locations of the ambiguous targets between the different looks. In some cases, along-track undersampled and multilook processed images still contain quite visible grating lobe targets (see pp299-300 [31]).

The idea of multi-look processing is much broader than the range-Doppler application implies. The holographic properties of the synthetic aperture data means that any 2-D segment of the offset Fourier data produces an image estimate. Thus, multi-look processing can be performed in range as well as along-track. Advanced multi-look techniques use overlapped, windowed sections of the offset Fourier data. With an overlap of 40-60 percent, each look still retains statistical independence from the other looks as long as an appropriate window function is applied [107]. The overlapping technique results in a higher *equivalent number of looks* (ENL) [96, 107] than the non-overlapped technique. Moreira [107] presents a technique that combines a multi-look *high resolution* image produced using a low number of overlapped, windowed looks with *low resolution* multi-look image formed from a larger number of overlapped, windowed looks. When these two images are combined incoherently, the high resolution image improves the geometric resolution, while the low resolution image improves the radiometric resolution [107]. The ENL measure from this technique is up to 2.3 times better than the non-overlapped technique [107]. This technique is similar to the non-linear spatially variant apodisation (SVA) technique of Stankwitz *et al* [149] (see also Appendix D.3 of Carrara [21]). Still more techniques exist where pixel averaging [47, p80] or local smoothing of a single-look image is used to improve image appearance [96].

The original interpretation of multi-look in terms of Doppler filtering is due to the fact that in radar high range resolution is harder to achieve than high along-track resolution. High range resolution is hard to achieve due to two aspects of spaceborne radar systems; the first is the high speed of light, eg., the Seasat SAR used a pulse bandwidth of 19MHz, giving a slant-range resolution of 5.1m. The second problem is the slant-range to ground-plane correction; the angle from the nadir to the near swath for Seasat was about 18°, the sine correction required for ground-plane conversion drops the range resolution to 25m [47, p115]. The Seasat SAR carried a 10.7m real aperture, so the along-track resolution of a single-look image was 5.3m. To produce images with equal pixel sizes, the Seasat SAR used four-looks in *along-track* to produce images with 25m×25m resolution.

In sonar, the much lower propagation speed of sound relative to the speed of light means that high range resolution is easier to achieve with much smaller pulse bandwidths. For example, the Kiwi-SAS obtains 4cm resolution with a 20kHz pulse. However, as the Kiwi-SAS carries a 32cm real aperture, the best along-track resolution achievable is 16cm. Thus, the Kiwi-SAS could produce reduced speckle images with 16cm×16cm resolution using four-looks in *range* (when the effects of wavenumber domain windowing are included, this resolution is more likely to be 20cm×20cm).

Spotlight SAR also uses the offset Fourier nature of the data to produce multi-look images. In

spotlight mode if an n-look image is required, then it is usually a simple matter to record n times the angular diversity required for single look imagery with equal resolution in range and along-track. Alternatively, the offset Fourier data of an image with equal range and along-track resolution can be split up into n equal-sized overlapping or non-overlapping sections of Fourier space suitable for processing into n independent images. These lower resolution images can then be incoherently summed into an n-look image that still has equivalent range and along-track resolution, and has improved speckle characteristics. See pages 112-121 of Jakowatz *et al* [82] for examples of multi-look spotlight SAR images.

A final caution before leaving this section. Many references equate the multi-look Doppler filtering approach as being equivalent to producing *subaperture* images [7]. This terminology is misleading as in strip-map mode it refers to producing subapertures of the *real* aperture, not the synthetic aperture (for example, see Fig. 5.2 on p218 of [31]). In spotlight SAR however, the term subaperture is used to refer to the division of the *synthetic* aperture into various subapertures. The one-to-one relationship between u and k_y seen in the tomographic formulation of spotlight SAR makes the spotlight terminology appropriate. If a strip-map synthetic aperture was divided into subapertures, then different images would be produced! A slight misuse of the term 'look' is seen in Truong [155] where they *coherently* added four 'looks' to produce a 6.6m×25m Seasat image. The division of the Doppler bandwidth into four sections was necessary to implement processing and had nothing to do with speckle reduction. The term 'look angle' is used in Elachi [47] and Curlander and McDonough [31] to refer to the incidence angle of the boresight of the real aperture to the scene center measured from the nadir (eg. see Fig. 1.6 on p14 of [31]), and does not refer to the squint angle of a single-look in a multi-look process.

4.8   PROCESSING REQUIREMENTS OF A REALISTIC STRIP-MAP SYNTHETIC APERTURE SYSTEM

Spotlight mode processing using the tomographic or plane wave inversion has a major advantage over strip-map processing in that all the raw spotlight scene data goes into processing the final spotlight image. In strip-map mode, the data is limited in along-track extent only by the distance over which data is collected. No strip-map processor would ever attempt to process the final strip-map image in one block process. The memory limitations of digital devices also requires that the final image be processed from smaller blocks of the raw data.

In a strip-map system, the overall radiation pattern of the real apertures limits the along-track distance over which an individual target is valid for processing. This limited along-track distance is called the *synthetic aperture length* of the target and it corresponds to the 3dB width of the beam pattern at the lowest operating frequency, i.e, $L_{\mathrm{sa}} = x\lambda_{\max}/D$, where x is the range of the target from the platform position, D is the effective aperture length, and $\lambda_{\max}$ is the maximum wavelength of the
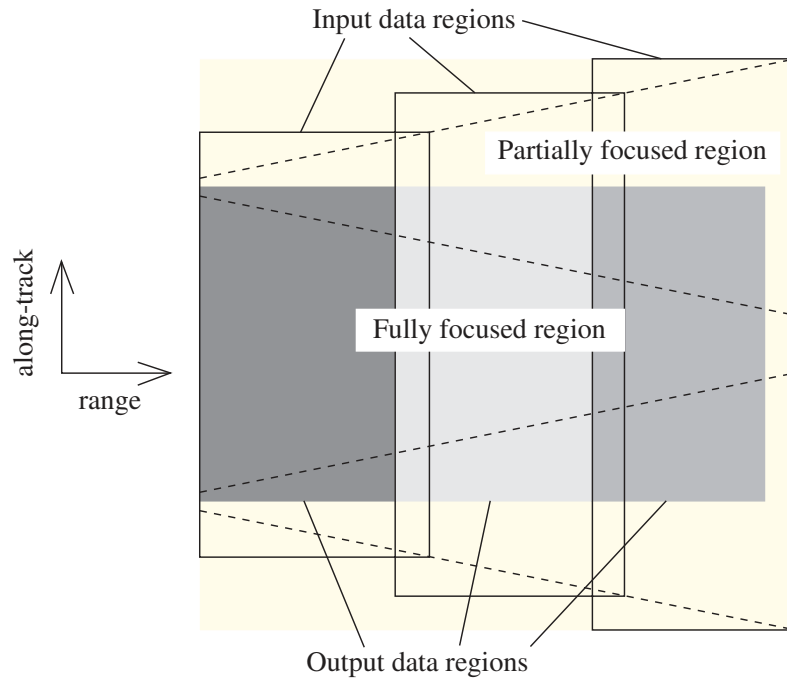
Figure 4.8 The distribution of data necessary to obtain an equal load on each processor and to produce rectangular shaped processed blocks (see the text for more details).

transmitted signal. Regardless of the fact that the Doppler wavenumber bandwidth is independent of wavelength, the spatial coverage of the beam pattern does change with wavelength, the necessary step of taking in limited spatial domain data is equivalent to a spatial windowing operation. This spatial window filters out some of the Doppler spectrum for targets near the edges of the along-track segment of spatial data. Alternatively, not all of their along-track locus is passed to the processor for focusing. This means that when processing the data in blocks, the 2-D blocks have to be overlapped in range and along-track (remember that the range overlap was necessary to counter the effects of range migration). Figure 4.8 shows an example of a realistic processor consisting of three parallel processors. Each processor is set up so that it processes the same volume of raw data, where each data set is overlapped to account for range migration effects. The outlined blocks in Fig. 4.8 contain these equivalent areas of raw data, while the shaded areas inside correspond to the blocks of fully focused data after processing. The extent of the 3dB beam pattern is also indicated with a dashed line. When the processors load up for the next block, the area within the lower edge of the beam pattern shown at the top of the figure has to be retained to fully focus the range lines currently lying just above the shaded areas of processed data.

The real-time ACID/SAMI sonar processor described in [1,2,131] uses a similar load balancing idea. However, it is important to note that the ACID method only produces a *single* processed along-track line at the end of each block process. This is due to the spatial-temporal, time-delay and sum, beamforming

approach of its synthetic aperture image processor.

4.9   SUMMARY

The development of synthetic aperture sonar processors, to date, has been focused on spatial-temporal domain, time-delay and sum, beamforming approaches or fast-correlation approaches. The synthetic aperture techniques presented in this chapter offer methods that are several orders of magnitude faster at producing final focused imagery. This is especially true of the *accelerated* chirp scaling algorithm developed in this thesis.

The use of a unified mathematical notation and the explicit statement of any change of variables via mapping operators assists in the clear presentation of the available processing algorithms.

Chapter 5

APERTURE UNDERSAMPLING AND MAPPING RATE IMPROVEMENT

This chapter begins by analysing the effects of sampling on the synthetic aperture signal model. Range and along-track ambiguity constraints are defined and their limiting effect on the mapping rate is shown. The levels of alias energy present in the raw data due to range and along-track signal aliasing are quantified by the range ambiguity to signal ratio (RASR) and the along-track ambiguity to signal ratio (AASR). An analysis of the cause, and methods for the suppression of, grating lobe targets is given. These suppression methods include; the transmission of wide bandwidth, low-Q, waveforms, aperture diversity, pulse diversity, frequency diversity, ambiguity suppression via inverse filters, and digital spotlighting.

5.1 SAMPLING EFFECTS IN SYNTHETIC APERTURE SIGNALS

A digitally based synthetic aperture system samples the spatial-temporal domain over a rectangular grid of spacing (Δ_t, Δ_u), where $\Delta_t = 2\pi/\omega_s$ is the sampling period of the receiver operating at sampling radian frequency ω_s and $\Delta_u = v_p \tau_{\mathrm{rep}}$ is the *approximately* constant along-track sample spacing that is set by the platform velocity v_p and the pulse repetition period τ_{rep}. The along-track sample spacing corresponds to a sampling Doppler wavenumber, $k_{us} \equiv 2\pi/\Delta_u$, which adequately samples all signals with Doppler wavenumbers $k_u \in [-\pi/\Delta_u, \pi/\Delta_u]$. Any spectral energy at $|k_u| > \pi/\Delta_u$ aliases back into the region $k_u \in [-\pi/\Delta_u, \pi/\Delta_u]$ and generates *alias targets* or *grating lobe targets* in the final image. The level of these alias targets in *adequately* sampled systems is insignificant. This section explains the along-track signal and its adequate sampling requirements, while other sections in this chapter deal with reducing the level of alias targets in poorly sampled systems.

The grid sampled every (Δ_t, Δ_u) samples a *rectangular* region of the (ω, k_u)-domain. However, due to the temporal-frequency dependent nature of the along-track chirp signal, any rectangular area of the spatial-temporal domain does not map to a rectangular area of the frequency-wavenumber domain. This
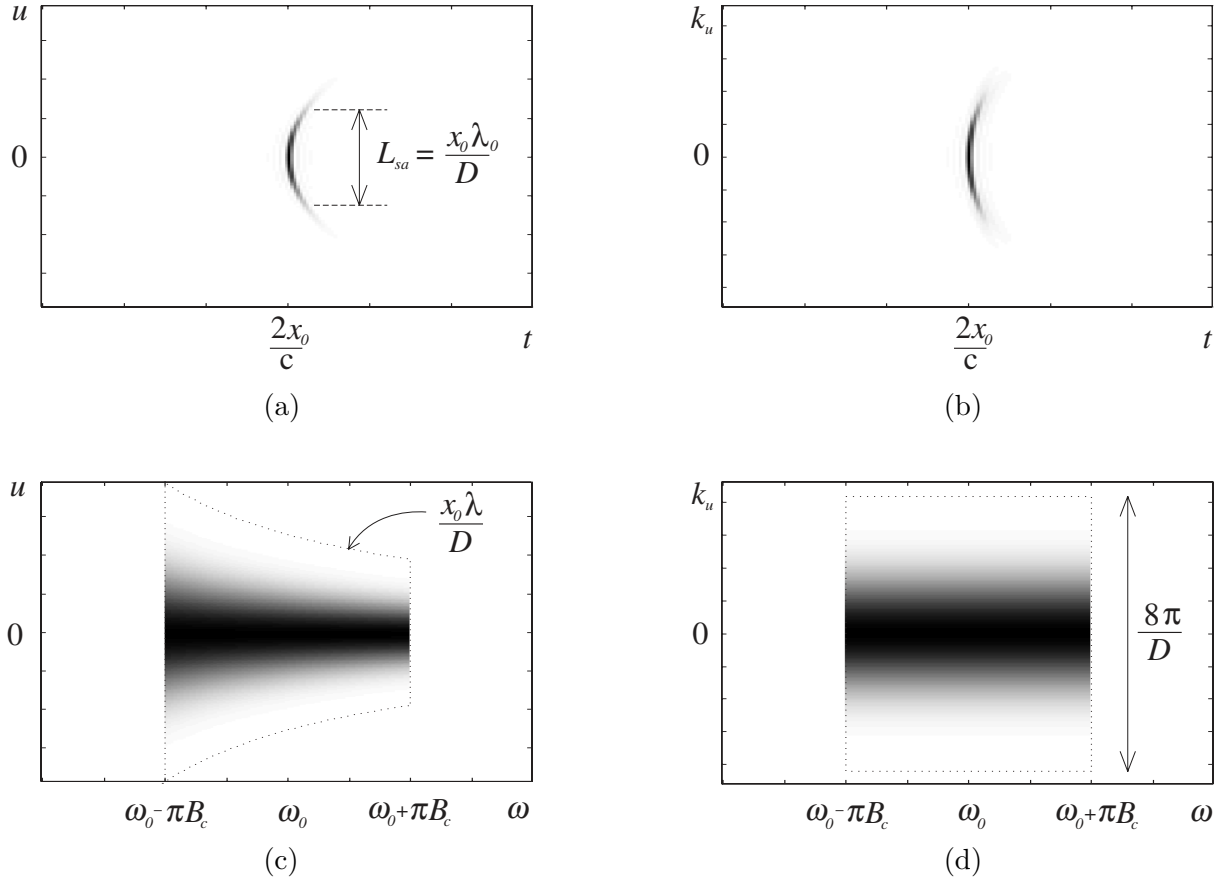
Figure 5.1 Signals of the point target, (a) spatial-temporal signal, $|ss_\delta(t,u)|^2$, (b) range-Doppler signal, $|sS_\delta(t,k_u)|^2$, (c) temporal frequency content of along-track echoes, $|Ss_\delta(\omega,u)|^2$, (d) ω-k_u domain, $|SS_\delta(\omega,k_u)|^2$. Dashed regions in (c) and (d) outline the extent of the system pulse bandwidth and first null of the radiation pattern in along-track and Doppler. All images displayed as linear greyscale. The slight broadening of the range-Doppler response for wavenumbers away from $k_u = 0$ is due to the LFM term that is corrected by secondary range compression (SRC).

statement is illustrated graphically in Fig. 5.1 and is presented mathematically shortly. Figure 5.1(a) shows the (pulse compressed) (t,u)-domain response of a point target located at $(x,y) = (x_0,0)$, Figs. 5.1(b), (c), and (d) are the transforms of the pulse compressed data, $ss_\delta(t,u)$ (see (4.12)), in the range-Doppler, (ω,u), and (ω,k_u) domains respectively. The target locus in Fig. 5.1(a) shows the effects of range cell migration. The amplitude modulation of this locus in along-track is caused by the aperture impulse response. The non-linear mapping of the along-track signal is best demonstrated in Fig. 5.1(c) and (d). The outlined area in Fig. 5.1(c) encompasses a signal of bandwidth B_c (Hertz) and the first null of the radiation pattern of an aperture of length D. The first null migrates through the (ω,u)-domain according to $u \approx x_0\lambda/D$. The frequency dependent width of the radiation pattern halves in the figure as the synthetic aperture system is assumed to have a Q of 1.5, eg. a 20kHz bandwidth

pulse at a 30kHz carrier for a sonar system, or a 200MHz pulse about a 300MHz carrier for a foliage penetrating SAR [147]. The underlying along-track chirp contained in the envelope of this radiation pattern causes the outlined area in Fig. 5.1(c) to transform to the rectangular region in Fig. 5.1(d) when spatial Fourier transformed. The radiation pattern consists of a scaled *amplitude function* (i.e., the AM part) in each domain, with an underlying phase function (i.e., the PM part) that controls the bandwidth in the along-track Fourier domain. Any spatial energy in Fig. 5.1(c) that exists *outside* of the outlined area aliases back into the rectangular area shown in the Fig. 5.1(d). Since the outlined area in Fig. 5.1(d) encompasses the main lobe of the radiation pattern, the aliased energy is due to sidelobe energy only. If the aperture illumination functions are weighted, then the level of this alias energy is low. Note that the area shown in Fig. 5.1(d) corresponds to an along-track sample spacing of $D/4$, so that any synthetic aperture system employing a sample spacing of $D/2$ can be seen to alias the signal energy either side of the 3dB width of the beam pattern back into a rectangular area in Doppler space where, for a $D/2$ spacing, $|k_u| \leq 2\pi/D$.

If we assume that the transmitted pulse spectrum is uniform over bandwidth B_c, and the transmit and receive apertures are the same length, the pulse compressed signal in the (ω, u)-domain for the point target located at $(x, y) = (x_0, 0)$ is

$$
\begin{aligned}
Ss_\delta(\omega, u) &= |P_m(\omega)|^2 \cdot A(\omega, x_0, -u) \cdot \exp\left(-j2k\sqrt{x_0^2 + u^2}\right) \\
&\approx \text{rect}\left(\frac{\omega - \omega_0}{2\pi B_c}\right) \cdot \text{sinc}^2\left(\frac{kD}{2\pi} \cdot \frac{u}{\sqrt{x_0^2 + u^2}}\right) \cdot \exp\left(-j2k\sqrt{x_0^2 + u^2}\right).
\end{aligned}
\tag{5.1}
$$

Note that in this chapter, unless otherwise noted, spectral signals are calculated with reference to the t and x axes, not the translated t' and x' axes (thus, the spectral functions are not primed). The conversion of these functions to the form required for a digital implementation is straightforward, Chapter 4 gives many examples.

In the (ω, k_u)-domain the spatial Fourier transform of (5.1) is

$$
\begin{aligned}
SS_\delta(\omega, k_u) &= |P_m(\omega)|^2 \cdot A(k_u) \cdot \exp\left(-j\sqrt{4k^2 - k_u^2} \cdot x_0\right) \qquad \text{for} \qquad |k_u| \leq 2k \\
&\approx \text{rect}\left(\frac{\omega - \omega_0}{2\pi B_c}\right) \cdot \text{sinc}^2\left(\frac{k_u D}{4\pi}\right) \cdot \exp\left(-j\sqrt{4k^2 - k_u^2} \cdot x_0\right).
\end{aligned}
\tag{5.2}
$$

where the function is only defined over visible Doppler wavenumbers, $|k_u| \leq 2k$.

Both (5.1) and (5.2) are *continuous* functions that are sampled by the digital processor in the spatial-temporal domain at spacings given by (Δ_t, Δ_u). This sampling *always* results in some form of aliasing. To determine the detrimental effects of this aliasing it is necessary to know what effects the extent of the data and the sampling frequencies have on each domain. The range signal is assumed to
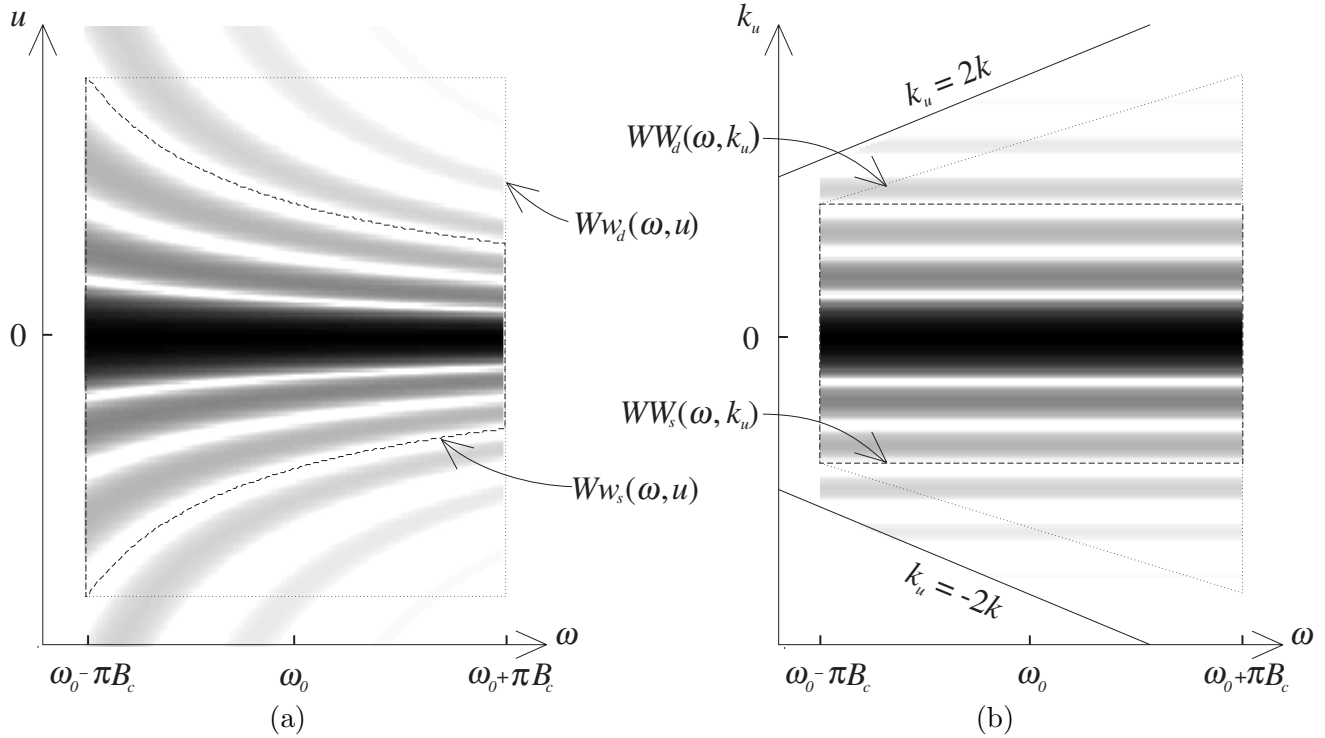
Figure 5.2 The effect of windowing on the point target response, (a) $|Ss(\omega, u)|^2$, (b) $|SS(\omega, k_u)|^2$. Each image is displayed as a logarithmic greyscale image clipped at -50dB.

have been appropriately sampled in the receiver, so the effects we are interested in are with respect to the along-track signal. Figure 5.2 is similar to Fig. 5.1(c) and (d), however, this time the figures are displayed on a logarithmic scale to show the sidelobe detail of the radiation pattern. There are also two different areas outlined in each figure. One outline obviously follows the third null of the radiation pattern, while the other shows the distortion of a rectangular area in (ω, u) when it transforms to (ω, k_u).

The rate of change of the spatial phase in (5.1) gives an ω dependent instantaneous Doppler wavenumber for any along-track position in the (ω, u)-domain; that is,

$$k_{ui}(\omega, u) = \frac{-2ku}{\sqrt{x_0^2 + u^2}}. \tag{5.3}$$

If the outlined rectangular area in Fig. 5.2(a) is limited in along-track to $u \in [-L/2, L/2]$, then (5.3) indicates that the maximum Doppler wavenumber in the outlined region is $k_u^{\max} = k_{ui}(\omega_0 + \pi B_c, -L/2)$. To sample the outlined region without aliasing requires a $k_{us} \geq 2k_u^{\max}$. Looking at Fig. 5.2(b), the maximum Doppler wavenumber lies at the highest temporal frequency and lies at the sixth null of the radiation pattern. The sample spacing required to sample this bandwidth is $D/24$, a much finer sample spacing than $D/4$. In a practical synthetic aperture system, a sampling rate of $D/4$ causes the sidelobe

energy outside of the sampled region to alias, in systems sampled at $D/2$, some of the main lobe energy also aliases. If the rectangular *data* window in (ω, u) is defined as

$$Ww_d(\omega, u) = \text{rect}\left(\frac{\omega - \omega_0}{2\pi B_c}\right) \cdot \text{rect}\left(\frac{u}{L}\right), \tag{5.4}$$

then this window contains the Doppler wavenumbers in the (ω, k_u) domain outlined by

$$
\begin{aligned}
WW_d(\omega, k_u) &= \text{rect}\left(\frac{\omega - \omega_0}{2\pi B_c}\right) \cdot \text{rect}\left(\frac{k_u\sqrt{x_0^2 + (L/2)^2}}{2kL}\right) \\
&\approx \text{rect}\left(\frac{\omega - \omega_0}{2\pi B_c}\right) \cdot \text{rect}\left(\frac{k_u x_0}{2kL}\right).
\end{aligned}
\tag{5.5}
$$

A similar argument follows for a rectangular area in the (ω, k_u) domain. A signal of bandwidth B_c and a Doppler sampling wavenumber of k_{us} defines a *sampled* windowed region of the (ω, k_u)-domain given by

$$WW_s(\omega, k_u) \equiv \text{rect}\left(\frac{\omega - \omega_0}{2\pi f_s}\right) \cdot \text{rect}\left(\frac{k_u}{k_{us}}\right) \tag{5.6}$$

To determine the extent of this region in the (ω, u)-domain requires the instantaneous position of the signal in (5.2) for any point in the (ω, k_u) domain; that is,

$$u_i(\omega, k_u) = \frac{-k_u x_0}{\sqrt{4k^2 - k_u^2}}. \tag{5.7}$$

The evaluation of this function along $k_u = \pm k_{us}/2$ is used to determine an ω-dependent along-track rect function that outlines the extent of the adequately sampled (ω, u)-domain; that is,

$$
\begin{aligned}
Ww_s(\omega, u) &\equiv \text{rect}\left(\frac{\omega - \omega_0}{2\pi f_s}\right) \cdot \text{rect}\left(\frac{u\sqrt{4k^2 - (k_{us}/2)^2}}{k_{us} x_0}\right) \\
&\approx \text{rect}\left(\frac{\omega - \omega_0}{2\pi f_s}\right) \cdot \text{rect}\left(\frac{2ku}{k_{us} x_0}\right).
\end{aligned}
\tag{5.8}
$$

Looking again at Fig. 5.2, consider a system which extracts the rectangular region $Ww_d(\omega, u)$ shown in Fig. 5.2(a) and samples at a fixed Doppler wavenumber corresponding to the rectangular region $WW_s(\omega, k_u)$ shown in Fig. 5.2(b), several observations can be made; the radiation pattern in the (ω, u) domain varies considerably across the transmitted bandwidth, the radiation pattern in the (ω, k_u) domain is independent of ω, the rectangular data block $Ww_d(\omega, u)$ contains Doppler wavenumbers that exceed the sampling wavenumber, and the amount of Doppler wavenumber aliasing (i.e., the triangle
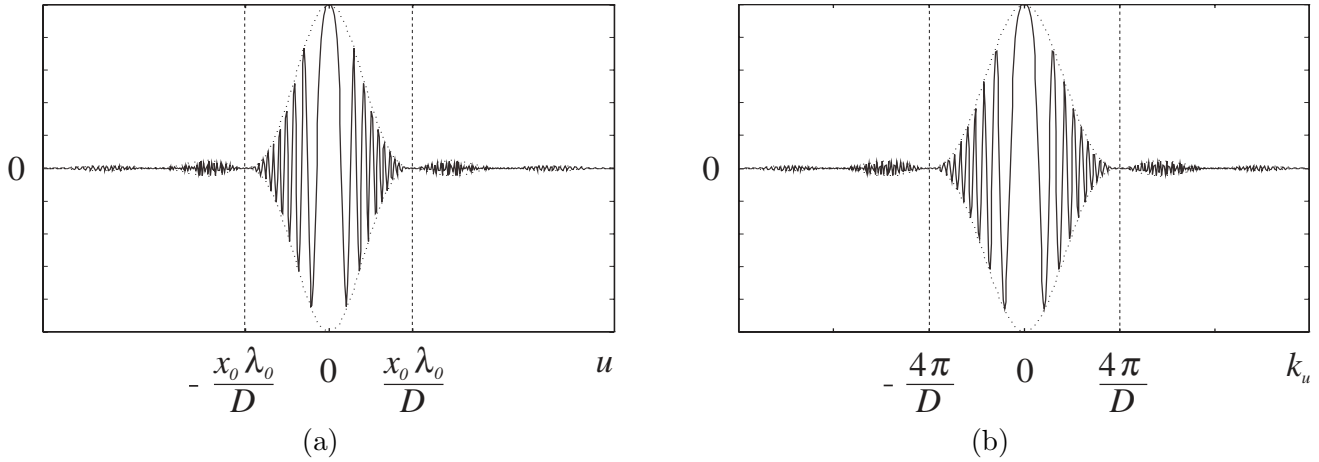
Figure 5.3 Cross-section of a well sampled ($\Delta_u \ll D/4$) point target response at the carrier frequency, (a) real $\{Ss_\delta(\omega_0, u)\}$, (b) real $\{SS_\delta(\omega_0, k_u)\}$. The scaling properties of the aperture radiation pattern can be seen in these figures, i.e., the amplitude response in both domains is sinc-like with an underlying linear FM-like signal structure.

regions of $WW_d(\omega, k_u)$ above and below the rectangular region of $WW_s(\omega, k_u)$ in Fig. 5.2(b)) changes with frequency ω.

Note that as we are only dealing with a single point target at a known location, it is possible to avoid Doppler wavenumber aliasing by setting all pixels outside the region $Ww_s(\omega, u)$ defined by (5.8) to zero before transforming into the k_u domain. However, this windowing operation does not apply in general when processing images with random target locations.

If the (ω, k_u)-domain signal in Fig. 5.2(b) is Stolt mapped, then it forms a half-ring shape on the positive k_x side of the wavenumber domain with an inner radius $2k_0 - 2\pi B_c/c$ and outer radius $2k_0 + 2\pi B_c/c$ (see for example Fig. 4.7(d)). The k_u dependent radiation pattern in the (ω, k_u)-domain becomes a k_y dependent pattern in the (k_x, k_y)-domain, so the radiation pattern is independent of k_x in the wavenumber domain.

Figure 5.3 shows the cross-section of the point target return at the carrier frequency of the pulse for the spatial domain and Doppler domain signal. Both signals are well sampled ($\Delta_u \ll D/4$) so minimal spatial aliasing occurs. In a situation where the signal is sampled at $\Delta_u = D/4$, the spatial signal in Fig. 5.3(a) appears practically unchanged, however, the signal in Fig. 5.3(b) repeats at spacings $k_u = 2\pi/\Delta_u$. This figure demonstrates the scaling effect of the aperture radiation pattern between domains. The radiation pattern is sinc-like in both domains, with the along-track scaling being controlled by the underlying phase function.

The $D/4$ along-track sampling requirement stated throughout this thesis (and the increase to $D/3$ detailed in Section 5.4.1.1) is somewhat unconventional with regard to much of the synthetic aperture literature (the exception to this observation is found in the work by Soumekh [144, 146, 147]). The

origin of the $D/2$ statement in the SAR literature is due in part to the erroneous analysis of Tomiyasu covered in Section 3.5. However, the most compelling reasons for employing a $D/2$ sample spacing is that; in most (but, not all) situations the resulting alias targets are not observed in the final focused imaged, and that there is usually a commercial requirement to map as fast as possible. A $D/4$ sample spacing requires that the imaging platform halve its speed, or that the swath width being imaged is halved (i.e., the PRF is increased), both of these options are undesirable in a commercial situation. A $D/3$ or $D/4$ sample spacing still results in some spatial aliasing, however, the aliased energy is from the sidelobes of the radiation pattern only. As most apertures are apodised (weighted) in some way, these sidelobes contain very little energy. A $D/2$ spacing results in aliasing of mainlobe energy, image dynamic range will always be compromised with such a sample spacing (though this compromise may often be acceptable when weighed against commercial requirements).

5.2 RANGE AND ALONG-TRACK AMBIGUITY CONSTRAINTS IN SYNTHETIC APERTURE SYSTEMS

The maximum along-track resolution a synthetic aperture system can achieve is given by $\delta y_{3\mathrm{dB}} \approx D/2$, this suggests that arbitrarily fine along-track resolution can be attained by reducing the aperture size. However, an arbitrarily small aperture has limited power handling capability and, more importantly, the pulsed nature of synthetic aperture systems places limits on the along-track sample spacing and the maximum possible range swath. The along-track spacing for constant along-track platform velocity is given by $\Delta_u = v_p \tau_{\mathrm{rep}}$, where $\tau_{\mathrm{rep}} = 1/\mathrm{PRF}$ is the pulse repetition period and the pulse repetition frequency (PRF) is the Doppler sampling frequency in Hertz. To uniquely distinguish the returns from a slant-range plane swath of width X_s (see Fig. 4.2 in Chapter 4), it is necessary to record all the target returns before transmitting the next pulse. Thus, the slant-range swath width is bounded by the *range ambiguity constraint* [31, p21]; that is,

$$X_s = R_{\max} - R_{\min} < \frac{c\tau_{\mathrm{rep}}}{2} = \frac{c}{2 \cdot \mathrm{PRF}}, \tag{5.9}$$

where $R_{\min}$ and $R_{\max}$ are the inner and outer edges of the swath in the slant-range plane, and c is the wave propagation speed. The *along-track ambiguity constraint* requiring along-track samples to be spaced $\Delta_u \leq D/(2\eta)$ (where $\eta = 1$ in the SAR literature and this thesis recommends $\eta = 1.5$ to 2), in conjunction with the range constraint, bounds the PRF by

$$\frac{2\eta v_p}{D} = \frac{\eta v_p}{\delta y_{3\mathrm{dB}}} \leq \mathrm{PRF} = \frac{v_p k_{us}}{2\pi} < \frac{c}{2X_s} \tag{5.10}$$

or alternatively,

$$\frac{2v_p}{c}X_s < \Delta_u \le \frac{D}{2\eta} = \frac{\delta y_{3\text{dB}}}{\eta}, \tag{5.11}$$

which requires that the swath width *decreases* as the along-track resolution *increases* [31, p21]. Thus, the real aperture length ultimately effects the *mapping rate* of the system, where mapping rate (in m^2/s) is defined as the ground plane swath width multiplied by the platform velocity:

$$\text{mapping rate} = v_p X_g = \frac{v_p X_s}{\sin\phi_i} \le \frac{Dc}{4\eta\sin\phi_i}, \tag{5.12}$$

where X_g is the ground-plane swath width and ϕ_i is the incidence angle (angle from nadir to boresight).

Using the argument that the real aperture width, W, must be such that the apertures elevation radiation pattern only illuminates the swath of interest in the ground-plane, the inequality in (5.11) also sets the following approximate lower bound on the aperture area (eg. see p21 of [31]):

$$\text{aperture area} = DW > \frac{8\eta\pi v_p r_0}{\omega_0}\tan\phi_i, \tag{5.13}$$

where r_0 is the mid-swath range, i.e., the stand-off range, in the slant-range plane.

The swath limiting relationship in (5.11) allows an informative comparison of spaceborne and airborne SAR to SAS. For example, a spaceborne SAR with an aperture of length 10m travelling at 7.5km/s can map a slant range swath of 50km ($X_s \approx Dc/(8v_p)$), an airborne SAR with an aperture of 2m travelling at 80m/s can map a swath of 940km, and a SAS with a 0.53m aperture travelling at 1m/s can map a 100m swath. The maximum swath widths predicted for spaceborne SAR and SAS are close to those imaged in practice, therefore these systems are limited by the along-track sampling requirements of the synthetic aperture. Airborne systems typically image swaths of 10km and operate with PRFs of around 1kHz, i.e., along-track samples are spaced $\Delta_u \approx 8\text{cm}$ apart which is much less than the $D/4$ requirement. Because of this, airborne SAR systems do not have problems with along-track undersampling. However, the along-track data rate is far higher than necessary for image reconstruction. Even though they sample a Doppler bandwidth much higher than the null-to-null bandwidth of the aperture (i.e., $k_{us} = 2\pi/\Delta_u \gg 8\pi/D$), it is still only practical to process the 3dB width of Doppler bandwidth of the aperture. To avoid unnecessary storage of samples, most airborne systems employ *presummers* or *prefilters* [16, 87]. Prefiltering consists of low-pass filtering the along-track signal (this substantially reduces alias target energy), followed by resampling of the low-pass along-track signal at a much lower sample rate. If multi-look processing is being employed, then each look can be individually generated using band pass filters spaced across the Doppler bandwidth being processed. Presumming is also used in practical spotlight systems where the patch diameter set by the real aperture is much larger than that

necessary for processing. The full Doppler bandwidth still needs to be sampled to avoid target aliasing. However, once sampled, the data can be presummed to a sampling rate appropriate for the patch size being imaged, again this results in significant reduction in processing time for real-time airborne spotlight SAR systems. The essential point to be gained from this comparison is that SAS systems suffer from similar along-track ambiguity constraints to those of spaceborne SAR systems. Spaceborne SAR systems have extra PRF constraints that are related to the interlacing of pulses due to the large offset range relative to the range swath, nadir echo avoidance, and blanking times due to the use of the same aperture for transmission and reception (eg. see p112-p119 of [47] and pp305-307 in [31]).

5.2.1   Along-track ambiguity to signal ratio (AASR)

Along-track sampling is always a tradeoff of aliased energy versus practical imaging requirements. The images produced using spaceborne SAR or SAS systems with along-track sample spacings of more than $D/4$ contain ambiguous targets due to aliasing of the *main lobe* energy in the Doppler wavenumber spectrum. The sampling constraint of $D/4$ can be relaxed slightly, as long as the processed Doppler bandwidth is reduced, eg. if $k_{us} = 6\pi/D$ ($k_u \in [-3\pi/D, 3\pi/D]$ and $\Delta_u = D/3$), then energy from the main lobe (which exists for $|k_u| \leq 4\pi/D$) aliases back into the sampled bandwidth, however, the bandwidth within $|k_u| \leq 2\pi/D$ only contains alias energy from the aperture sidelobes (which should be low due to apodisation of the real aperture), so processing the 3dB Doppler bandwidth should result in low alias target levels. The level to which the PRF can be reduced is usually determined using the *along-track ambiguity to signal ratio* (AASR). The AASR is defined as the ratio of the ambiguous target energy folded into the image processing bandwidth relative to the energy of the main signal within the processing bandwidth. This ratio is estimated using the following equation [31, p298]:

$$\text{AASR} \approx 10\log_{10}\left\{ \frac{\displaystyle\sum_{m=-\infty, m\neq 0}^{\infty} \int_{-B_p/2}^{B_p/2} A^2(k_u + mk_{us})dk_u}{\displaystyle\int_{-B_p/2}^{B_p/2} A^2(k_u)dk_u} \right\} \tag{5.14}$$

or, more generally, when a window/weighting function $W(\cdot)$ is applied over the processing bandwidth, the AASR is

$$\text{AASR} \approx 10\log_{10}\left\{ \frac{\displaystyle\sum_{m=-\infty, m\neq 0}^{\infty} \int_{-\infty}^{\infty} \left[W\left(\frac{k_u}{B_p}\right) A(k_u + mk_{us})dk_u \right]^2}{\displaystyle\int_{-\infty}^{\infty} \left[W\left(\frac{k_u}{B_p}\right) A(k_u)dk_u \right]^2} \right\} \tag{5.15}$$

where $A(k_u)$ is the combined aperture response, the integration is limited to visible wavenumbers $|k_u| \leq 2k$, the target reflectivity has been assumed uniform, and B_p is the processed Doppler bandwidth (in rad/m). In spaceborne SAR systems, the PRF and processed Doppler bandwidth are usually selected such that the AASR is -18 to -20dB [31, p296]. For example, the Seasat SAR operated with a 10.7m aperture, and took along-track samples every $\Delta_u = v_p/\text{PRF} = 7454/1647 = 4.53\text{m}$, if it is assumed that the aperture was not apodised in any way and the processed bandwidth is not weighted, then the AASR for 100% and 70% processed bandwidths is -17dB and -24dB. Figure 5.4 shows how the AASR is calculated for the 70% case. The figure shows the aperture function repeated in Doppler wavenumber space due to the along-track sampling. The darker shaded areas overlaying the lighter shaded processing bandwidth B_p, correspond to the regions of aliased target energy. Because the system samples at $D/2$, the predominant aliased energy in the *sampled* bandwidth is due to main lobe aliasing. By processing only 70% of this bandwidth, most of this main lobe alias energy is excluded. If the processed bandwidth is weighted by a Hamming window before processing then the AASR ratio drops to -24dB and -28dB for the two cases. The AASR for 100% processing drops the most when windowed due to the fact that the bulk of the ambiguous energy lies close to the Doppler folding wavenumber. When multilook processing is used to reduce the final image speckle (see Section 4.7), each look is extracted from a different section of the unweighted processing bandwidth. The looks extracted from the edges of the processing bandwidth contain the highest level of ambiguous energy. The coherent addition of these multilook images acts to suppress the ambiguous targets in the final lower resolution image. In summary, the use of reduced bandwidth processing trades off image resolution for image dynamic range (lower alias levels) and a relaxed along-track sampling requirement.

In scenes where the target field is composed of bright targets near bland targets, the approximation in (5.14) that the target reflectivity is uniform does not hold. For scenes containing, for example, bridges over lakes or rivers, an AASR of 10dB is not unusual and grating lobe targets are quite obvious in the SAR images. For examples of ambiguous targets in spaceborne SAR images see pp299-300 of Curlander and McDonough [31] and the paper by Li [98]. For examples of ambiguous targets in synthetic aperture sonar images see the work by Hayes and Gough [74, 75] and the work by Rolt [132, 133]. The reduced bandwidth processing discussed in the paper by Rolt [132] is essentially an investigation into the AASR for synthetic aperture sonar. By reducing the processed bandwidth, he reduces the level of ambiguous energy, hence the grating lobe targets get smaller.

5.2.2   Range ambiguity to signal ratio (RASR)

The problems encountered due to range ambiguities are different for each synthetic aperture system. In airborne SAR systems, range ambiguities are not considered significant [31, p303]. The large standoff range of spaceborne SAR systems means that they have several pulses in transit at any one time and
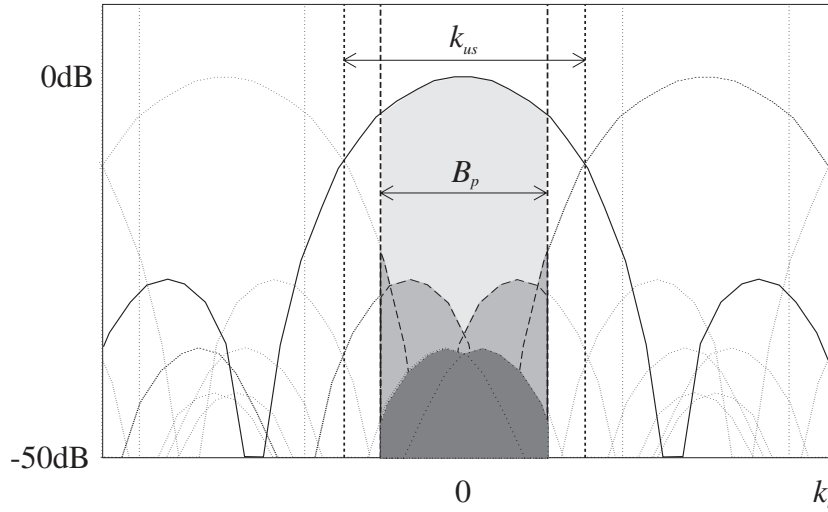
Figure 5.4   The along-track ambiguity signal to noise ratio (AASR). This figure corresponds to a system that samples in along-track every $\Delta_u = D/2$. The sampled bandwidth, $k_{us} = 4\pi/D$, spans only the 3dB width of the aperture radiation pattern, so main lobe and sidelobe energy is aliased (darker grey areas). By processing 70% of the sampled bandwidth, $B_p = 0.7k_{us}$ (light grey area), the majority of the aliased main lobe energy is excluded, and the level of alias targets is reduced in the final image estimate. When a window function is applied across the processing bandwidth, this alias level is reduced further still. Reduced bandwidth processing trades off image resolution for image dynamic range and a relaxed along-track sampling requirement.

the width of the apertures employed on spaceborne SARs are designed so that they only illuminate the range swath of interest. Range ambiguities then occur due to interference from previous and successive pulse returns off targets that lie outside of the swath illuminated by the 3dB main lobe of the apertures *elevation* radiation pattern (see for example Fig. 6.26 on p297 in [31]). The RASR is defined on p304 of Curlander and McDonough [31] as the maximum ratio of the ambiguous power received from targets outside the swath to the power received from targets inside the desired swath.

Synthetic aperture sonar systems typically use low grazing angles, so it is difficult to operate the sonar such that it insonifies only the range of interest. The varying radiation patterns of the apertures in systems that employ wide bandwidth, low-Q signals also makes limited swath insonification a difficult task. Short range sonars operating in shallow waters do not interlace pulses and have a maximum range given by $R_{\max} = c\tau_{\text{rep}}/2$, range ambiguities are then only experienced from the returns of *previous* pulses. For example, if a target was detected at range x_0 in pulse i, there is a chance that it came from a reflection of the previous pulse $(i-1)$ from a target located at range $R_{\max} + x_0$. Two factors operate to suppress this ambiguous target; in an operational sonar time-varying gain (TVG) circuits account for two-way *amplitude* spreading $1/x^2$ losses for target ranges $x \in [R_{\min}, R_{\max}]$, and synthetic aperture focusing only correctly focuses the target located at ranges $x \in [R_{\min}, R_{\max}]$. To examine the effects of TVG and focusing, this time consider two targets of equal reflectivity; one located at range x_0, the other at range $R_{\max} + x_0$. The signal received from the target at x_0 has had its signal strength reduced

by $1/x_0^2$ (i.e., signal power is reduced by $1/x_0^4$). This reduction is offset by the TVG. However, the ambiguous targets signal strength after TVG is proportional to $x_0^2/(R_{\max} + x_0)^2$, which for targets at say $x_0 = 50$m and $R_{\max}+x_0 = 250$m is a difference of 28dB. The locus for the ambiguous target is spread over a significant area in along-track and range, and it is not matched to the focusing parameters for a target lying at x_0, so there is little or no improvement in the ambiguous target's strength after focusing. Similar observations were made for range ambiguities in the ACID project [93]. Since spreading losses reduce the possible level of range ambiguities far below the levels expected from along-track ambiguities, they are not considered a problem and are no longer discussed in this thesis.

5.3   GRATING LOBE TARGETS IN SYNTHETIC APERTURE IMAGES

To determine methods for observing, suppressing, and removing ambiguous targets in synthetic aperture images, it is first necessary to understand the properties of the aliased along-track signals and the modifications required in the inversion schemes. To investigate the effects of along-track undersampling, consider the along-track sampled point target response:

$$ss_s(t, u) = ss_\delta(t, u) \cdot \sum_m \delta(u - m\Delta_u) \tag{5.16}$$

where $\Delta_u = v_p \tau_{\text{rep}}$ is the along-track sample spacing. It is not necessary to consider the range sampling requirement as it is satisfied in the receiver. Transforming to the (ω, u)-domain gives:

$$Ss_s(\omega, u) = Ss_\delta(\omega, u) \cdot \sum_m \delta(u - m\Delta_u) \tag{5.17}$$

Transforming this signal to the (ω, k_u)-domain gives:

$$
\begin{aligned}
SS_s(\omega, k_u) &= SS_\delta(\omega, k_u) \odot_\omega \frac{2\pi}{\Delta_u} \cdot \sum_n \delta\left(k_u - \frac{2\pi n}{\Delta_u}\right) \\
&= \frac{2\pi}{\Delta_u} \cdot \sum_n SS_\delta\left(\omega, k_u - \frac{2\pi n}{\Delta_u}\right)
\end{aligned}
\tag{5.18}
$$

The sampled signal in the (ω, k_u)-domain or the (t, k_u)-domain is composed of scaled, coherently interfering, repeated versions of the continuous signal $SS_\delta(\omega, k_u)$. Figure 5.5(a) shows an example of these repeated along-track spectra for the carrier temporal frequency. These repeated spectra coherently interfere to produce the corrupted (aliased) signal shown in Fig. 5.5(b). Figure 5.5(c) shows the range-Doppler domain signal for the sampled point target. The figures in Figure 5.5 were generated by a system with an along-track sample spacing of $3D/4$, i.e. $k_{us} = 8\pi/(3D)$.

In undersampled systems, it is still possible to achieve an along-track resolution of $D/2$. To achieve
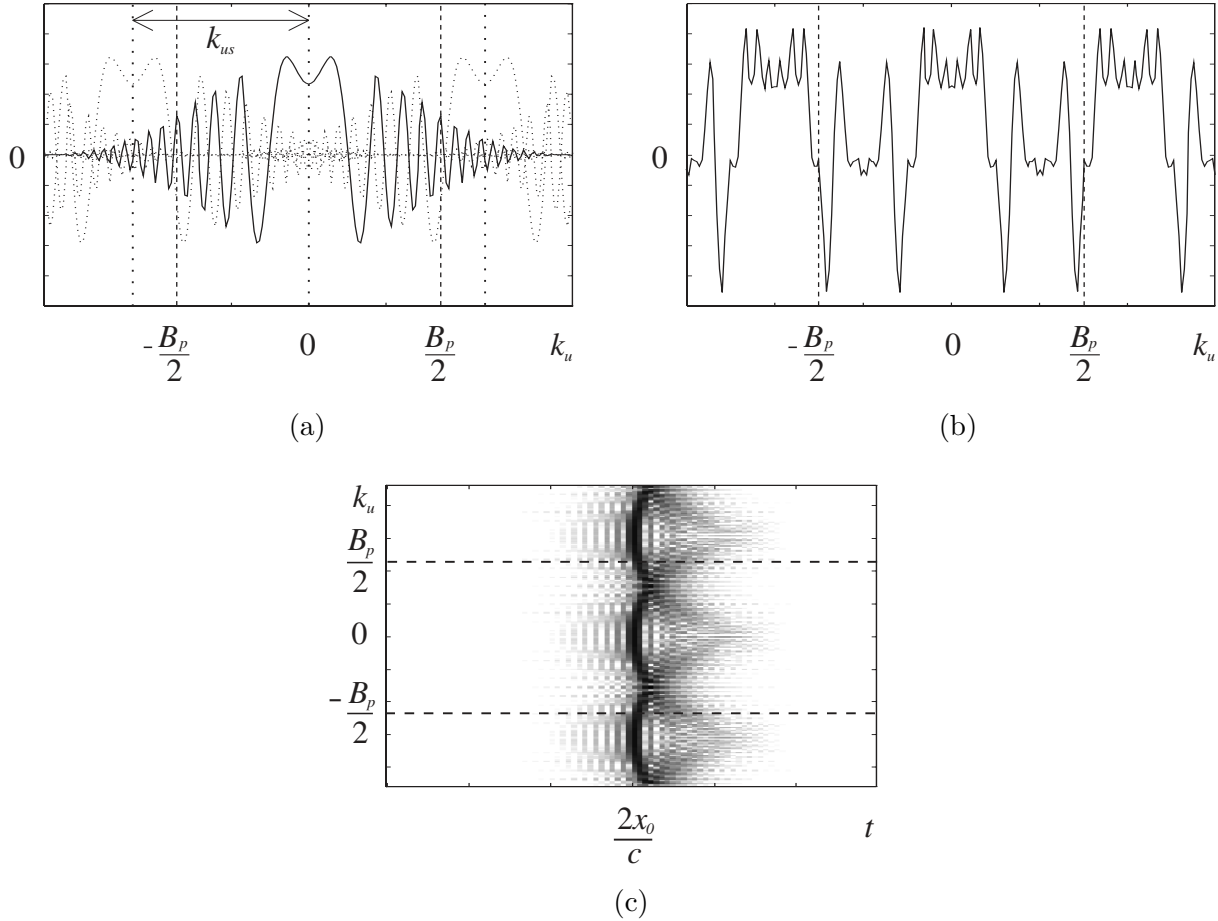
Figure 5.5   The effects of undersampling in the along-track direction, (a) repeated versions of real $\{SS_\delta(\omega_0, k_u)\}$, (b) resulting undersampled signal spectrum, (c) undersampled signal in the range-Doppler domain (logarithmic greyscale with 30dB dynamic range).

this resolution it is necessary to process the 3dB bandwidth of the aperture with *correctly sampled* filter functions. A physical interpretation in terms of an array gives some insight into how this operation is performed. Consider an undersampled synthetic aperture as a thinned array of length $L_{\mathrm{sa}} = x_0\lambda/D$. When this array is focused, it has an ultimate resolution of $D/2$. Because this focusing is performed digitally, it is necessary to focus onto a grid with an along-track spacing that meets, or exceeds the ultimate resolution limit of the array. To meet this requirement, the along-track spacings of the processing grid are set such that the raw data corresponds to some of the along-track spacings, while the others are left blank. This operation can be interpreted as placing new elements into the thinned array, thus producing an adequately sampled array, albeit with 'deaf' or 'blind' new elements. Each pixel in this resampled array is then focused with the correctly sampled delay-and-sum beamforming operations.

An equivalent interpretation can be given from a Fourier transform point of view. A correctly

sampled filter implies a filter with a spatial frequency (Doppler wavenumber) response that spans the processing bandwidth B_p. In undersampled systems that have sample spacings greater than $D/2$, the full processing bandwidth is not available. Fourier transformation of this type of undersampled along-track signal via the FFT produces only one section of the repeated k_u-space signal (see the discussion in Section 2.1). Because the required processing bandwidth *exceeds* the sampled bandwidth, i.e., $B_p > k_{us}$, the significant aliasing of the main lobe energy that has occurred results in a high AASR. To obtain the required repeated sections of k_u-space it is possible to either; repeat the FFT samples until the number of repeated Doppler wavenumber segments exceeds the processing bandwidth, or the equivalent spatial domain operation is interlacing vectors of zeros in along-track until the sample spacings are less than $D/2$ (i.e., insert blind elements into the array). To generate the range-Doppler signal shown in Fig. 5.5, the Fourier space was repeated three times by interlacing two vectors of zeros between every valid range line vector.

5.4   TECHNIQUES FOR GRATING LOBE TARGET SUPPRESSION AND AVOIDANCE

An operational SAS system that is designed to achieve high-along track resolution (less than 50cm) over a wide swath (hundreds of meters) is severely limited by the limitations of towing vessels. To traverse the synthetic aperture with a straight tow path typically requires a tow speed in excess of that allowed by the relationship in (5.11). To avoid or reduce the effects of target aliasing the following techniques can be employed. In some cases these techniques may also be applicable to spaceborne SAR systems, however, due to the physical limitations of spaceborne SAR systems, some remain unique to SAS.

5.4.1   Wide bandwidth, low-Q waveforms

The transmission of wide bandwidth signals has been seen as a method for increasing the mapping rate in SAS systems. This increase in mapping rate causes spatial undersampling, so it comes at the expense of decreased dynamic range in the final image estimate. The image dynamic range is decreased due to the fact that the location of the grating lobe targets about an actual target in an image produced from undersampled data is frequency dependent. The transmission of a wide bandwidth signal causes these frequency dependent grating lobe targets to smear into a continuum. This section shows the effects of undersampling on such wide bandwidth, low carrier frequency-to-bandwidth ratio (low-Q), systems and derives the peak to grating lobe ratio (PGLR) as a method for estimating the dynamic range in the images produced by undersampled systems.

Few practical investigations of wide bandwidth SAS systems had been performed prior to the development of the Kiwi-SAS discussed in this thesis. During the development of this system, the low

frequency, low-Q, 'ACoustical Imaging and Development' (ACID) project was discovered to be underway (the ACID system is now referred to as the SAMI (Synthetic Aperture Mapping and Imaging) system). Other narrow bandwidth SAS systems were also being investigated. These systems are reviewed in Section 1.6.2.

The first wide bandwidth synthetic aperture sonar developed was the University of Canterbury's previous SAS system. The development and results obtained using this system are covered in the thesis by Hayes [75] and the papers by Hayes and Gough [62, 63, 74]. A simulation study, based in part on this system, is covered in Rolt's thesis and paper [132, 133]. The results of simulation studies performed as part of the ACID/SAMI project are presented in the papers by Zakharia and Chatillon [22, 23, 167], while results from the operational ACID/SAMI system are covered in the papers by Adams, Riyait, and Lawlor [2, 93, 130, 131].

Though all of the references dealing with the practical implementation of a SAS processor note that the synthetic aperture technique is a mature technique in the field of radar [2, 22, 23, 63, 74, 75, 131–133, 167], only one of the references found by this author actually implemented one of the commonly used SAR processing algorithms for a sonar application. This was the application of the wavenumber algorithm to a theoretical 15kHz bandwidth, 150kHz carrier frequency system by Huxtable and Geyer in 1993 [78]. The synthetic aperture processors used in previous practical SAS investigations were all based on spatial-temporal domain delay-and-sum beamformers or 2-D frequency-domain fast-frequency correlators. Both of these processors are covered in Section 4.3.1 and it is noted that this form of processing is inefficient due to the fact that each pixel in the final image is processed on an individual basis, as opposed to the block processing methods used in the common SAR algorithms; range-Doppler, wavenumber, and chirp scaling. The delay-and-sum beamforming processor was chosen on the basis that SAR algorithms were narrow band or that time-domain processing was exact. Both of these assumptions are not quite correct. Though it is necessary to account for extra modulation functions in a wide bandwidth application of the commonly used SAR algorithms (see Chapter 4), so that images can be calibrated correctly, the SAR algorithms in their original formulation would still have produced accurately focused images (though the sidelobes in these images would not have been at uniform or easily predictable levels). For a delay-and-sum beamformer to produce accurately focused images, the required delays must be calculated via interpolations or via frequency domain phase shifts. If an interpolator is used, then the algorithm is not longer exact. If 2-D frequency fast-correlation is used, then the algorithm is still exact, but it is extremely inefficient. An alternative to interpolation has been used in the ACID/SAMI beamformer. The ACID/SAMI system obtains the required delays by oversampling the range signal [2], thus the amount of data that needs to be processed to form the final image is substantially increased (resulting in an inefficient processor) and, even then, the *exact* delay is not being used, only the nearest sample to the required delay.

Previous investigations of wide bandwidth systems experienced difficulty in analysing and simulating

the effects of the radiation patterns generated during the transmission and reception of a wide bandwidth signal. Through intuitive arguments, most of the previous investigators noted that the overall effect of the wide bandwidth radiation pattern on an individual target was an ensemble of the individual radiation patterns at each frequency, however, none of these investigators gave a clear mathematical formulation of the effect. Chapter 3 presents the theory of wide bandwidth aperture theory and in Chapter 4, this theory is applied to synthetic aperture systems.

The range resolution in a synthetic aperture image is dependent on the bandwidth of the transmitted waveform via $\delta x_{3\text{dB}} \approx c/(2B_c)$ and the along-track resolution is dependent only on the aperture length via $\delta y_{3\text{dB}} \approx D/2$, so what parameters are related to the instantaneous transmitted frequency within the transmitted pulse? In SAS, frequency dependent attenuation means that better SNR is achieved at lower frequencies (see pp96-104 in Urick [156]). The derivation of the wavenumber algorithm in Chapter 4 showed that the along-track *improvement factor* due to along-track focusing, i.e., the space-bandwidth product of the the along-track chirp for a target at range x, is

$$\text{IF}_{\text{sa}} = K_a L_{\text{sa}}^2 = \frac{2x\lambda}{D^2} \tag{5.19}$$

The dependence of the synthetic aperture length L_{sa}, and hence the improvement factor, on frequency means that longer apertures are generated for lower frequencies, yielding higher improvement factors, thus higher SNR. The use of lower frequencies is encouraged by both attenuation and improvement factor dependence on frequency. The downside of the longer apertures comes if the coherency time of the medium is short, however, if autofocus methods are successful, then longer apertures produce more samples for phase error estimation. The use of wide bandwidth signals also allows for a wider spectral classification of sea floors, different compositions of materials on the sea floor have scattering parameters that depend on aspect angle and transmitted frequency. Many targets encountered in practice have some sort of aspect dependence so that when imaging with a narrow beam system these targets can be missed entirely. The use of wide bandwidths and wide beam widths (small apertures) in the synthetic aperture system leads to very high-resolution synthetic aperture images containing a wealth of target information.

In undersampled synthetic apertures, the location of grating lobe targets is also dependent on the transmitted frequency. Therefore it is critical to chose an appropriate band of frequencies for the required range resolution that also optimizes the suppression of grating lobe targets. The location of the ith grating lobe target in the along-track direction is determined by the wavenumber folding frequency. The location of the grating lobe of a point-like target at $(x, y) = (x_0, 0)$ in the along-track dimension of the (ω, u)-domain is $u = i\Delta_g$ where

$$\Delta_g(\omega) = \frac{2x_0(\pi/\Delta_u)}{\sqrt{4k^2 - (\pi/\Delta_u)^2}} \approx \frac{\pi x_0}{k\Delta_u} = \frac{x_0\lambda}{2\Delta_u} \tag{5.20}$$

The migration of the grating lobe with frequency allows wide bandwidth systems to *smear* the effects of the grating lobe targets. This smearing effect was first suggested by de Heering in 1984 for use in synthetic aperture sonar systems [38]. However, de Heering incorrectly stated that the along-track ambiguity constraint was a narrowband assumption and that wide bandwidth systems could *remove* spatial ambiguities [38, p278]. The along-track ambiguity criteria is a spatial sampling constraint that is *independent* of frequency. If this constraint is exceeded, wide bandwidth signals only act to smear the grating lobes produced. This smeared energy introduces a form of self-clutter which limits the final image dynamic range.

To ensure adequate smearing of the grating lobes, de Heering proposed that if the transmitted signal bandwidth is chosen such that the along-track location of the first grating lobe at the lowest frequency equals the along-track location of the second grating lobe at the highest frequency then the grating targets smear into a continuum. Solving for this case using Δ_g, it is found that the highest frequency is twice the lowest frequency, i.e., an octave of frequency is required. This gives a system *quality factor* $Q = 1.5$ (not 0.6 as stated in de Heering's paper [38] or $\sqrt{2}$ as stated Hayes' thesis [75, p48]).

The level of the smeared grating lobe energy in the images produced using wide bandwidth synthetic aperture systems is best determined from the analysis of a point target response. Figure 5.6 shows the processing domains of a point target that has been illuminated/insonified with an octave transmitted signal, and has been sampled at along-track spacings of $\Delta_u = 3D/4$ (the simulation parameters used in this chapter are those of the Kiwi-SAS given in Table 7.4). The raw range-Doppler data for this signal is shown in Fig. 5.5(c). Figure 5.6(a) and (c) clarify the definition of two terms appropriate for describing the response of the alias targets in synthetic aperture images; grating lobe targets and grating lobe energy. The term *grating lobe target* is appropriate for describing the peaks either side of the target response in along-track y at a given range wavenumber $k_x = \sqrt{4k^2 - k_u^2} - 2k_0 \approx 2(k - k_0)$. Figure 5.6(c) clearly shows the frequency dependent migration of the first grating lobe target in the (k_x, y)-domain. The image has been displayed as the real part of the signal to emphasize the rotating phase of the grating lobe targets relative to the non-rotating response of the actual target along $y = 0$ (note that the time and range origins have been shifted to the target location x_0 so that the spectra can be calculated via the FFT, for more details see Section 2.3 and note that $r_0 = x_0$ for Fig. 5.6). The width of the actual target in Fig 5.6(c) along $y = 0$ is $D/2$, while the width of the grating lobe targets is close to D. The *grating lobe energy* is the smeared energy shown either side of the target response in Fig. 5.6(a). This energy smears in along-track due to the frequency dependence of the grating target locations in Fig. 5.6(c), while the smearing in range occurs due to the reduced k_x bandwidth of any $y =$ constant slice through the migrating grating targets. It is the maximum level of this grating lobe energy that determines the loss in image dynamic range in undersampled systems. The location of this maximum is best shown in Fig 5.6(b), the phase matching of the data in Fig. 5.5(c) has smeared the ambiguous energy of the upper and lower repeated spectra due to mismatch in the phase matching
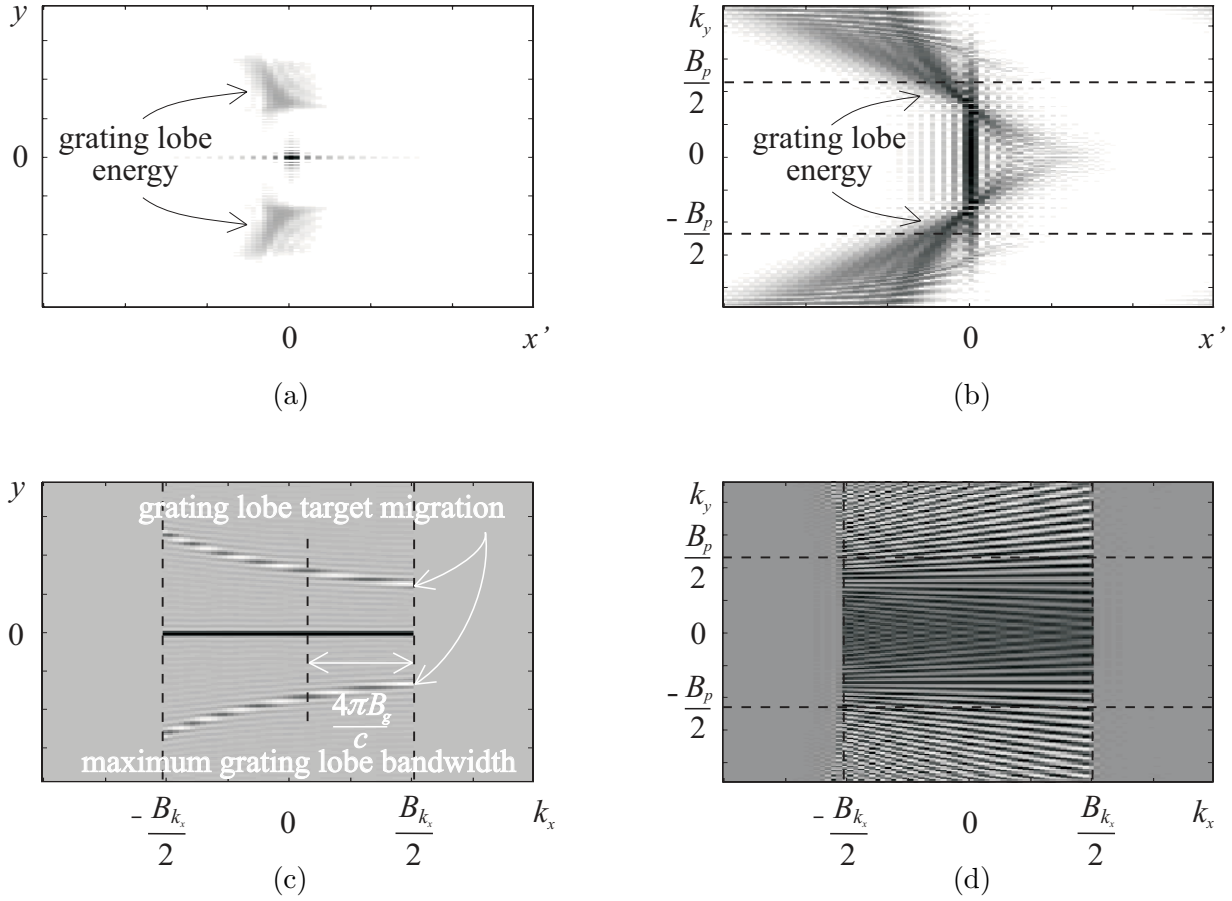
Figure 5.6 Signal domains of an undersampled point target, (a) final image estimate, $|ff_\delta(x', y)|$, where the x' origin is located at $x = x_0$, (b) (x', k_y)-domain (range-Doppler), (c) (k_x, y)-domain, (d) (k_x, k_y)-domain. Images (a) and (c) have been windowed so that only energy within the Doppler processing bandwidth $B_p = 4\pi/D$ and the range wavenumber bandwidth $B_{k_x} \approx 4\pi B_c/c$ is processed. Images (b) and (d) show the Doppler signal prior to windowing. To emphasize the smeared grating lobe energy, images (a) and (b) are displayed as logarithmic greyscale with 30dB dynamic range. To emphasize the phase rotations in the undersampled data, images (c) and (d) show only the real components of the signals. Distortion of the wavenumber domain in (d) is due to the data remapping. Each inversion algorithm produces this effect. Note that, other than to deconvolve the radiation pattern effects, no range or along-track weightings have been applied to reduce sidelobes.

filter parameters, i.e., too much range migration correction has been applied to the ambiguous target energy. The smeared energy shown above and below the processing bandwidth is not processed, it is simply shown to emphasize the effect of the filter mismatch on the ambiguous targets. The point at which this ambiguous target energy is a maximum is the point at which the energy is most concentrated in range, i.e., the point at which the k_x-bandwidth is a maximum. With reference to Fig 5.6(b), this point is seen to be located in line with the target response at range x_0, i.e., $x' = 0$, at the folding Doppler wavenumber, $k_u = k_y = \pi/\Delta_u$, of the original undersampled data. To determine the along-track location and value of this peak, consider the response in Fig 5.6(c). The point at which the alias

target response has the widest bandwidth is at the highest k_x values where the grating lobe migration is changing the slowest. Consider the inverse Fourier transform of this (k_x, y)-domain signal at $x' = 0$, i.e. the response in along-track through the target maximum:

$$
\begin{aligned}
ff_\delta(0, y) &= \mathcal{F}_{k_x}^{-1} \left\{ Ff_\delta'(k_x, y) \right\} \big|_{x'=0} \\
&= \int_{-B_{k_x}/2}^{B_{k_x}/2} Ff_\delta'(k_x, y) dk_x
\end{aligned}
\tag{5.21}
$$

from this equation it can be seen that the target peak at $(x', y) = (0, 0)$ is the sum (integral) of $Ff_\delta'(k_x, y)$ along k_x. Since this spectrum is of uniform height, the peak in the (x', y)-domain is increased by the factor $B_{k_x} \approx 4\pi B_c/c$ (pulse compression scaled to spatial coordinates). Similarly, the maximum grating lobe migration bandwidth can be determined from the criteria that the along-track location of the grating lobe peak at the maximum frequency must not migrate any more than the width of the aperture D, this gives the maximum grating lobe bandwidth (in Hertz) of

$$
B_g = f_{\max} - \left(\frac{2\Delta_u D}{x_0 c} + \frac{1}{f_{\max}} \right)^{-1}
\tag{5.22}
$$

where $f_{\max} = f_0 + B_c/2$ is the maximum frequency in the transmitted signal. Relative to the level of the grating lobe target migration, the peak of the maximum grating lobe target energy is increased by the factor $4\pi B_g/c$ (grating lobe target compression). If the radiation pattern is not deconvolved across the processed Doppler bandwidth, the amplitude of the grating lobe targets relative to the main target peak in the (k_x, y)-domain are approximately bounded by the function

$$
A_g(k_x, x_0, y) = \exp\left[-2\left(\frac{kD}{2\pi} \cdot \frac{y}{x_0} \right)^2 \right]
\tag{5.23}
$$

where $k \approx k_0 + k_x/2$. Evaluating (5.23) at the first grating target location $y = \Delta_g$ gives a predicted grating target level of $\exp[-2(D/(2\Delta_u))^2]$. Thus, the image dynamic range is limited by the *peak to grating lobe ratio* (PGLR) which is given in decibels as

$$
\text{PGLR} = 20 \log_{10} \left\{ \exp\left[-2\left(\frac{D}{2\Delta_u} \right)^2 \right] \cdot \frac{B_g}{B_c} \right\}
\tag{5.24}
$$

Figure 5.7 shows the y cross-section through $x' = 0$ of a point target generated in a similar fashion to Fig 5.6(a), and the PGLR as predicted by (5.24). Also shown on the figure is the level of bandwidth smear, L_{BWS}, predicted by the wide bandwidth grating target analysis in Rolt's thesis [133, p104]; that
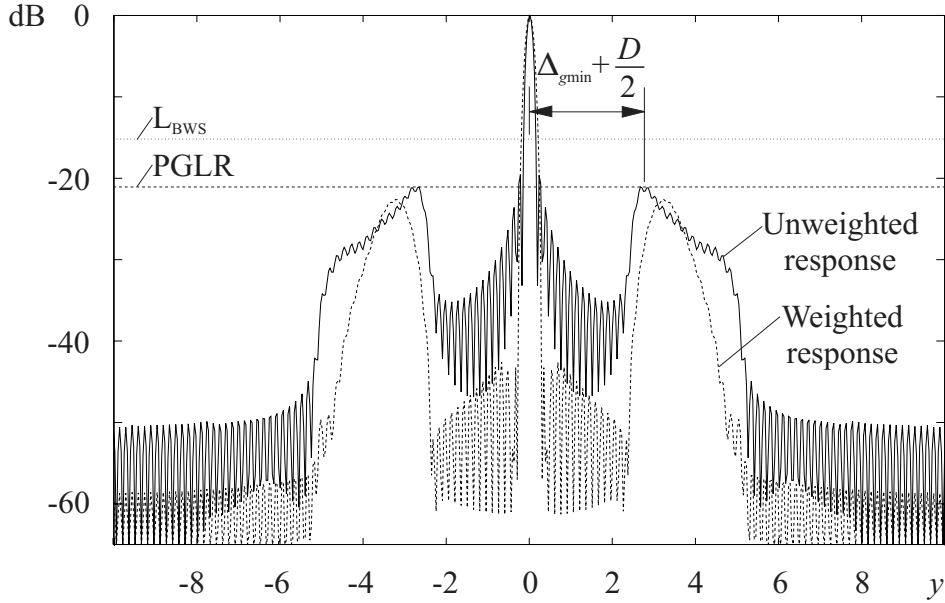
Figure 5.7   Along-track cross-section of an undersampled point target response. The target was generated using the Kiwi-SAS parameters in Table 7.4 for a target at $x_0 = r_0 = 30$m at an along-track sample spacing of $\Delta_u = 3D/4$. The location of the maximum of the grating lobe energy is $\Delta_{g\min} + D/2 = \Delta_g(\omega_{\max}) + D/2$ away from the main lobe, while the height of the grating lobe relative to the main lobe is accurately predicted by the *peak to grating lobe ratio* (PGLR). The grating lobe level predicted by the level of bandwidth smear, L_{BWS}, from Rolt's thesis is shown for comparison. The unweighted response was obtained from a spectral estimate that had the frequency dependant modulation removed, but still had the weighting of the radiation pattern over the along-track processing bandwidth. The weighted response had the along-track weighting deconvolved, followed by 2-D Hamming weighting. See the text for more details.

is,

$$L_{BWS} = 10 \log_{10} \left(\frac{D\Delta_u f_{\min} f_{\max}}{2x_0 c B_c} \right) \tag{5.25}$$

(Note that Rolt's thesis states this equation in dB magnitude, as opposed to dB power as stated here). Figure 5.7 also shows the along-track cross-section of a point target processed with Hamming weighting in both temporal frequency and Doppler wavenumber. A comparison of the two target cross-sections shows that spectral weighting only mildly reduces the level of the grating lobe energy. The main features of Figure 5.7 and others like it are used in the next section to analyse the along-track sampling requirements of synthetic aperture systems.

5.4.1.1   Resolving the D/2 versus D/4 sampling argument

It has been mentioned several times throughout this thesis that the SAR community recommends a $D/2$ sample spacing in the along-track direction of the synthetic aperture. In Section 3.5 and in Chapter 4 it is shown that, if it is assumed that it is necessary to adequately sample the along-track wavenumber
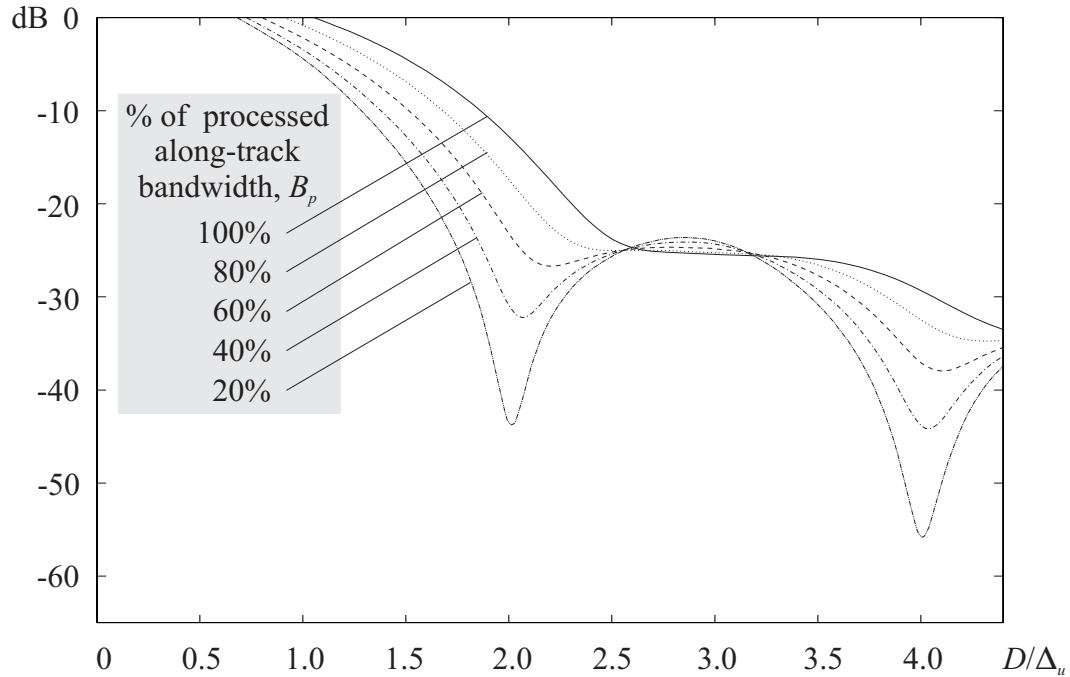
Figure 5.8 The effect of reducing the processing bandwidth on the along-track ambiguity to signal ratio (AASR) for a given number of samples per aperture length (D/Δ_u). The nominal processed bandwidth $B_p = 4\pi/D$ results in an image estimate with $D/2$ along-track resolution, a reduction in the processing bandwidth causes a reduction in the along-track resolution of the final image estimate. See the text for more details.

signal within the null-to-null width of the radiation pattern in wavenumber space, then a $D/4$ sample spacing is required. An image with a final along-track resolution of $D/2$ is obtained by processing the along-track wavenumber bandwidth that lies within the 3dB bandwidth of the radiation pattern, i.e., within the processing bandwidth of width $B_p = 4\pi/D$ (rad/m). Because of this bandwidth limiting or filtering operation, it is possible to increase the along-track sample spacing to the order of $D/3$ without introducing substantial grating lobe targets into the final image estimate. A increase in the sample spacing to $D/3$ causes aliasing of the mainlobe of the radiation pattern, however, this aliased energy does not fall into the processed bandwidth B_p (a low level of alias energy from the radiation pattern sidelobes still enters the processing bandwidth). Using the AASR and the PGLR, this section shows how the sample spacing impacts the dynamic range of synthetic aperture images.

Figure 5.8 shows the along-track ambiguity to signal ratio (AASR) for an aperture with a constant illumination function for different sample spacings and different processed bandwidths (this plot is valid for an arbitrary synthetic aperture system). The AASR is plotted versus the *number of samples per aperture length*, D/Δ_u, eg. if the sample spacing is $\Delta_u = D/2$, there are 2 samples per aperture length. This ordinate can also be scaled to along-track sample rate via $k_{us} = 2\pi/D \cdot D/\Delta_u$. This form of figure was used in the reference by Mehlis [103] to investigate the sampling requirements of a synthetic

aperture radar system. In Fig. 5.8, note that at a sample spacing of $D/2$ that as the processing bandwidth is *decreased*, the AASR also decreases. At a sample spacing of $D/2$, mainlobe energy is aliased into the sampled bandwidth. If $D/2$ image resolution is required in the final image estimate (i.e., $B_p = 4\pi/D$), then the sampled bandwidth *equals* the processed bandwidth and all of the aliased mainlobe energy is passed to the image processor. By reducing the processed bandwidth relative to the sampled bandwidth, less alias energy is processed, so the AASR decreases. (Similar features occur at $D/\Delta_u = 4$ due to aliasing of energy from the first sidelobe of the radiation pattern). At a sample spacing of $D/3$ the mainlobe energy does not get aliased into the processing bandwidth, so all of the curves are at similar levels.

Though the AASR indicates how much *energy* aliases into the processed bandwidth, it does not indicate how this alias energy affects the final image dynamic range. The peak to grating lobe ratio (PGLR) is a more useful figure of merit for determining the loss of image dynamic range due to alaising. Figure 5.9 shows the variation of the PGLR with respect to; sample spacing, target range, carrier frequency, bandwidth, and spectral windowing. Figures 5.9(a) and (b) were generated using the parameters of the previous University of Canterbury's SAS system; $f_0 = 22.5$kHz, $B_c = 15$kHz, $D = 0.3$m, for targets at $r_0 = 30$m and 60m respectively. Figures 5.9(c) and (d) were generated using the Kiwi-SAS parameters; $f_0 = 30$kHz, $B_c = 20$kHz, $D = 0.3$m, for targets at $r_0 = 30$m and 60m respectively. The images used to produce Figs. 5.9(a)-(d) were limited to $B_p = 4\pi/D$ in along-track wavenumber, the 3dB sinc^2-weighting of the radiation pattern was not removed, and the $\sqrt{1/k}$-weighting in range was deconvolved. Figures 5.9(e) and (f) were generated using the Kiwi-SAS parameters. The images used to produce Figs. 5.9(e) and (f) were limited to $B_p = 4\pi/D$ in along-track wavenumber, the 3dB sinc^2-weighting of the radiation pattern was deconvolved, the $\sqrt{1/k}$-weighting in range was deconvolved, and a 2-D Hamming weighting was applied to the processed range and along-track bandwidths. Plotted in each of the graphs in Fig. 5.9 is; the grating lobe level measured from simulated images, the PGLR as predicted by (5.24), the L$_{\text{BWS}}$ as predicted by (5.25), and the maximum sidelobe response at the grating lobe location.

The most interesting aspect of Fig. 5.9 is that in every figure, at a sample spacing of $D/2$ (i.e., $D/\Delta_u = 2$), the smeared grating lobe energy due to the first grating lobe target is above the maximum sidelobe level. It is only when the along-track sample rate is greater than $D/\Delta_u = 2.5$ to 3 that the grating lobe level drops beneath the sidelobe level. Figure 5.9 allows one to determine the loss of dynamic range for a given sample spacing, or it allows one to determine the minimum sample spacing for a desired dynamic range. For example, if 30dB dynamic range was desired in a Kiwi-SAS image, and spectral windowing was to be used, then Figs. 5.9(e) and (f) show that a sample spacing of $D/2$ would be adequate—grating lobe targets will exist in the image estimate, but they will be below the required image dynamic range.

Other observations made using Fig. 5.9 are; the PGLR gives a reasonable estimate of the image
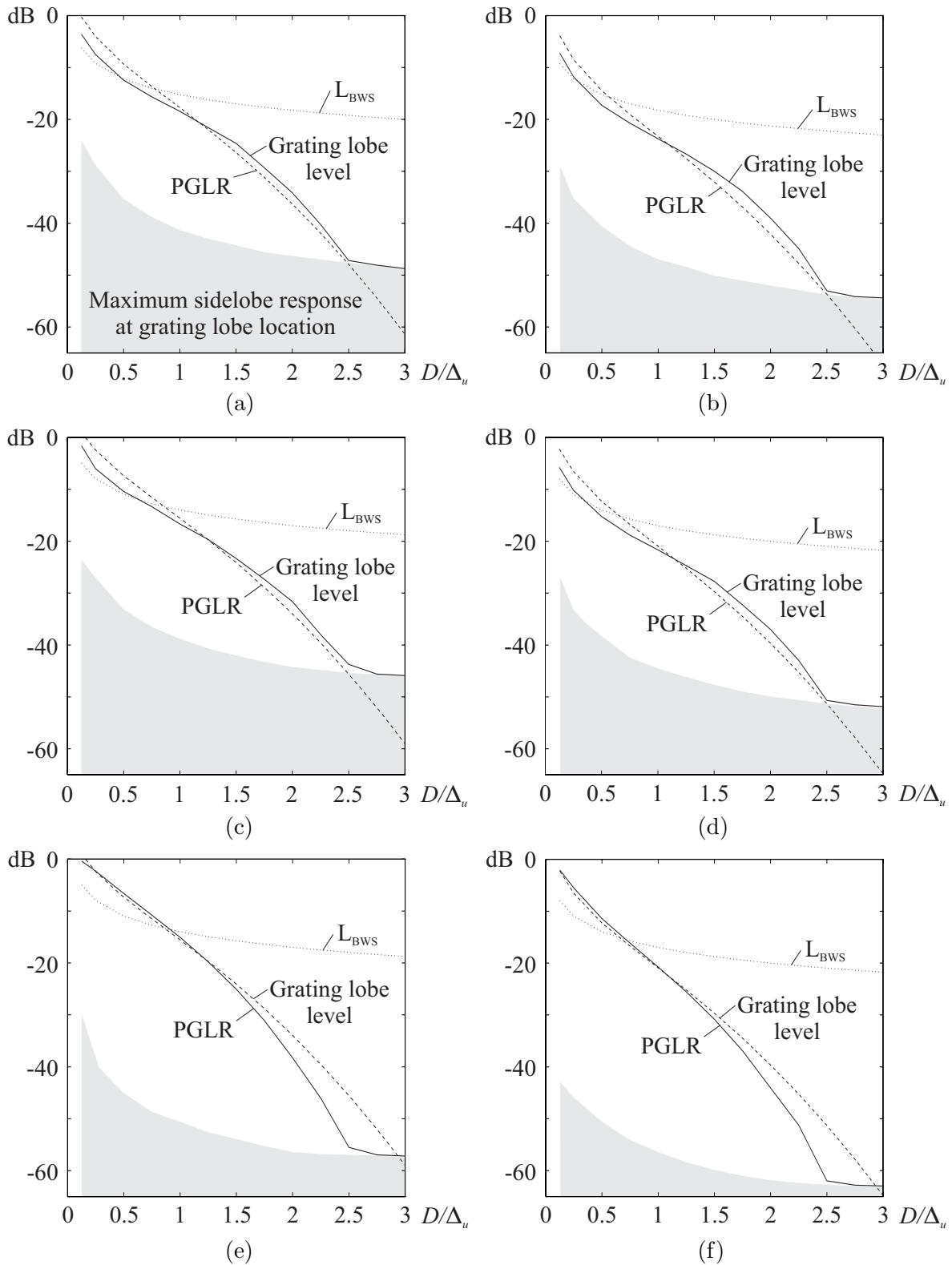
Figure 5.9 Using the PGLR to determine the optimum sample spacing. (a) and (b) were generated using the previous Canterbury SAS parameters; $f_0 = 22.5$kHz, $B_c = 15$kHz, $D = 0.3$m, for targets at $r_0 = 30$m and 60m respectively. (c) and (d) were generated using the Kiwi-SAS parameters; $f_0 = 30$kHz, $B_c = 20$kHz, $D = 0.3$m, for targets at $r_0 = 30$m and 60m respectively. (e) and (f) were generated using the Kiwi-SAS parameters and spectral windowing. See the text for more details.

dynamic range in all cases, the L$_{\text{BWS}}$ gives a reasonable estimate of the image dynamic range for highly undersampled situations, but underestimates the dynamic range for more reasonable sample spacings. The effects of frequency can be seen by comparing Figs. 5.9(a) and (b) to Figs. 5.9(c) and (d); the previous University of Canterbury SAS employed lower frequencies, at lower frequencies the grating lobe targets are further from the main peak, so the sidelobe levels are lower in Figs. 5.9(a) and (b). A similar argument applies to the range dependence seen in a comparison between Figs. 5.9(a) and (c) to Figs. 5.9(b) and (d); at further ranges, the grating lobe targets are further from the main peak, so the maximum sidelobe level is lower in Figs. 5.9(b) and (d). The effects of spectral windowing can be seen by comparing Figs. 5.9(c) and (d) to Figs. 5.9(e) and (f); though the sidelobe level drops, the grating lobe level is not affected nearly as much, in fact it is worse for highly undersampled systems.

The main conclusion that can be drawn from this section, is that synthetic aperture imaging systems that sample at $D/2$ always alias energy into the processing bandwidth. An analysis based on sampling the full Doppler bandwidth produced across the mainlobe of the radiation pattern would seem to suggest that a sample spacing of $D/4$ is required, however, since only the 3dB width of the radiation pattern needs to be processed to produce an image with $D/2$ along-track resolution, a $D/3$ sample spacing is adequate and results in images with high dynamic range.

5.4.2   Aperture diversity (multiple apertures)

In an attempt to decouple the along-track sampling requirement from the range and along-track ambiguity constraints, a number of authors have suggested the use of multiple apertures. The suggestions can be split into three categories; an aperture that is used to form multiple beams that insonify different range swaths with different frequencies [94, 117, 161], an aperture with multiple beams squinted in along-track [32], and a single transmit aperture combined with multiple horizontal (along-track) receive apertures also known as a vernier array [56, 89, 94, 133]. This section shows that the vernier system is the only practical system. The increase in complexity of hardware and processing required to image with systems employing multiple beams and frequencies in range or multiple squinted beams in along-track makes the use of these systems unlikely.

The concept of forming multiple beams to insonify different non-overlapped range swaths with different non-overlapped frequencies was suggested by Walsh in 1969 [161], and Lee in 1979 [94] and was analyzed for application by Nelander in 1989 [117]. Each beam is formed by phasing an area on a large aperture such that the beam formed is steered in elevation to insonify a different range swath. The pulses used to insonify each swath cover sections of non-overlapping bandwidth with a center frequency that increases for the swaths further out. By increasing the frequency with standoff range, the synthetic aperture lengths are kept at a similar along-track length, and the width of the aperture required for forming the non-overlapped range beam is reduced. This method suffers from a number
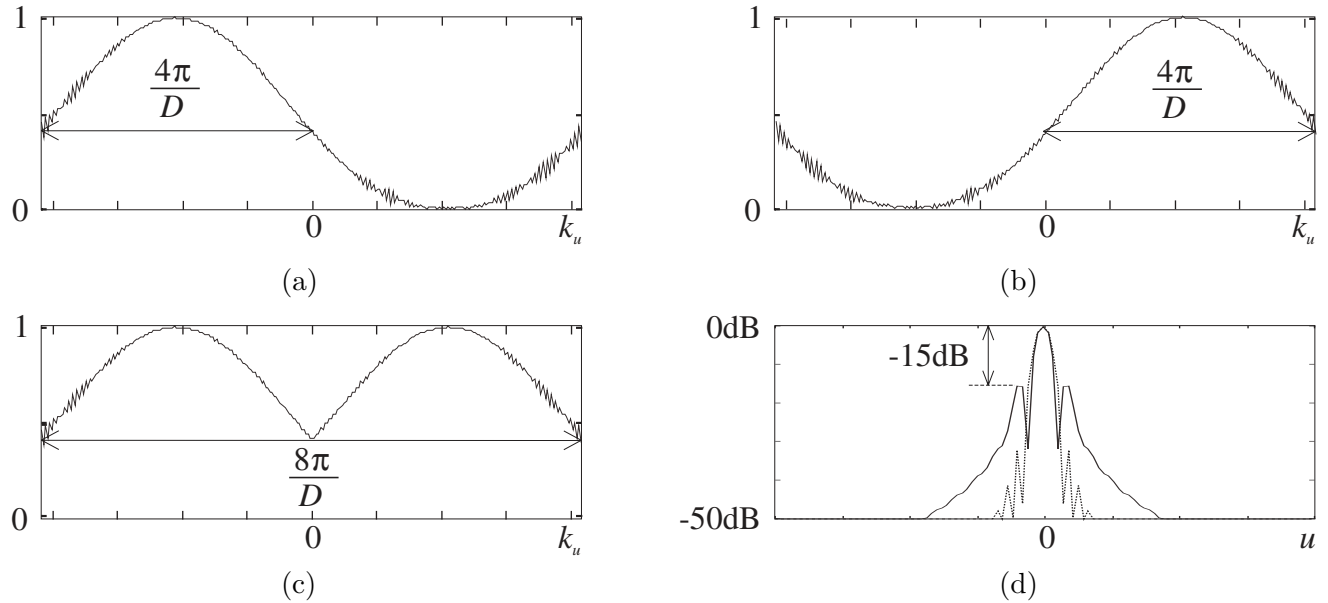
Figure 5.10 Cutrona's multibeam system. (a) Doppler spectrum produced by fore beam, (b) Doppler spectrum produced by aft beam, (c) Combined Doppler spectrum with amplitude modulation due to the radiation pattern of the real apertures, (d) Image cross-section; heavy line = image produced with amplitude modulated spectrum, dashed line = image produced if AM deconvolved (Hamming windows were applied in range and along-track during processing). Doppler spectra are cross-sections of the respective (ω, k_u)-domain signals at ω_0 with a $D/4$ along-track sample spacing.

of problems; for each beam to insonify only the swath of interest, it is necessary for each beam to have high directivity and thus a wide phased array, the attitude and height of the multi-beam system needs to be carefully maintained, and the use of multiple non-overlapped frequencies requires a very wide bandwidth transducer, or multiple transducers. In terms of signal processing, the swaths further out are imaged using interlaced pulses, in much the same way as a spaceborne SAR operates, and it is unlikely that low range-ambiguity levels can be achieved with the shallow grazing angles employed by SAS systems. The RASR of a multiple range beam system has not been analysed [117]. Also, if a single phased array is being used to image each range swath, it cannot simultaneously form all beams at once due to the different steer angles, different number of phased elements required and different frequencies of each beam [117]. The necessity of containing multiple or wide apertures on the limited space provided by a towfish, along with the stringent towing requirements, increased signal processing and probable high RASR makes the use of multiple range beams unlikely.

The use of a single long aperture that has separately phased channels producing multiple squinted beams was suggested by Cutrona in 1975, and 1977 [32, 33]. In fact, Cutrona cites this as "the single most important result" in reference [32]. A number of authors (see [18,37,42,75,93,133]) have referenced

this technique as being a possible method for avoiding spatial undersampling or increasing the along-track resolution, or have stated that it is essentially the same as the vernier technique. However, none of these authors have commented on the amplitude modulation induced on the Doppler spectrum due to the radiation patterns of the individual beams. For example, consider a system with an aperture of length D phased into two beams, $n = 2$, with one beam squinted forward by $\theta_{\text{fore}} = \sin^{-1}[k_{u0}/(2k)]$, the other squinted backwards (aft) by $\theta_{\text{aft}} = \sin^{-1}[-k_{u0}/(2k)]$, where the Doppler centroid $k_{u0} = 2\pi/D$ of each beam is such that the 3dB points of the aperture functions lie at zero Doppler independent of frequency ω. With this setup, the fore-beam produces the negative k_u-spectrum in the 2-D (ω, k_u)-domain, and the aft-beam produces the positive Doppler spectrum. Figure 5.10(a) and (b) show the k_u spectra of each fore and aft channels at ω_0, sampled at an along-track spacing of $\Delta_u = D/4$. The parts of these two spectra that lie between the 3dB points can be combined to produce a processing bandwidth of $B_p = 8\pi/D$, which is twice that normally available from a single beamed system, so an along-track resolution of $\delta y_{3\text{dB}} = D/4 = D/(2n)$ is achievable. However, Fig. 5.10(c) and (d) show that the amplitude modulation caused by the radiation pattern of the real apertures produces -15dB sidelobes in the final image (these sidelobes are at -10dB if Hamming windowing is not used during processing). These sidelobes are removed by deconvolving the radiation patterns of the real apertures from the Doppler spectrum as is shown by the dashed cross-section in Fig. 5.10(d). The increased hardware required for phasing, and the increased signal processing steps of Cutrona's method when compared to the vernier technique makes the use of the multibeam technique unlikely. (Note that when the along-track sample spacings are $D/2$ apart, or when more than two beams are being used, the Doppler spectra from each receiver channel must be demodulated before the correct 3dB bandwidth can be extracted to produce the processing bandwidth, B_p).

Kock 1972 [89], Gilmour 1978 [56], Lee 1979 [94], and de Heering 1982 [37] all suggested the use of a single small transmit aperture combined with multiple receive apertures for use in synthetic aperture radar and/or sonar systems. The single small transmitter produces a wide Doppler bandwidth that is then spatially sampled by multiple receive apertures. The receive apertures are spaced in along-track and are unphased. Sheriff 1992 [138], has proposed another method that employs multiple along-track receivers. Instead of being used to increase the sampling rate, the two apertures employed by Sheriff's system sample at the same spatial location. Any phase difference between the two signals recorded at the same spatial location is attributed to a motion error *in range* which can then be removed, thus producing perfectly focused images. Multiple receiver systems offer the most flexibility as they can be connected as a single hydrophone, or can be left as multiple hydrophones, beam forming does not have to be performed in real time, phase differences between pairs of apertures, or physical aperture images formed from the multiple receivers can be used as a form of motion compensation. The multiple receiver idea is again a trade-off between hardware and processing complexity, however, there is a minimal increase in the hardware complexity compared to the previous two multi-beam methods.

Of the publications on ocean-going SAS systems, only the researchers from the Sonatech Corp. and the University of California at Santa Barbara (UCSB) have published detailed information on their application of multiple apertures [40–43, 141]. The UCSB system, developed circa 1992, consisted of a 600kHz sonar with 4kHz bandwidth (Q=150), the transmitter aperture was 10cm long, and there were 10 receiving apertures each of 10cm length. The flexibility of multiple apertures was exploited using a novel motion compensation scheme that had similarities to Sheriff's algorithm. The UCSB motion compensation algorithm is known as the *cascade algorithm*. For each pulse transmitted, 10 receive channels were available, these 10 channels were focused to form a complex valued *preliminary* (lower resolution) physical aperture image. In the initial version of the cascade algorithm [40], the cross-correlation of these preliminary images in along-track produced a function that peaked at the along-track spatial translation distance of the sonar, i.e., the raw data contained an *estimate* of the spatial sample spacings over which the image was obtained. Using this spacing, the preliminary physical aperture images were registered, aligned and coherently combined, producing the final high-resolution synthetic aperture image [40]. Further motion parameter estimation, including range errors, is covered in [42, 141]. It is important to note that the along-track spacing of the UCSB system was such that they correctly sampled the Doppler spectrum at $k_{us} \approx 8\pi/D$, i.e., the multiple receivers sampled the spatial wave field produced by the transmitter at an effective rate that corresponded to sample spacings of close to $D/4$ where $D = 10$cm. In their 1993 and 1994 papers [41, 141], the UCSB group suggests a multilook technique that incoherently adds the physical aperture images. This is a misuse of the multilook concept as normally applied to synthetic aperture systems. Images produced in this way are identical to standard physical aperture images produced by a focused 600kHz sonar with a 1m aperture. The correct application of the multilook technique would be to coherently sum the preliminary images to produce a single high resolution image estimate, Fourier transform this estimate in the along-track direction, and then segment the resulting Doppler spectrum to produce looks in the normal fashion (see Section 4.7). If these looks are then incoherently combined, the final image has improved speckle statistics and has an along-track resolution that is independent of range (the resolution in this multilook image is lower than the resolution in the initial full bandwidth image estimate). Comments in Douglas' thesis [43] and their 1994 [141] paper could also mean that the assumptions made in the development of the cascade algorithm only make it applicable to narrow beam width, high frequency mapping systems. However, it is likely that it can be modified to work with wide bandwidth, low-Q, multiple receiver systems.

The UCSB system is nearly identical to the multiple receiver system proposed by Gilmour in his 1978 patent [56]. Gilmour's system also forms complex preliminary images/beams which can be detected and displayed as a normal side-looking sonar image, or they can be used to form synthetic beams for improved along-track resolution.

A point which is covered, but not always made clear in publications dealing with multi-receiver

sonars, is the consideration of the effect of the fixed path from the transmitter-to-target combined with the variable target-to-receiver path on the phase of the Doppler spectrum obtained for each transmitted pulse. The normal single transmitter, single receiver Doppler bandwidth is produced from the rate of change of spatial phase of a phase quantity that is proportional to $2kR$ where R is the one-way path length that is considered identical for the transmitter-to-target and target-to-receiver paths. For example, consider the UCSB SAS with a target at 30m broadside from the transmit element. The phase path from the transmitter-to-target is obviously 30m, but the return path to the outermost receive element is 30.0042m, i.e., it is 4.2mm further. A comparison of this path difference to the wavelength of 2.5mm (600kHz) shows that this extra delay is of the order of a wavelength, i.e., the range swath lies in the near-field of the physical aperture. This near field effect must be considered during the development of an appropriate model and inversion scheme for the processing of multi-receiver SAS data. This is why the UCSB system forms complex physical aperture images before coherently combining the data from the full synthetic aperture. Huxtable and Geyer's 1993 paper [78] contains a detailed simulation of a 150kHz carrier, 15kHz bandwidth, 10cm long transmitter, 12×10cm long receiver system that incorporates highly accurate motion compensation via INU and autofocus algorithms. Their paper also briefly mentions the necessity of removing the near field effects of the real multi-aperture receivers [78, p129].

High-resolution medical imaging systems using the synthetic aperture technique also employ multiple receiver arrays [86, 119, 154]. In Karaman's paper [86], the transmitter is also made up of multiple elements. Using a quadratic defocusing phase, he generates a radiation pattern similar to that of a much smaller aperture. The quadratic phase acts as a LFM chirp in spatial frequency and generates a wide Doppler bandwidth. Sampling and processing this Doppler bandwidth with multiple receivers results in images with high along-track resolution, and access to multiple transmit transducers means that higher power relative to a much smaller unphased array can be transmitted. The defocusing phase becomes insignificant when in the far-field of the transmitter [86, p431] (aperture defocus is covered in Section 3.2.1). In a similar way to the UCSB SAS, these medical systems also employ motion compensation based on the correlation of lower resolution subaperture images [86, p441], [154].

The high frequency operation and narrow bandwidth (high-Q) of the multi-aperture systems developed to date have allowed the developers to avoid radiation pattern effects in the multi-receiver models. Application of the multi-receiver technique to wide bandwidth SAS would need to include the effects of the wide bandwidth radiation pattern. Given the success of the multi-receiver systems covered in this section, it is likely that any realisable SAS of commercial value will employ multiple along-track receivers.

5.4.3   Pulse diversity

Another method that increases the along-track sampling frequency is the transmission of coded pulses [13, p145], [28, Ch. 8], [142, p10.18]. This method increases the processing complexity without increasing hardware complexity. Each pulse repetition period is split up into N equal periods and within each period a different orthogonal code is transmitted. Maximum time-bandwidth gain is achieved if each code is transmitted for the full time τ_{rep}/N allocated to it. On reception, the reflected signal can be pulse compressed using each code, with the result that the along-track sampling frequency is increased by the number of codes used. In general, the conditions of orthogonality are relaxed so that each code has good auto-correlation properties, i.e., high, narrow main peak, with low sidelobes, and reasonable cross-correlation properties, i.e., a uniform level. Because of the relaxed cross-correlation property of the coded pulses, systems that use coded pulses introduce a level of self-clutter that is determined by the cross-correlation properties of the codes used [129, p327].

In any application that uses coded pulses, the carrier frequency, pulse bandwidth and spectrum shape must be identical for each code. This is because the Doppler signal exploited by the synthetic aperture technique relies on a phase history in along-track that is dependent on the transmitted frequency and spectrum shape. If the codes introduce different carriers and amplitude modulation effects, the Doppler bandwidth aliases and error free inversion is not possible. The proof of this requirement is given by way of an example in the next section.

5.4.4   Frequency diversity

It is not possible to increase the along-track sampling rate of a synthetic aperture system by employing multiple codes with different frequency bands and hence different carriers for imaging the *same* range swath. The reason why is explained by way of an example.

In 1986, Gough published the incorrect statement that a continuous transmission frequency modulated (CTFM) based synthetic aperture sonar does not have its forward velocity fixed by the PRF [59, p334]. Other than the fact that CTFM is identical to FM-CW, and hence constrained by the same spatial sampling requirements as spatial based synthetic apertures (see Section 4.5), consider the following scenario: Since a CTFM sonar is linearly swept in frequency as the sonar moves linearly from one sampling point (start of each transmission) to the next, Gough suggested that it was possible to trade-off range resolution for along-track resolution. To trade-off range resolution for along-track resolution the LFM pulse can be considered to be N separate LFM pulses each of bandwidth B_c/N transmitted at points $u = m\Delta_u + n\Delta_u/N; n = 0, 1, \ldots, N-1$. Each pulse when demodulated covers a baseband band of frequencies $\omega_b = [-\pi B_c/N, \pi B_c/N]$. This argument has made two inaccurate assumptions; that each pulse reflects off the target identically regardless of aspect angle and frequency, and that the phase content of the demodulated signals correctly samples the Doppler wavenumber domain. The radiation

patterns shown in Fig. 5.2 show that the amplitude modulation in along-track is not independent of frequency or aspect angle. This assumption can however be satisfied if the combined radiation pattern of the apertures is made invariant with frequency, thus the amplitude pattern imposed on the demodulated spectra are slowly varying with u and hence a spatial Fourier transform can be performed using the principle of stationary phase. To determine what Doppler wavenumber spectrum is produced, consider a system imaging a point target at $(x, y) = (x_0, 0)$ with the LFM pulse sectioned in two. The continuous signal representing the baseband pulse compressed signals is given in the (ω_b, u)-domain by

$$
\begin{aligned}
Ss_{bi}(\omega_b, u) &= A(\omega_{\max}, x_0, -u) \cdot P_b(\omega_b) \cdot \exp\left(-j2(k_b + k_i)\sqrt{x_0^2 + u^2}\right) \\
&= \mathrm{sinc}^2\left(\frac{k_{\max}D}{2\pi} \cdot \frac{u}{\sqrt{x_0^2 + u^2}}\right) \cdot \mathrm{rect}\left(\frac{\omega_b}{\pi B_c}\right) \cdot \exp\left(-j2(k_b + k_i)\sqrt{x_0^2 + u^2}\right)
\end{aligned}
\tag{5.26}
$$

where $k_i = k_1$ or k_2 represents the carrier wavenumbers (center wavenumbers) of the pulse segments, k_b represents the baseband wavenumbers in the two segments, $A(\omega_{\max}, x_0, -u)$ is the frequency invariant radiation pattern fixed to the pattern of the highest frequency, and the pulse spectrum is assumed uniform. The continuous spatial Fourier transform of this signal is

$$
SS_{bi}(\omega_b, k_u) = \mathrm{sinc}^2\left(\frac{k_{\max}D}{2\pi} \cdot \frac{k_u}{2k}\right) \cdot \mathrm{rect}\left(\frac{\omega_b}{\pi B_c}\right) \cdot \sqrt{\frac{\pi x_0}{j(k_b + k_i)}} \cdot \exp\left(-j\sqrt{4(k_b + k_i)^2 - k_u^2} \cdot x_0\right)
\tag{5.27}
$$

where the radiation pattern in the 2-D frequency domain is now frequency dependent. The sampled signals representing the segmented LFM pulse are then

$$
Ss_{bs}(\omega_b, u) = Ss_{b1}(\omega_b, u) \cdot \sum_m \delta(u - m\Delta_u) + Ss_{b2}(\omega_b, u) \cdot \sum_m \delta\left(u - \left(m + \frac{1}{2}\right)\Delta_u\right)
\tag{5.28}
$$

and

$$
\begin{aligned}
SS_{bs}(\omega_b, k_u) &= SS_{b1}(\omega_b, k_u) \odot_{k_u} \frac{2\pi}{\Delta_u} \cdot \sum_n \delta\left(k_u - \frac{2\pi n}{\Delta_u}\right) \\
&\quad + SS_{b2}(\omega_b, k_u) \odot_{k_u} \frac{2\pi}{\Delta_u} \cdot \sum_n \delta\left(k_u - \frac{2\pi n}{\Delta_u}\right) \exp(-j\pi n) \\
&= \frac{2\pi}{\Delta_u} \cdot \sum_n SS_{b1}\left(\omega_b, k_u - \frac{2\pi n}{\Delta_u}\right) + \frac{2\pi}{\Delta_u} \cdot \sum_n SS_{b2}\left(\omega_b, k_u - \frac{2\pi n}{\Delta_u}\right) \exp(-j\pi n)
\end{aligned}
\tag{5.29}
$$

where the difference of half a sample between the two signals causes the sampled signal SS_{b2} to have an extra rotation of πn in each repeated spectrum.

Figure 5.11 shows what effect the carrier has on the Doppler spectra of the samples obtained at the
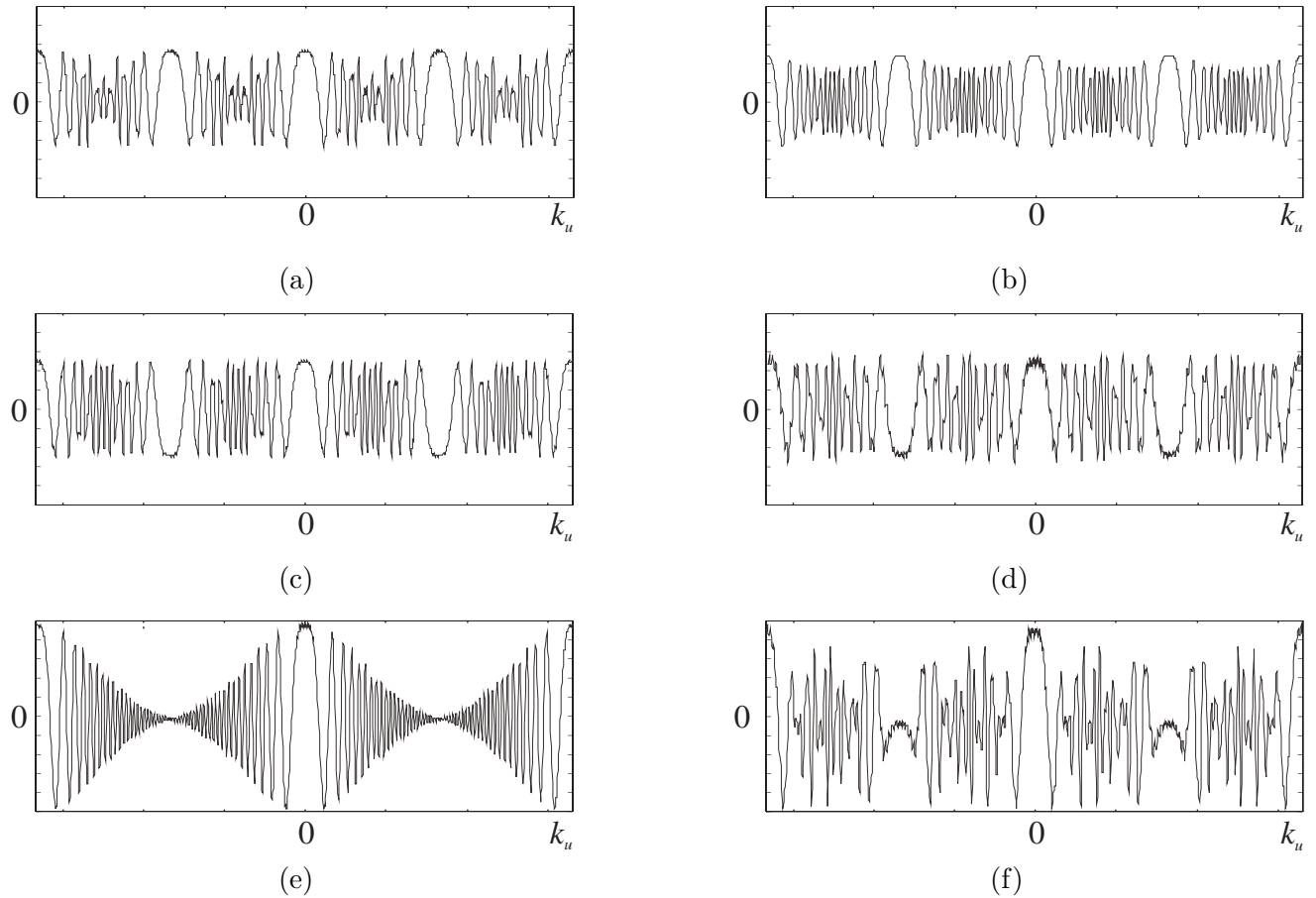
(a)

(b)

(c)

(d)

(e)

(f)

Figure 5.11 The effect of different carrier frequencies on the Doppler signal. (a), (c), and (e) represent the k_u-spectra produced by samples at carrier $\omega_1 = \omega_2 = \omega_0$ for $m\Delta_u$ ($\Delta_u = 3D/4$), $(m + 1/2)\Delta_u$, and the combined samples. The extra samples at the same carrier frequency coherently interfere and reduce the effect of aliasing. Similarly, (b), (d), and (e) represent the k_u-spectra produced by samples at carrier $\omega_1 = \omega_0 - \pi B_c/2$ for $m\Delta_u$, $\omega_2 = \omega_0 + \pi B_c/2$ for $(m + 1/2)\Delta_u$, and the combined effect. Because these samples are taken using signals with different carriers, the coherent interference of the two signals does not reduce the effects of k_u-aliasing.

along-track locations $m\Delta_u$ and $(m + 1/2)\Delta_u$ ($\Delta_u = 3D/4$). Figures 5.11(a), (c), and (e) show how the introduction of more samples in along-track at the *same* carrier frequency coherently interferes to reduce the level of along-track aliasing. Conversely, Figs. 5.11(b), (d), and (f) show how the introduction of more samples in along-track at a *different* carrier frequency does not result in coherent interference that reduces aliasing. Because the analytic form of the signal in 2-D Fourier space is known, it is possible to match filter the signal. Due to the fact that the radiation pattern is now frequency dependent in the 2-D domain, it is necessary to reduce the processing bandwidth to the 3dB bandwidth of the lowest radiation pattern in the (ω_b, k_u)-domain, i.e., $B_p = 4\pi/D \cdot k_{\min}/k_{\max}$. For an octave transmitted signal, $B_p = 2\pi/D$, so that $\delta y_{3\mathrm{dB}} = D$ is the maximum possible resolution achievable from this system. The

image produced from the matched filtering of the signals shown in (5.28) and (5.29) is identical to coherently adding two undersampled images. The coherent interaction of these two images does not act to remove aliasing effects, in fact sidelobes of -19dB are introduced close to the main target response (unweighted processing). If the full LFM pulse was not segmented and was processed as an undersampled signal, then the fact that only half the normal processing bandwidth was processed results in alias levels that are below -30dB anyway! In both cases, since a reduced section of the Doppler spectrum was produced, the level of the grating lobe targets was below -30dB.

In summary, CTFM is no different than the standard SAR methods and pulse diversity using different carrier frequencies does not result in an increased along-track sample rate.

5.4.5   Ambiguity suppression

In a series of papers, Moreira suggested the use of *ideal* or *inverse* filters to suppress ambiguous targets in undersampled images [108, 109, 112]. Moreira's filters are inverse filters that deconvolve the amplitude and phase errors due to spatial aliasing from within the processing bandwidth of systems that have been sampled at $\Delta_u = D/2$, i.e., they deconvolve the aliased energy in systems where the processing bandwidth equals the sampled bandwidth. Moreira's derivation and application was narrow band, this section generalizes Moreira's work to wide bandwidth systems and places it within the consistent mathematical framework of this thesis.

To develop the form of the inverse filter, consider for a moment a single point target. Moreira observed that arbitrary shifts of the sampling grid left the main target response unchanged, but caused rotations of the aliased sections of the target response. The mathematical formulation for a point target sampled at arbitrary locations is represented by the complex valued baseband signal

$$Ss_s(\omega_b, u) = Ss_\delta(\omega_b, u) \cdot \left\{ \delta(u - \delta u) \odot_u \sum_m \delta(u - m\Delta_u) \right\} \tag{5.30}$$

where δu represents an arbitrary shift of the sampling function and $Ss_\delta(\omega_b, u)$ represents the continuous form of the data recorded from a point target. Note that arbitrary shifting of the point target by δu is *not* the same as shifting the sampling function.

The spatial Fourier transform of the sampled point target signal is given by

$$SS_s(\omega_b, k_u) = \frac{2\pi}{\Delta_u} \cdot \sum_n SS_\delta \left(\omega_b, k_u - \frac{2\pi n}{\Delta_u} \right) \exp \left(-j2\pi n \frac{\delta u}{\Delta_u} \right) \tag{5.31}$$

where m represents the repeated spectra. This signal can be inverse spatial Fourier transformed to
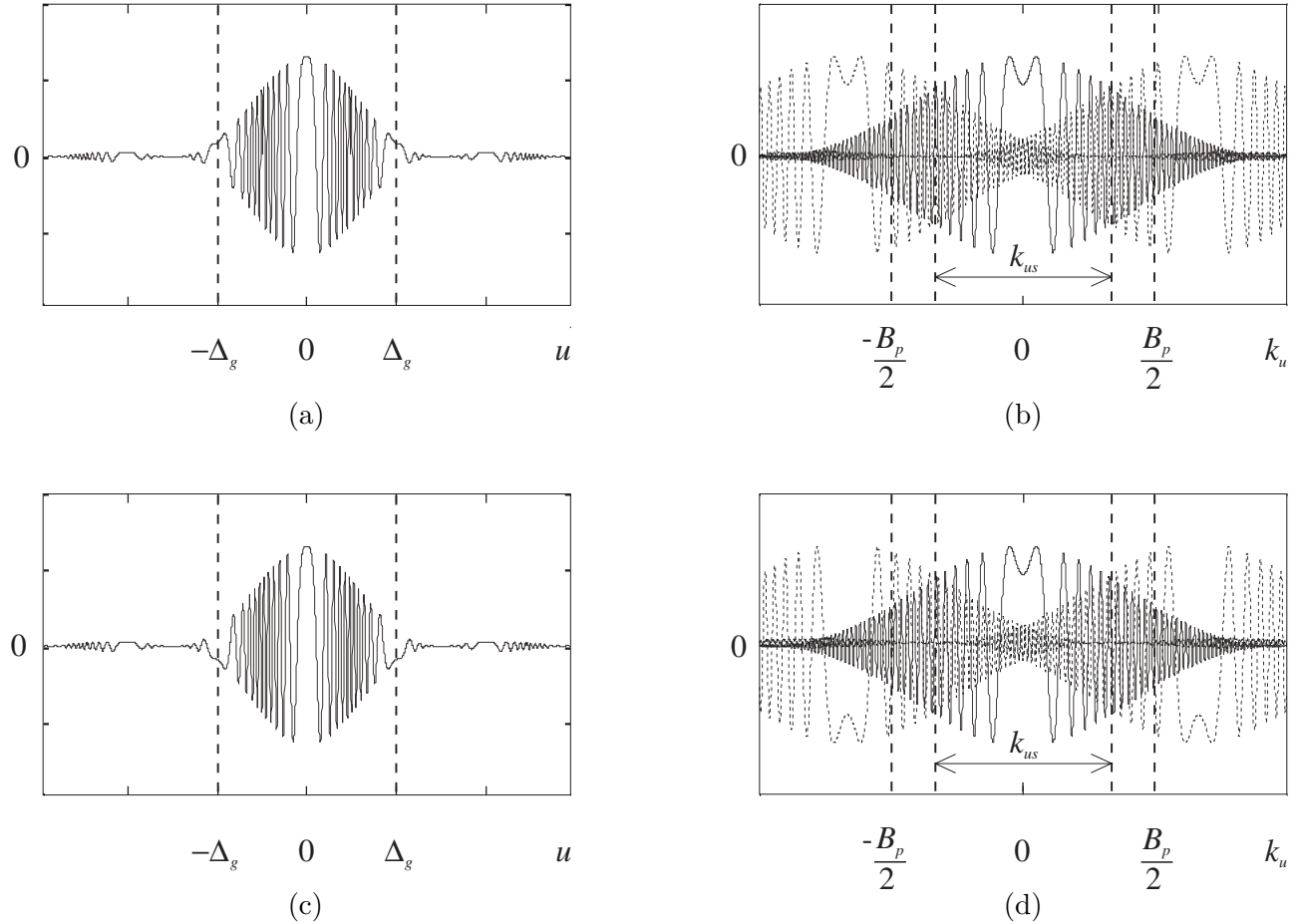
Figure 5.12 Rotation of the aliased energy in undersampled images. Figure (a) shows the along-track signal at frequency ω_0 for a sample spacing of $\Delta_u = 3D/4$ and a symmetric sampling function ($\delta u = 0$). Figure (b) shows this aliased signal in k_u-space. The repeated spectra interfere coherently to produce the signal shown in Fig. 5.5. Figures (c) and (d) show similar signals for non-symmetric sampling ($\delta u = \Delta_u/2$). The shift of half a sample in the sampling function causes a rotation of the first grating lobe. The analytic forms of the functions shown in (b) and (d) are used to generate the deconvolution filters.

determine the effects of these repeated spectrum on the spatial domain,

$$Ss_s(\omega_b, u) = \sum_n Ss_\delta(\omega_b, u) \exp\left(j\frac{2\pi n u}{\Delta_u}\right) \exp\left(-j2\pi n\frac{\delta u}{\Delta_u}\right) \operatorname{rect}\left(\frac{u - n\Delta_g}{\Delta_g}\right) \qquad (5.32)$$

where n represents the grating lobe targets, not the sample locations. The first exponential in (5.32) acts to demodulate the grating lobe energy so that it lies within the sampled spatial bandwidth, the second exponential is the arbitrary rotation due to sampling function location, and the rect function limits the grating lobes in along-track. The spacing of these grating lobes is given in (5.20).

Figure 5.12 shows the essence of the ideal filter concept for the along-track signal at the carrier

frequency ω_0. Figures 5.12(a) and (c) show the along-track signal when sampled with $\delta u = 0$ and $\Delta_u/2$. The shift of $\delta u = \Delta_u/2$ in Fig. 5.12(c) causes a rotation of the first grating lobe. An equivalent rotation is seen to occur in the Doppler wavenumber domain signals shown in Figs. 5.12(b) and (d). Figures 5.12(b) and (d) are the Doppler wavenumber signals *before* they coherently interfere, they are plotted this way for clarity. Knowledge of this rotating characteristic can be exploited to develop two separate inverse filters. If the actual signal obtained was symmetrically sampled, then deconvolution using the symmetrically sampled signal removes the aliased energy completely. However, deconvolution with the non-symmetrically sampled signal *reinforces* the grating lobe targets. Given that the actual sampling grid is completely arbitrary, two image estimates are generated using the two deconvolving filters. These image estimates have the first ambiguous target out of phase by 180°. These two estimates are then used to suppress the first ambiguous targets in an image estimate that has been formed using a conventional phase only matched filter.

The form of the aliased signal with no phase rotation ($\delta u = 0$) is

$$RR^{0°}(\omega_b, k_u) = \sum_{n=-1}^{1} SS'_\delta \left(\omega_b, k_u - \frac{2\pi n}{\Delta_u} \right) \tag{5.33}$$

and with a phase rotation of π ($\delta u = \Delta_u/2$) is

$$RR^{180°}(\omega_b, k_u) = \sum_{n=-1}^{1} SS'_\delta \left(\omega_b, k_u - \frac{2\pi n}{\Delta_u} \right) \exp(-j\pi n) \tag{5.34}$$

(where the use of the FFT to calculate the spectra is assumed). These two functions can be used to define two ideal (inverse) filters:

$$RR^{0°}_{\text{ideal}}(\omega_b, k_u) = \frac{1}{RR^{0°}(\omega_b, k_u)}$$
$$RR^{180°}_{\text{ideal}}(\omega_b, k_u) = \frac{1}{RR^{180°}(\omega_b, k_u)} \tag{5.35}$$

alternatively these inverse filters can be defined as Weiner filters [57, p231].

The 'normal' filter function is defined as,

$$RR_a(\omega_b, k_u) = \text{rect}\left(\frac{k_u}{B_p} \right) \cdot \frac{1}{A(k_u)} \cdot \sqrt{\frac{k}{k_0}} \cdot \exp\left[j \left(\sqrt{4k^2 - k_u^2} - 2k \right) \cdot r_0 \right] \tag{5.36}$$

where the radiation pattern is deconvolved as part of the filter function (the two ideal filters deconvolve radiation pattern effects too) and the phase function is due to the implementation of the processing via the FFT and is equal to the phase of the target impulse response (see Section 4.3.3).

If the raw, pulse compressed data is given by $ss_0(t, u)$, then three estimates of the image spectrum are obtained,

$$\widehat{FF}'_b(k_x, k_y) = WW(k_x, k_y) \cdot \mathcal{S}_b^{-1} \left\{ SS'_0(\omega_b, k_u) \cdot RR_a(\omega_b, k_u) \right\}$$

$$\widehat{FF}^{0°}_{\text{ideal}}(k_x, k_y) = WW(k_x, k_y) \cdot \mathcal{S}_b^{-1} \left\{ SS'_0(\omega_b, k_u) \cdot RR^{0°}_{\text{ideal}}(\omega_b, k_u) \right\} \qquad (5.37)$$

$$\widehat{FF}^{180°}_{\text{ideal}}(k_x, k_y) = WW(k_x, k_y) \cdot \mathcal{S}_b^{-1} \left\{ SS'_0(\omega_b, k_u) \cdot RR^{180°}_{\text{ideal}}(\omega_b, k_u) \right\}$$

where $WW(k_x, k_y)$ is a 2-D Hamming function limiting the estimates to the processed spatial bandwidths. The window in k_y reduces image artifacts due to the spectral discontinuities caused by the inverse filters and lowers along-track sidelobes, while the k_x window reduces range sidelobes.

The (k_x, y)-domain signals of these spectral estimates can be used to define an image mask (magnitude only) that contains mainly ambiguous target energy:

$$Ff_{\text{ambig}}(k_x, y) = \frac{\left| Ff^{180°}_{\text{ideal}}(k_x, y) - Ff^{0°}_{\text{ideal}}(k_x, y) \right|}{2} \qquad (5.38)$$

Subtracting (5.38) from the image magnitude data gives an image mask with the ambiguities suppressed:

$$Ff_{\text{ambig\_sup}}(k_x, y) = \left| \widehat{Ff}'_b(k_x, y) \right| - Ff_{\text{ambig}}(k_x, y) \qquad (5.39)$$

The definitions of (5.38) and (5.39) are slightly different to Moreira's definitions due to his narrowband approximation allowing him to operate on signals in the (x, y)-domain [109].

A limiting function that forces the processing to always result in some alias suppression is defined as:

$$10^{(-S/20)} \leq Ff_{\text{limit}}(k_x, y) = \frac{Ff_{\text{ambig\_sup}}(k_x, y)}{\left| \widehat{Ff}'_b(k_x, y) \right|} \leq 1 \qquad (5.40)$$

where S is the maximum dB level of suppression required (typically 5dB-15dB in Moreira's application).

The final image estimate with the first ambiguity suppressed is then given by

$$ff_{\text{ambig\_sup\_limit}}(x', y) = \mathcal{F}^{-1}_{k_x} \left\{ Ff_{\text{limit}}(k_x, y) \cdot \widehat{Ff}'_b(k_x, y) \right\} \qquad (5.41)$$

where it should be noted that the magnitude only image mask obtained from (5.40) is multiplied by the complex (k_x, y)-domain image estimate data. Moreira could apply his image mask directly to the image estimate in the (x, y)-domain [109, p892].

Figure 5.13 shows the application of ambiguity suppression to an image that has been sampled in along-track at $\Delta_u = 3D/4$. The transmitted signal covers an octave bandwidth. Figures 5.13(a) and
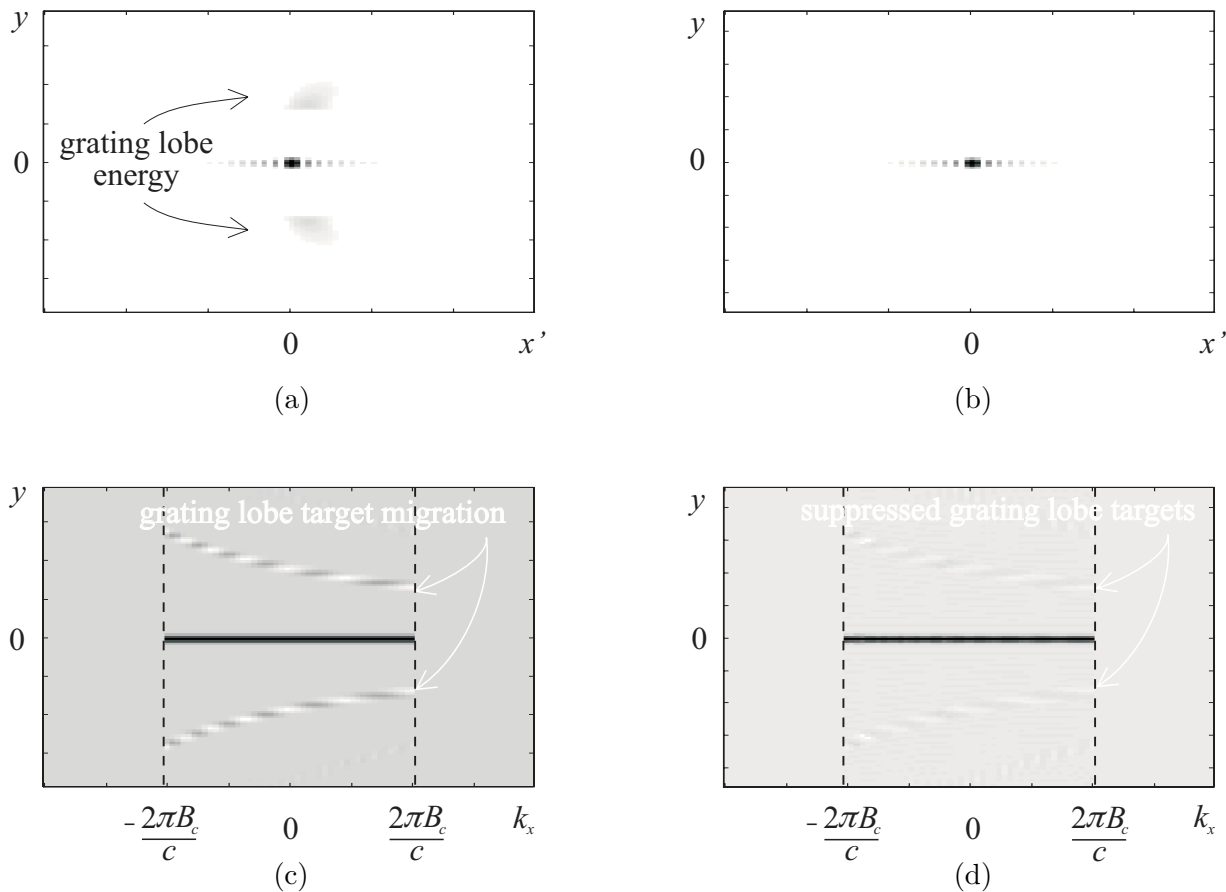
Figure 5.13   Ambiguity suppression of undersampled images, (a) and (c) represent the (x', y) and (k_x, y)-domain signals of a target processed in the normal fashion, (b) and (d) represent the same signals with the first grating lobe suppressed (but not removed). PGLR in (a) is -25dB, and -39dB in (b). Images (a) and (b) are logarithmic greyscale showing 30dB dynamic range, while (c) and (d) are the real part of the (k_x, y)-domain signals against a linear greyscale. Hamming windowing has been applied in along-track, but not in range.

(c) show the aliased image and its (k_x, y)-domain signal, while Figs. 5.13(b) and (d) show the successful suppression of the grating lobe energy. Hamming windowing has been applied in along-track to reduce the effects of the spectral discontinuity caused by the deconvolution, no Hamming weighting was applied in range. The PGLR of the undersampled image dropped from -25dB to -39dB, increasing the image dynamic range by 14dB.

Reasonable levels of ambiguity suppression were obtained up to a sample spacing of about $1.5D$ for simulations of a system based on the Kiwi-SAS parameters. When the along-track samples were wider than $1.5D$, the deconvolution filters did not work well due to the coherent interference of the repeated k_u-spectra and little or no ambiguity suppression was observed. The application of this technique to distributed targets did not yield promising results, Moreira also found this to be the case for his version

of this technique [112]. In summary then, Moreira's technique and the wide bandwidth version are limited in application to point targets, or targets of limited extent, that have been undersampling by a small amount.

5.4.6   Digital spotlighting

The spotlight models and inversion schemes presented in Chapter 4 can be modified to process the data produced from a strip-map system. In military applications where an aircraft carries a strip-map SAR that operates with parameters that also satisfy the tomographic formulation of spotlight mode, it is common for the SAR processor to produce spotlight image estimates from within the recorded strip-map data [116, p2144]. Often the along-track dimension of the aperture D produces a Doppler wavenumber bandwidth much greater than that required for reconnaissance imaging and presummers are employed to filter the excess Doppler bandwidth sampled by the SAR system([5, p381], [87, p331-333] [16]). The application of the generalized spotlight algorithm for forming these spotlight images is also possible. In the application of the generalized algorithm for wide bandwidth systems the presummer translates to an (ω, k_u)-domain filter. This filter is used after along-track compression to *digitally spotlight* the Doppler wavenumbers to those produced by the targets lying within the spotlight radius X_0. This radius is either less than or equal to the null-to-center width Y_0 of the real aperture radiation pattern [145]. Soumekh's article on a foliage penetrating (FOPEN) UHF band SAR contains an application of digital spotlighting. The FOPEN-SAR was a strip-map system that had been correctly sampled in along-track. Digital spotlighting was used to extract the signatures of targets within a region of this strip-map data [147]. The signature extraction method used a cross-range gating filter, i.e., the (ω, k_u)-domain filter, to remove aliased energy due to target signatures that lay outside of the digitally spotlight area. The aliasing discussed was not due to along-track undersampling. Digital spotlighting was also used in the FOPEN article to extract single target signatures. These signatures were then mapped back to the original (t, u)-domain of the raw data to assist in target classification.

This digital spotlighting method can also be used to recover undersampled strip-map data in situations where the target of interest has compact support and lies on a bland background. This situation would be typical for say, moored or proud mines lying below the sea surface near the sea floor in a harbour. The instantaneous Doppler frequency of targets lying within a radius X_0 at a standoff range r_0 from the sonar after along-track compression of the signal is given by (4.88). Given that the sampled Doppler wavenumbers are $k_u \in [-\pi/\Delta_u, \pi/\Delta_u]$, the radius of the correctly sampled area is given by

$$X_0 \approx \frac{\pi r_0}{2k\Delta_u} = \frac{\Delta_g}{2} = \frac{L_{\text{sa}}}{\gamma} \tag{5.42}$$

where $\gamma = \Delta_u/(D/4)$ can be interpreted as the *undersampling factor* of the null-to-null Doppler band-

width of the aperture. Targets that lie outside of the region defined by X_0, but within the null-to-center width of the aperture radiation pattern, fold their spectral energy into the undersampled Doppler spectrum. Thus the requirement that the undersampled target lies on a bland background. In this case, the aliased spectral energy causes minimal image degradation.

Digital spotlighting of compact undersampled targets makes use of the *a priori* knowledge that the amplitude function of the target is well sampled but the phase function is not. By deramping using along-track compression, the chirped phase of the target is removed leaving only the amplitude function and residual modulation of parts of the target away from broadside of the synthetic aperture radiation pattern. This lower frequency data is adequately sampled, so an image estimate is recovered. This technique introduces no new information to the scene and the patch over which it works is limited to a radius of $X_0 = \Delta_g/2$ about the digitally spotlight center.

Chapter 6

APERTURE ERRORS AND MOTION COMPENSATION

The formation of a synthetic aperture requires a coherent phase history to be maintained during the time it takes the platform to traverse the synthetic aperture. Up to this point we have considered the motion of the platform to be perfectly rectilinear, however, in a realistic situation the imaging platform is subject to motion perturbations. In fact, even when the platform follows a rectilinear path, atmospheric and ionospheric inhomogeneities effect the path lengths of reflected electromagnetic pulses used in SAR and medium layering, salinity, and varying sound speed profiles all lead to ray bending and path length fluctuations of the acoustic pulses used in SAS applications. If these path length variations or timing errors are not corrected to less than an effective range error of $\lambda_0/8$ then image blurring and loss of dynamic range occurs. Uncompensated high frequency motion or path length errors cause a rise in sidelobe level and resolution loss, while low frequency errors cause geometric distortion and poor map accuracy [49, 88].

6.1 BULK MOTION COMPENSATION

Airborne SAR systems were the first to implement motion compensation schemes. In airborne SARs, the PRF is normally slaved to Doppler navigation units which measure the aircraft speed v_p to keep the along-track sample spacing $\Delta_u = v_p/\text{PRF}$ constant [35]. Bulk motion compensation is typically based on the outputs of inertial navigation units (INUs) or inertial measurement units (IMUs) [87,88,106]. The outputs of the accelerometers in these INUs are double-integrated to yield the platform motions. Any drift or offset errors in the accelerometers translate to displacement errors that are proportional to time squared. In SAR applications, the time of synthetic aperture formation is typically a few seconds, so that motion parameters derived from INU measurements are often adequate for low resolution mapping. In sonar applications where the integration time can be on the order of minutes, or in SAR situations where high resolution is required, the output errors of INUs causes significant image degradation. No motion compensation suite based on INUs alone can account for delays due to medium inhomogeneities and electronic instabilities of the transmitter/receiver circuitry, so all practical synthetic aperture systems

employ some form of motion compensation based on the data itself, the algorithms used for estimating these motion perturbations are generically referred to as *autofocus* algorithms.

Digital processing of synthetic aperture data has meant that data collected along any path can be inverted as long as the synthetic aperture platform path is accurately estimated via INU and autofocus measurements. Thus, synthetic aperture imaging with maneuvers is possible [106] and almost arbitrary paths can be used to image scenes [146, p300]; see for example the circular flight path used in [148].

6.2   OVERVIEW OF AUTOFOCUS ALGORITHMS

6.2.1   Doppler parameter estimation, contrast optimization, multi-look registration, map-drift and subaperture correlation autofocusing

Spaceborne SAR motion is often well described by accurate ephemeris and orbital data. However, effective aperture squint due to orbital eccentricity and rotation of the earth causes a variation of the *Doppler centroid* and *Doppler rate*. The determination of these varying parameters are required for the accurate focusing of spaceborne SAR data. Tracking of the Doppler centroid is often referred to as *clutterlock*, while tracking of the Doppler rate is referred to as *autofocus* [31, 47, 48, 97, 100, 102].

The Doppler centroid is estimated using the knowledge that the Doppler spectrum of at least a synthetic aperture length of data resembles the overall radiation pattern of the real apertures, i.e., the spectrum envelope for broadside operation is ideally $A(k_u)$ [50] (this assumes that the Doppler wavenumber spectrum is well sampled, so that aliasing of the mainlobe of the radiation pattern does not distort the amplitude of the Doppler wavenumber spectrum). By correlating the known radiation pattern with the recorded Doppler spectrum, any offset of the beam center, i.e., any squint of the aperture, can be estimated and corrected for. Doppler centroid estimation can also be used in squinted operation to determine effective squint parameters. The Doppler rate (spatial chirp rate) is estimated by filtered/dividing the Doppler spectrum into a number of "looks" with each look corresponding to a certain bandwidth of the along-track Doppler wavenumber data. As with multi-look processing (see Section 4.7), each of these Doppler sections corresponds to a different "look" at the object field. Each look is focused using an initial estimate of the Doppler parameters to give a set of images with reduced along-track resolution. These low resolution image overlay exactly if the optimal Doppler parameters have been correctly estimated. However, if the Doppler rate is incorrect or if other parameters used by the focusing algorithm were in error, the resulting low resolution images are displaced relative to each other. The amount of image displacement is measured by cross-correlating each of the low resolution images, the displacements measured are then used to obtain an estimate of the parameter error.

Doppler parameter estimation is also used in the focusing of airborne SAR data. Autofocus techniques such as; *contrast optimization*, *multi-look registration*, *map-drift*, and *subaperture correlation* are

all methods that derive range varying estimates of the Doppler parameters [9–11, 31, 120]. Because the Doppler centroid and Doppler rate are parameters relating to the collection geometry, they can be considered to be deterministic quantities, i.e., they are completely predictable in cases where the platform speed and the aperture squint is known. Low frequency or low order polynomial approximations to non-deterministic phase errors caused by non-rectilinear motion and medium inhomogeneities can be estimated using these techniques, however, they can not estimate residual high frequency errors.

Chapter 6 of Carrara *et al*, 1995 [21], develops the map-drift algorithm (for quadratic errors, 2-6 iterations required), the multiple aperture map-drift algorithm (for higher order errors, 2-6 iterations), and the phase difference algorithm (similar to map-drift, but iteration is not required). Images estimates prior to, and after, autofocus are presented for a spotlight system. Chapter 4 of Jakowatz *et al*, 1996 [82], shows the ability, and inability, of map-drift algorithms to focus quadratic, and higher order aperture errors.

Section 6.7 details some important considerations regarding the application of these forms of autofocus to low-Q systems.

6.2.2   Phase gradient autofocus

Phase gradient autofocus (PGA) [44–46, 81, 82, 158] is typically used in the processing of spotlight-SAR images, but has recently been adapted to produce the phase curvature autofocus (PCA) algorithm for strip-map systems [159]. PGA models the along-track error as a phase-only function of the Doppler wavenumber in the range-Doppler domain. By using redundancy over range in the image, an estimate of this phase error can be obtained and removed from the image. PGA is often used in spotlight SAR systems to remove non-deterministic aperture errors in synthetic aperture data.

Chapter 4 of Jakowatz *et al*, 1996 [82], presents examples of images blurred with high and low frequency aperture errors along with the image estimates obtained after the successful removal of aperture errors by autofocusing with PGA. Chapter 4 of Carrara *et al*, 1995 [21], shows similar results.

PGA and Doppler parameter estimation techniques are normally applied iteratively to produce well focused images. The (tomographic) spotlight formulation of PGA is generally the fastest, typically converging in 2-3 iterations.

Phase gradient autofocus, the modified phase gradient algorithm and the phase curvature algorithm are all developed shortly.

6.2.3   Prominent point processing

Prominent point processing (PPP) is described in Carrara *et al* [21] as a method for estimating *arbitrary* motion parameters in SAR data. This form of algorithm has applications in inverse synthetic aperture

(ISAR) imaging and aspects of the algorithm are used in Chapter 7 to focus echo data recorded by the Kiwi-SAS system from point-like calibration targets.

Prominent point processing is formulated for spotlight systems that satisfy the tomographic paradigm. The PPP algorithm proceeds by selecting a single prominent point in the motion corrupted range compressed spotlight data, $\widetilde{ss}_b(t_s, u)$, where the tilde indicates a corrupted version of the recorded echos and the notation follows the conventions defined in Section 4.4. Section 4.4 shows that the spectral data $Ss_b(\omega_b, u)$, obtained from a spotlight system, can be interpreted as an estimate of the 2-D wavenumber spectrum of the image, i.e., $\widehat{FF}_b(k_x, k_y)$ (the two are related via a polar transform). Thus, the corrupted data, $\widetilde{ss}_b(t_s, u)$, could also be seen as a corrupted version of the range-Doppler space data $\widehat{fF}_b(x, k_y)$, i.e., $\widetilde{fF}_b(x, k_y)$. This duality of domains sometimes makes the description of spotlight algorithms confusing to follow.

The uncorrupted echo data, $ss_b(t_s, u)$, obtained from a hypothetical scene containing a single target at the stand-off range r_0, would have the appearance of a single peak located at range gated time $t_s = 0$ for all along-track locations u. In a corrupted version of this hypothetical scene, this straight peak has 'wiggles' in it due to the motion perturbations. To remove the effects of motion perturbations, PPP first selects a prominent point and then a reference echo pulse against which other echo pulses are correlated. Any shift of the correlation peak from $t_s = 0$ indicates erroneous motion *in range* relative to the reference peak. Correlation of all the echo pulses gives an estimate of the motion errors in the range direction for all along-track locations u. Once the peak shifts are determined in the (t_s, u)-domain (which is approximately equal to the (x, k_y)-domain), they can be removed using a phase multiply in the (ω_b, u)-domain (i.e., in the (k_x, k_y) domain as a 2-D phase multiply). After the phase multiply necessary to straighten the prominent point is applied, the resulting data is polar remapped and inverse Fourier transformed to give the image estimate (note that the first prominent point becomes the center of the image as a consequence of the peak straightening phase multiplies—see Carrara [21] for more details). The selection and processing of two more prominent points in the scene allows the rotation rate and scaling of the target to be estimated [21] (the addition of more prominent points is useful in ISAR imaging of unknown targets, such as flying aircraft).

6.3   POSSIBLE LIMITATIONS IN THE APPLICATION OF SAR AUTOFOCUS ALGORITHMS TO SAS

SAR systems have a much higher number of along-track samples, or hits-on-target, relative to SAS applications due to their large standoff range (esp. spaceborne SAR) and their high PRFs (esp. airborne SAR). The system geometry employed by SAR systems means that they have a much higher space-bandwidth product or synthetic aperture improvement factor relative to SAS systems.

Doppler parameter estimation or autofocus algorithms that divide the along-track data up into

subapertures and use correlation techniques for parameter estimation work well in SAR systems that suffer low frequency aperture errors due to the high number of hits-on-target in these systems. However, in SAS systems operating with a single receiver, the number of hits-on-target is much smaller so that the division of the synthetic aperture into smaller apertures yields few along-track samples in each subaperture. Each subaperture image then has a poor SNR and each parameter estimate has a high variance. Similar comments were made regarding the application of the multiple aperture map drift algorithm to spotlight data in [21, p254]. Thus, it is unlikely that Doppler parameter estimation techniques will work well with single receiver SAS systems. These algorithms may be useful as a preliminary autofocus step to remove the bulk of low frequency aperture errors. This partially focused image could then be passed through an autofocus procedure which removes the high frequency errors.

Multiple receiver SAS systems have access to more than one receiver channel per transmitted pulse. The UCSB-SAS publications and Douglas's thesis contain information on the *cascade* autofocus algorithm which is used in multi-receiver systems [40–42, 141]. In cascade autofocus, the echoes received by each of the multiple apertures are used to form preliminary complex physical images. Correlation based techniques are then used with these preliminary images to determine estimates of motion parameters. Huxtable and Geyer's 1993 paper covers the end-to-end simulation of a multiple receiver system incorporating a high-quality motion compensation suite and autofocusing routines [78]. Both the map-drift and PGA autofocus methods were successfully used to remove residual motion errors. The system simulated by Huxtable and Geyer and the UCSB SAS are high frequency, high-Q systems.

Further important considerations regarding autofocus algorithms are covered in Section 6.7.

6.4   MODELLING MOTION AND TIMING ERRORS IN SYNTHETIC APERTURE SYSTEMS

To obtain tractable mathematical solutions, residual motion errors and path length variations through the air or water are often parameterized as unknown timing errors in range [24, 32, 82, 120, 138]. This treatment is analogous to the introduction of an unknown random lens into the imaging system [120]. If the fluctuations of this lens occur on a scale that is large compared to the synthetic aperture then Doppler parameter estimation techniques can be used to correct for the unknown lens. When the fluctuations occur on a much smaller spatial scale a more accurate model of the timing error is required. Spotlight images showing the effects of low frequency (quadratic) and high frequency (power-law spectrum) aperture errors are shown on pp233-237 of Jakowatz *et al*, 1996 [82]. The low frequency errors cause a loss of image resolution, while the high frequency errors produce a rise in sidelobe levels, reducing image contrast.

If the motion errors are parameterized as a displacement in range given by $X(u)$, then the received

echoes in the (ω_b, u) domain are given by $(X(u) \ll \text{range to target})$

$$
\begin{aligned}
\widetilde{Ee}'_b(\omega_b, u) &\approx \exp\left[j2kX(u)\right] \cdot Ee'_b(\omega_b, u) \\
&= Zz(\omega_b, u) \cdot Ee'_b(\omega_b, u),
\end{aligned}
\tag{6.1}
$$

where $Zz(\omega_b, u) = \exp\left[j2kX(u)\right]$ represents the error function to be determined by the autofocus procedure and $Ee'_b(\omega_b, u)$ represents the complex valued, baseband, error free data (calculated using the FFT).

The error model transforms to the (ω_b, k_u)-domain as

$$
\widetilde{EE}'_b(\omega_b, k_u) \approx ZZ(\omega_b, k_u) \odot_{k_u} EE'_b(\omega_b, k_u),
\tag{6.2}
$$

where it can be seen that the error function acts as a modulation function that spreads the 2-D spectrum in Doppler wavenumber. The image spectral estimate formed from this corrupted data using say, a wavenumber processor, is given by

$$
\begin{aligned}
\widetilde{FF}'_b(k_x, k_y) &\approx \mathcal{S}_b^{-1}\left\{ \sqrt{\frac{k}{k_0}} \cdot \exp\left[j\left(\sqrt{4k^2 - k_u^2} - 2k\right) \cdot r_0\right] \cdot P_b^*(\omega_b) \cdot \widetilde{EE}'_b(\omega_b, k_u) \right\} \\
&= \mathcal{S}_b^{-1}\left\{ \sqrt{\frac{k}{k_0}} \cdot \exp\left[j\left(\sqrt{4k^2 - k_u^2} - 2k\right) \cdot r_0\right] \cdot P_b^*(\omega_b) \right. \\
&\quad \left. \cdot \left[ZZ(\omega_b, k_u) \odot_{k_u} \sqrt{\frac{\pi}{jk}} \cdot P_b(\omega_b) \cdot A(k_u) \right. \right. \\
&\quad \left. \left. \cdot \exp\left[-j\left(\sqrt{4k^2 - k_u^2} - 2k\right) \cdot r_0\right] \cdot \mathcal{S}\left\{FF'_x(k_x, k_y)\right\} \right] \right\},
\end{aligned}
\tag{6.3}
$$

where the windowless version of (4.44) and the baseband version of (4.41) have been used. The convolution in k_u, the phase multiplies, and the Stolt mappings make it difficult to describe the effect of the timing errors on the final image or the final image spectrum as a simple convolution or multiplication.

Most autofocus algorithms make the assumption that the along-track error can be modelled as a one-dimensional function in the (t, u)-domain or the (t, k_u) domain, or similarly as a one dimensional function in the (x, y)-domain or the (x, k_y)-domain. This assumption limits the application of SAR autofocus procedures to SAS, however, aspects of the SAR algorithms can be used to formulate wide bandwidth autofocus schemes.

6.5 PHASE GRADIENT AUTOFOCUS (PGA)

The publications dealing with PGA to date have applied the algorithm either to spotlight systems that satisfy the tomographic paradigm, or to narrow swath strip-map SAR systems. This section describes the PGA formulation in the generalized format of this thesis and presents the modifications necessary for PGA to work with the generalized spotlight formulation and general strip-map systems.

6.5.1 Timing errors in the tomographic formulation

PGA exploits characteristics of the tomographic formulation. The step of remapping the baseband deramped (ω_b, u)-domain data directly into wavenumber (k_x, k_y)-space allows the effect of the error to be modelled as

$$
\begin{aligned}
\widetilde{FF}'_b(k_x, k_y) &\approx \mathcal{P}^{-1} \left\{ \exp \left[j2kX(u) \right] \cdot Ss'_b(\omega_b, u) \right\} \\
&\approx \mathcal{P}^{-1} \left\{ \exp \left[j2(k_b + k_0)X(u) \right] \cdot Ss'_b(\omega_b, u) \right\} \\
&= \mathcal{P}^{-1} \left\{ \mathcal{F}_{t_s} \left\{ \exp \left[j2k_0X(u) \right] \cdot ss_b \left(t_s - \frac{2}{c}X(u), u \right) \right\} \right\}.
\end{aligned}
\tag{6.4}
$$

The airborne SAR systems that PGA was developed for typically deal with residual range errors on the order of wavelengths (2-3cm) and have range resolution cells on the order of 1m. These system parameters allow PGA to make the valid assumption that typical timing errors due to the motion error $X(u)$ are much less than a range resolution cell (note that this is a high-Q assumption). This assumption allows the timing error, or equivalently, the envelope shift of the pulse compressed response, to be ignored. This leaves only the carrier phase term:

$$
\widetilde{FF}'_b(k_x, k_y) \approx \mathcal{P}^{-1} \left\{ \exp \left[j2k_0X(u) \right] \cdot Ss'_b(\omega_b, u) \right\}.
\tag{6.5}
$$

In the tomographic formulation, the u-domain maps through to the k_y-domain via the polar remap operator (4.84), so that the error model becomes

$$
\widetilde{FF}'_b(k_x, k_y) \approx \exp \left[j2k_0X \left(u = -\frac{k_y}{k_x + 2k_0}r_0 \right) \right] \cdot \widehat{FF}'_b(k_x, k_y).
\tag{6.6}
$$

Due to the fact that only the phase error term of the final model is important, amplitude terms, such as the Jacobian, are ignored.

Spotlight SARs that satisfy the tomographic formulation are high-Q systems and they generally only subtend a few degrees to obtain equivalent resolution in range and along-track. The almost square region of (ω_b, u)-data obtained from such a system remaps to a region in the (k_x, k_y)-domain that is only slightly distorted by the polar transformation (for example see p179 of [82]). PGA thus assumes that

the polar mapping operation has a minor effect on the along-track error (see footnote p226 [82]). The high-Q nature of the spotlight system means that variation of k_x in the term $k_x + 2k_0$ in the denominator of (6.6) is small relative to $2k_0$, this fact allows the approximation $k_x + 2k_0 \approx 2k_0$ to be used giving

$$
\begin{aligned}
\widetilde{FF}'_b(k_x, k_y) &\approx \exp\left[j2k_0 X\left(u = -\frac{k_y}{2k_0}r_0\right)\right] \cdot \widehat{FF}'_b(k_x, k_y) \\
&= Z(k_y) \cdot \widehat{FF}'_b(k_x, k_y),
\end{aligned}
\tag{6.7}
$$

where the error is now modelled as a multiplicative k_y-dependent function,

$$
\begin{aligned}
Z(k_y) &= \exp\left[j2k_0 X\left(u = -\frac{k_y}{2k_0}r_0\right)\right] \\
&= \exp\left[j\phi(k_y)\right].
\end{aligned}
\tag{6.8}
$$

The final image estimate obtained via the tomographic formulation of the phase error is then

$$
\widetilde{ff}_b(x, y) = z(y) \circledcirc_y \widehat{ff}_b(x, y),
\tag{6.9}
$$

where it can be seen that the timing (motion) errors cause a spreading of the target response in the along-track direction (but not in range due to the high-Q assumption). Note that due to the assumptions made during the development of this tomographic model that this blurring function is *identical* for each target in the image. In the generalized spotlight mode, or in any strip-map application, this is not the case. The one-dimensional nature of the PGA timing (motion) error model allows the algorithm to exploit *redundancy over range*. A one-dimensional error estimate is obtained from every line of a function based on the corrupted range-Doppler estimate $\widetilde{fF}_b(x, k_y) = Z(k_y) \cdot \widehat{fF}_b(x, k_y)$. Each of these estimates are then combined to improve the final error estimate. This procedure is developed in the next section.

6.5.2 The PGA Algorithm

Using superposition, the target field $ff(x, y)$ can be modelled as a collection of point targets of reflectivity f_n:

$$
ff(x, y) = \sum_{n=0}^{\infty} f_n \delta(x - x_n, y - y_n).
\tag{6.10}
$$

The uncorrupted image estimate $\widehat{ff}(x, y)$ is the diffraction limited version of the target field. Ignoring diffraction limiting effects for the moment, the corrupted image estimate becomes

$$\widetilde{ff}(x, y) = \sum_{n=0}^{\infty} f_n \delta(x - x_n) z(y - y_n). \tag{6.11}$$

The phase gradient autofocus (PGA) algorithm proceeds as follows:

1. In the target scene $\widetilde{ff}(x, y)$, the N strongest targets are shifted to the center of the image, i.e., their y_n shifts are removed. This gives the shifted matrix

$$gg_0(x, y) = \sum_{n=0}^{N-1} f_n \delta(x - x_n) z(y) + \sum_{n=N}^{\infty} f_n \delta(x - x_n) z(y - y_n). \tag{6.12}$$

2. The weaker targets are removed by a windowing operation. If the degrading phase error is considered to be a low order polynomial then the window is determined from an integral (sum) of the intensity of the data in range:

$$w(y) = \begin{cases} 1 & \text{if} \quad 10 \log_{10} \left(\frac{\int_x |gg_0(x,y)|^2 dx}{\max\{\int_x |gg_0(x,y)|^2 dx\}} \right) > -10\text{dB} \\ 0 & \text{otherwise} \end{cases} \tag{6.13}$$

If instead of a polynomial type error, the phase error contains high-frequency components, then the window size is typically halved for each iteration.

The shifted windowed matrix is given by

$$\begin{aligned} gg_1(x, y) &= w(y) gg_0(x, y) \\ &= \sum_{n=0}^{N-1} f_n \delta(x - x_n) z(y) + nn(x, y), \end{aligned} \tag{6.14}$$

where only the N prominent targets and a noise term $nn(x, y)$ representing the interfering clutter around each of the selected point targets remains.

3. The non-zero elements within the shifted, windowed matrix $gg_1(x, y)$ are then padded to the next

power of 2 and along-track Fourier transformed to produce the following range-Doppler function:

$$
\begin{aligned}
gG_1(x, k_y) &= \mathcal{F}_y\left\{gg_1(x, y)\right\} \\
&= \sum_{n=0}^{N-1} f_n \delta(x - x_n) Z(k_y) + nN(x, k_y) \\
&= \sum_{n=0}^{N-1} a(x) \exp[j\phi(k_y)] + nN(x, k_y),
\end{aligned}
\tag{6.15}
$$

where $a(x) = f_n \delta(x - x_n)$ is an amplitude function in range representing the complex target reflectivity. Represented (6.15) in sampled form gives

$$
gG_1[p, q] = gG_1(p\Delta_x, q\Delta_{k_y}) = a[p] \exp\left(j\phi[q]\right) + nN[p, q],
\tag{6.16}
$$

where p and q are the indices of the range and along-track samples in the (x, k_y)-domain at sample spacings of (Δ_x, Δ_{k_y}).

The *phase gradient* of the phase error in (6.15) can be determined in a Maximum Likelihood sense as [82, p258]

$$
\widehat{\Delta\phi}[q] = \arg\left(\sum_{p=0}^{P-1}\left\{gG_1^*[p, q-1]gG_1[p, q]\right\}\right),
\tag{6.17}
$$

where the $\arg(\cdot)$ operator determines the phase of its argument. The original version of PGA used a slightly different phase estimation operation [44] [21, p267].

4. Once the phase gradient estimate is obtained for all along-track wavenumber indices, q, the entire aperture phase error is determined by integrating (summing) the $\widehat{\Delta\phi}[q]$ values:

$$
\widehat{\phi[q]} = \sum_{l=1}^{Q-1} \widehat{\Delta\phi}[l]; \qquad \widehat{\phi}[0] \equiv 0.
\tag{6.18}
$$

5. The phase estimate then has any linear trend and offset removed to prevent image shifting and it is upsampled to contain the same number of samples as the k_y dimension of the range-Doppler matrix representing $\widetilde{fF}(x, k_y) = \mathcal{F}_y\left\{\widetilde{ff}(x, y)\right\}$.

6. The phase error is removed from the original (or previous) image estimate by multiplying the range-Doppler image estimate with the complex conjugate of the measured phase error,

$$
\widetilde{fF}_{\text{new}}(x, k_y) = \exp\left\{-j\widehat{\phi}(k_y)\right\} \cdot \widetilde{fF}(x, k_y).
\tag{6.19}
$$

Table 6.1   Spotlight-SAR simulation parameters

Aperture length	m	L_{sp}	400
Spotlight area radius	m	X_0	64
Range to spotlight center	m	r_0	1000
Signal wavenumbers	radians/m	$k = \omega/c$	$2\pi[0.8828, 1.1172]$

Though it is not necessary for the operation of PGA, the original along-track track motion error function can be estimated from the phase estimate via

$$\widehat{X}(u) \approx \frac{1}{2k_0} \widehat{\phi}\left(k_y \approx -\frac{2k_0 u}{r_0}\right). \tag{6.20}$$

7. The starting point for the next iteration is then the new image estimate given by

$$\widetilde{ff}_{\mathrm{new}}(x, y) = \mathcal{F}_{k_y}^{-1}\left\{\widetilde{fF}_{\mathrm{new}}(x, k_y)\right\}. \tag{6.21}$$

Convergence of the algorithm can be determined by setting a tolerance on the RMS phase error to be recovered, or a minimum window size that must be reached. Typically the algorithm converges in 2-3 iterations.

6.5.3   Modified PGA for generalized spotlight systems (the failure of PGA)

Figure 6.1 and Fig. 6.2 show a spotlight target field simulated with the parameters shown in Table 6.1. These parameters *do not* satisfy the tomographic formulation and the system has a (reasonably low) quality factor of $Q = 300\mathrm{MHz}/70\mathrm{MHz} = 4.3$. These parameters were obtained from [144], other parameters not given here can be derived from equations in Chapter 4 or [144]. Because the system parameters do not satisfy the tomographic formulation of spotlight mode, the images in Fig. 6.1 and Fig. 6.2 were processed using a generalized processor (in this case, one based on the chirp scaling algorithm).

Figure 6.1(a) shows a corrupted image estimate. This image was formed by convolving (in along-track) an error function with an uncorrupted diffraction limited image estimate. Figures 6.1(b) and (c) show the focused image and error estimate obtained using PGA. The convolution in Fig. 6.1(a) acts to spread the target energy in along-track, but does not broaden the target in range (the high-Q assumption). PGA correctly recovers the phase error because the error function was applied to an *uncorrupted image* as a *convolution* in the along-track direction.

Figure 6.2(a) was generated using the same target field as Fig 6.1(a), however, this time the along-track motion error function was included as a *timing error* during the generation of the raw data (which
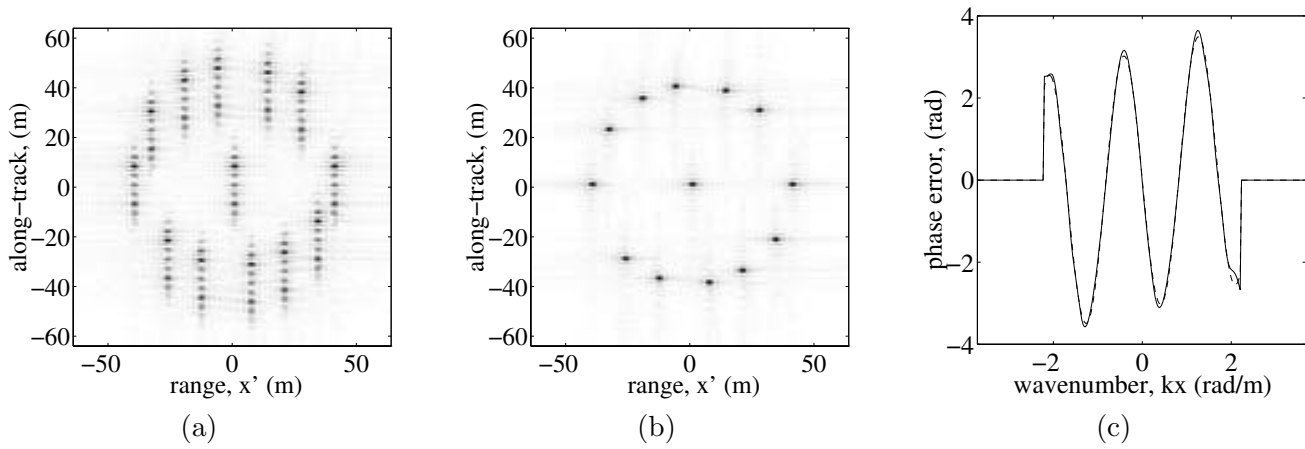
Figure 6.1   Phase Gradient Autofocus (PGA). (a) Initial image estimate, $|\widetilde{ff}(x',y)|$ (corrupted via a convolution in along-track y), (b) Image estimate after 2 iterations of PGA, (c) Corrupting phase (dashed line) vs recovered phase (solid line). Magnitude images are linear greyscale.
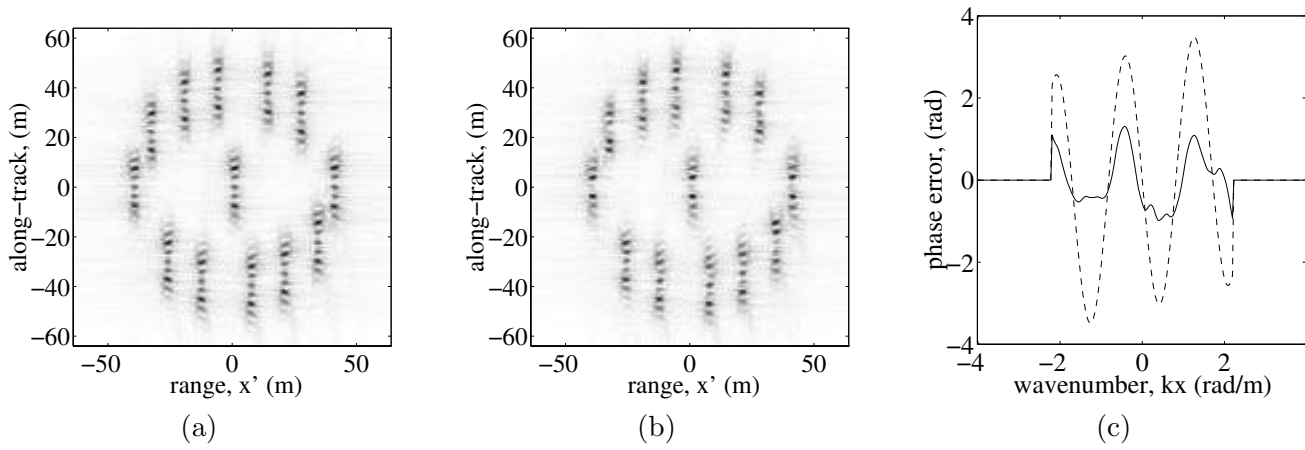

Figure 6.2   The failure of Phase Gradient Autofocus (PGA). (a) Initial image estimate, $|\widetilde{ff}(x',y)|$ (the raw echo data was corrupted with an along-track timing error, $2X(u)/c$), (b) target field after 10 iterations of PGA, (c) Motion error scaled to wavenumber error phase, i.e., $\phi(k_y) = 2k_0 X[u = -k_y r_0/(2k_0)]$, (dashed line) vs recovered phase (solid line). Magnitude images are linear greyscale.

was then focussed using a chirp scaling based generalized spotlight processor). The more accurate modelling of the motion error has resulted in range broadening of the blurred targets, however, the predominant blur is still in the along-track direction. Figure 6.2(b) and (c) show that the along-track error is not correctly recovered by PGA in this case. The modelling of the along-track error as an along-track convolution by PGA is not appropriate for generalized spotlight systems, i.e., the model is

not appropriate for systems where the plane wave approximation of the tomographic formulation fails. The ability of PGA to accurately model timing errors as a convolution for tomographic based systems is purely a consequence of being able to map the baseband $Ss_b(\omega_b, u)$ data directly to the wavenumber domain.

The reason why PGA does not work for generalized systems is due to the underlying phase of the targets in the (x, y)-domain of the corrupted image estimate. The error terms that are generated during the collection in the (t, u)-domain do *not* result in an error function that can be modelled as an invariant convolution in y. The phase of the nth target at (x_n, y_n) imaged by a platform located at u is

$$\phi(u) = -2k_0\sqrt{(x_n + r_0)^2 + (y_n - u)^2}, \tag{6.22}$$

where the distances x_n are distances relative to the origin of the x'-axis (required by the FFT) located at the standoff range r_0, i.e., at time $t = 2r_0/c$. The rate of change of this phase term with respect to u gives the instantaneous Doppler wavenumber of a target:

$$\begin{aligned} k_{ui}(x_n, y_n) &= \frac{2k_0(y_n - u)}{\sqrt{(x_n + r_0)^2 + (y_n - u)^2}} \\ &\approx \frac{2k_0}{r_0}y_n - \frac{2k_0}{r_0}u. \end{aligned} \tag{6.23}$$

where it is assumed that $\sqrt{(x_n + r_0)^2 + (y_n - u)^2} \approx r_0$, i.e., the target locations are small relative to the standoff range.

The approximate one-to-one relationship relating the aperture dimension u to the wavenumber dimension k_u, and hence k_y, is then

$$u \approx -\frac{k_y r_0}{2k_0} + y_n. \tag{6.24}$$

For the tomographic formulation, the equivalent one-to-one relationship assumed was: $u \approx -k_y r_0/(2k_0)$ (see (6.7)). In a generalized geometry, the phase centers of the along-track chirps of the individual targets do not lie at the wavenumber origin (this effect is seen in Fig. 4.7). This effect acts to produce a *modulation* of the target phase error function in the wavenumber domain, i.e., there is an extra phase modulation term $\exp[+j(2k_0 y_n/r_0) \cdot y]$ in the spatial domain (the origin of this term is given in (6.23)). This modulation term needs to be removed in the spatial domain prior to any phase estimation in the wavenumber domain. During the first step of PGA when the targets are *spatially* shifted to the scene center to produce the function $gg_0(x, y)$, their spectra should also be *demodulated* (using $\exp[-j(2k_0 y_n/r_0) \cdot y]$). The phase function estimated using steps 2-5 of Section 6.5.2 is then converted

back to an along-track motion error estimate via

$$\widehat{X}(u) \approx \frac{1}{2k_0}\widehat{\phi}\left(k_y = -\frac{2k_0 u}{r_0}\right).$$

(6.25)

This along-track error is then removed from the corrupted image *data* as a timing error, to give the new data estimate

$$\widetilde{ss}_{b\text{new}}(t_s, u) = \mathcal{F}_{\omega_b}^{-1}\left\{\exp[-j2k\widehat{X}(u)] \cdot \widetilde{Ss}_b(\omega_b, u)\right\}.$$

(6.26)

(this operation could have equivalently been performed on $\widetilde{ee}_b(t_s, u)$). This new data set is then used to form the new image estimate and the process is repeated until convergence occurs. As the process iterates, a cleaner raw data set is produced each time. Alternatively, the motion error estimate can be stored on a cumulative basis and can be applied to the *original* data set each time.

This *modified* PGA algorithm was applied to the corrupted image in Fig. 6.2(a) and in 2 iterations it obtained similar figures to those shown in Fig. 6.1(b) and (c). Showing that, at least for this example, the modified algorithm does produce an accurate *motion* error estimate. This modified PGA algorithm is still based on the narrow bandwidth, narrow swath-width assumption of (6.23), so its usefulness is limited (but not as limited as tomographic PGA).

6.6   STRIP-MAP PHASE CURVATURE AUTOFOCUS (PCA)

In 1994, Wahl *et al* [159] introduced the modifications necessary to allow a PGA-type algorithm to operate on corrupted data obtained using a strip-map system. The algorithm was presented for a high-Q SAR system where range migration is less than a resolution cell (Fresnel-approximation based imaging): this fact is implied by their assumption that a convolution in along-track with the appropriate spatial chirp reproduces the original data set (see p53 [159]). In low-Q systems where range migration is larger than a resolution cell, the reintroduction of the along-track chirp does not reproduce the original data set, what it does reproduce, is the *range migration corrected* data. The formulation of PCA is presented in this section, while the next section discusses whether the range migration corrected data can be used as the starting point of a low-Q autofocus routine.

Phase curvature autofocus (PCA) begins in the same way as PGA and isolates a number of point-like targets in the initial image estimate. These targets are then windowed (but not shifted) and convolved with the appropriate along-track spatial chirp to give the *range migration corrected* data; that is,

$$\begin{aligned}
cc(x, y) &= gg_2(x, y) \circledcirc_y \exp\left[-j2k_0\left(\sqrt{x^2 + y^2} - x\right)\right] \\
&= \mathcal{F}_{k_y}^{-1}\left\{gG_2(x, k_y) \cdot \exp\left[-j\left(\sqrt{4k_0^2 - k_y^2} - 2k_0\right) \cdot x\right]\right\},
\end{aligned}$$

(6.27)

where $gg_2(x, y)$ is a windowed but not shifted version of the initial motion corrupted image estimate containing only point-like targets. The reintroduction of the along-track chirp is performed using the second line of (6.27) in the range-Doppler domain. The derivation of the appropriate along-track chirp is found in Section 4.3.2, the range-Doppler domain chirp used in the second line of (6.27) is the conjugate of $qQ(x, k_x)$ (see (4.28)); the chirp that is used to perform the along-track compression in the range-Doppler processor.

The data resulting from the reintroduction of the along-track chirp in (6.27) is referred to as the range-migration corrected data as (if the effects of windowing are ignored) the data is equivalent to pulse compressed data that has had the range migration correction map, $\mathcal{T}^{-1}\{\cdot\}$, applied to it, but not the along-track compression (see Section 4.3.2). For the narrow beam width, high-Q, systems that PCA was developed for, it is not necessary to perform the range-migration correction map, however, to investigate the application of PCA to low-Q systems it is necessary to state this equivalence.

For low-Q systems, the range migration corrected data can also be formed using the chirp scaling and wavenumber algorithms. To produce the range migration corrected data using the chirp scaling algorithm, only the phase residual terms are removed during the last phase multiply (see (4.64)). If a wavenumber processor is used, then the along-track chirp has to be reintroduced as shown in (6.27). During the last step of the algorithm, instead of performing an inverse 2-D FFT to produce the image estimate, a 1-D inverse transform in k_x takes the data to the range-Doppler domain where the along-track chirp is reintroduced, then the data is inverse transformed in k_y to give the range migration corrected data. Range migration correction decouples the data coordinates (t, u) from the image coordinates (x, y) and allows them to be interchanged, i.e., the use of x or $ct/2$ and y or u in the system equations from this point on becomes arbitrary.

Given that the convolution in along-track in (6.27) spreads the target response out and 'unfolds' the effects of the motion error, and the high-Q assumptions that; the effects of range migration can be ignored, and the motion error only effects the carrier, (6.27) becomes

$$cc(x, y) \approx \exp[j2k_0 X(y)] \cdot \left\{ \widehat{ff}_n(x, y) \odot_y \exp\left[-j2k_0\left(\sqrt{x^2 + y^2} - x\right)\right]\right\}, \tag{6.28}$$

where $\widehat{ff}_n(x, y)$ is the error free, diffraction limited version of the n windowed targets represented in the same way as (6.10). Ignoring diffraction effects, (6.28) becomes

$$cc(x, y) \approx \exp[j2k_0 X(y)] \cdot \sum_n f_n \exp\left[-j2k_0\left(\sqrt{x_n^2 + (y - y_n)^2} - x_n\right)\right] \cdot \delta(x - x_n). \tag{6.29}$$

The use of this model by PCA to obtain an estimate of the along-track sway error is best explained by way of an example. Figure 6.3 shows images produced using the high-Q SAR system parameters given in Table 6.2. Figure 6.3(a) and (b) show the pulse compressed data and image estimate for data that

has been corrupted by a sinusoidal sway error (generated as a timing error during the image formation). Figures 6.3(c) and (d) show the same images for an error free flight path. Though Figs. 6.3(a) and (c) show the pulse compressed data, this data could also have been generated by applying (6.27) to the images in Figs. 6.3(b) and (d); Fig 6.3(a) is described by (6.29), while Fig 6.3(c) is described by (6.29) with $X(y) = 0$. The magnitude of the corrupting sinusoidal sway simulated was $\lambda_0 = 2$cm at a spatial frequency of 0.1 rad/m ($4\pi/128$ rad/m, i.e., 2 cycles along the aperture). Note that there is no obvious difference between the magnitude image of the pulse compressed corrupted data in Fig. 6.3(a) and the magnitude image of the pulse compressed uncorrupted data in Fig. 6.3(c), this is because the range cell resolution of 1m is 50 times greater than the sway error (and also much greater than the range cell migration). The fact that typical sway errors in SAR are much smaller than the range resolution cell size is why high-Q systems can make the valid assumption that the timing errors typically experienced during data collection only effect the *phase* of the echo signal, not the location of the echo *envelope*.

When dealing with diffraction limited imagery as in Fig. 6.3, the term $f_n\delta(x - x_n)$ in (6.29) corresponds to the sinc-like range responses of the pulse compressed targets in Figs. 6.3(a) and (c), with a width given by the range resolution (for a true point-like target). If the $n = 6$ targets in Fig. 6.3(c) were selected by PCA for use in phase error estimation, then the second phase term in (6.29) gives the value of the phase expected for along-track values $|y - y_n| \leq L_{\rm sa}/2$ about each of the selected targets at y_n, where $L_{\rm sa}$ is the synthetic aperture length of the nth target at range x_n. Any residual phase terms measured along the target responses in Fig. 6.3(a) are due to the along-track error phase (the first phase term in (6.29)). Measurement of these phase residuals gives an estimated of the phase error function for all points covered by the spread response of the n targets, i.e., each target provides a small y-segment of the overall error function. Redundancy over range is exploited for those targets that have similar along-track locations and extents but have different range locations x_n [159].

Because the phase error function is built up on a segment by segment basis, it is necessary to calculate the phase curvature (second differential) of the error function. This is so that any linear component of the phase error is also determined by the autofocusing algorithm [159, p54]. Once a phase error estimate $\widehat{\phi}(y) = \exp[j2k_0X(y)]$ is estimated, the complex conjugate, $\widehat{\phi}^*(y)$ is applied to the corrupted pulse compressed data (eg. applied to the data in Fig. 6.3(a)), the result is a new estimate of the pulse compressed data which is then along-track compressed (via a phase multiply in the range-Doppler domain), to give the next image estimate. Iteration of PCA is required as it is difficult to determine the exact y_n locations of the blurred targets in images like Fig. 6.3(b). For the example shown in Fig. 6.3, 4 iterations of PCA produced an image indistinguishable from Fig. 6.3(d).

As with PGA, PCA makes the high-Q assumption that the timing (motion) error can be treated simply as a phase error function. If however, the final phase error function is mapped back to a motion error, the correct timing correction can be applied to the raw data in much the same way as modified PGA. This comment does not effect the application of PCA to high-Q systems, however, it may be

Table 6.2   Simulation parameters for the high-Q SAR and low-Q SAS systems that were used to simulate Figs. 6.3 and 6.4.

Parameter	Symbol	Units	High-Q SAR	Low-Q SAS
Wave propagation speed	c	m/s	$3 \cdot 10^8$	$1.5 \cdot 10^3$
Carrier frequency	f_0	Hz	$15 \cdot 10^9$	$12 \cdot 10^3$
Carrier wavelength	λ_0	cm	2	12.5
Bandwidth	B_c	Hz	$150 \cdot 10^6$	$5 \cdot 10^3$
Sampling frequency	$f_s = 2B_c$	Hz	$300 \cdot 10^6$	$10 \cdot 10^3$
Aperture length	D	m	2	0.3
Along-track sample spacing	$\Delta_u = D/4$	cm	50	7.5
Standoff range	r_0	m	7000	30
Quality factor	$Q = f_0/B_c$	-	100	2.4
Image size		pixels	256×128	256×128
		m	128×64	19.2×9.6
Image resolution	$\delta x_{3\mathrm{dB}} \times \delta y_{3\mathrm{dB}}$	m	1×1	0.15×0.15

useful in determining whether it is possible to develop a PCA-like algorithm for low-Q systems.

6.7   SALIENT FEATURES OF AUTOFOCUS ALGORITHMS

To determine a valid *phase* estimate from the data set within any step of the synthetic aperture image formation algorithms, it is necessary to use data that has a reasonable *magnitude*. To begin to develop a phase estimation model for systems where the aperture error can not be modelled as a simple along-track convolution, it is necessary to determine the domains in which phase estimates can be considered to be reliable. The initial corrupted image estimate is often characterised by blurred point targets that have spurious peaks either side of what may, or may not, be the true target peak (see for example Fig. 6.3(c) and Fig. 6.4(c)). Thus, the initial image estimate is not a good candidate for phase retrieval. There are two possible methods of generating a magnitude function from which a reliable phase estimate may be obtained; the first function is the same as is used by PGA, i.e., the along-track Fourier domain of the image estimate, and the second is the method used by PCA, to reintroduce the along-track spatial chirp into the image estimate.

Reliable phase estimates are obtained in the tomographic formulation of PGA because a point target shifted to the center of the aperture has an along-track Fourier spectrum that is almost uniform. The spectrum is almost uniform due to the fact that spotlight SARs typically sample at a rate that
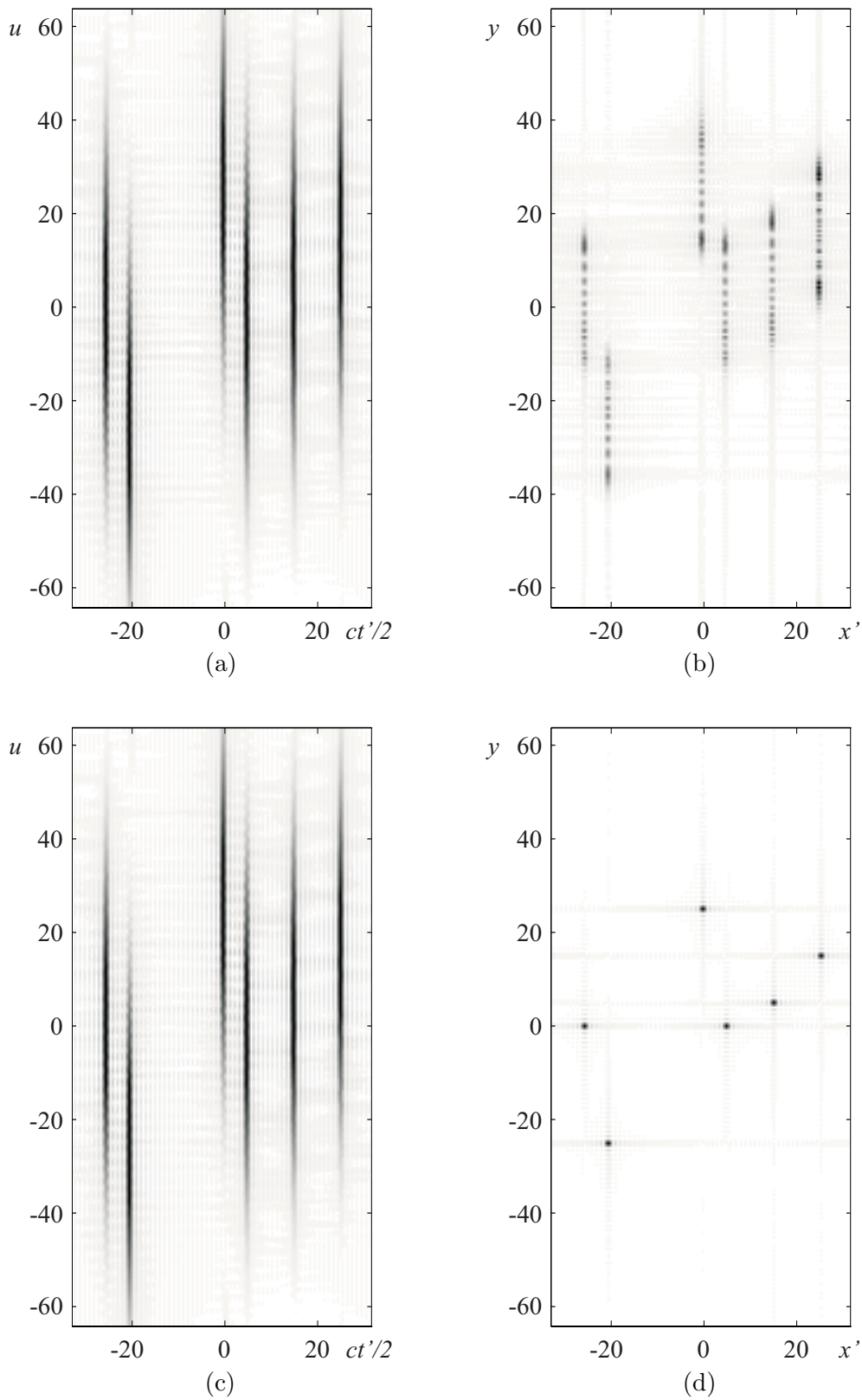
Figure 6.3 The effects of sway on high-Q, narrow beam width strip-map SAR images. (a) sway corrupted pulse compressed data, (b) sway corrupted image estimate, (c) sway free pulse compressed data, (d) sway free image estimate. See Table 6.2 for the system parameters and the text for more details.

oversamples the null-to-null Doppler bandwidth produced by the small real aperture carried by the aircraft. The echo data is presummed on reception to reduce the Doppler bandwidth to the extent required to image the spotlight patch. This presumming operation often reduces the Doppler bandwidth to less than the -0.1dB width of the aperture function $A(k_u)$, hence the spectrum is fairly uniform and the phase estimates are stable. This particular aspect of the Doppler spectrum may appear to be limited to spotlight systems alone, however, if sub-apertures of strip-map data, smaller than the synthetic aperture length of a target, are extracted, then an approximate one-to-one relationship exists between the wavenumber domain and the aperture domain and the Doppler spectrum has a reasonable magnitude. If the data is not segmented into subapertures, when the along-track Fourier transform is performed, the phase and magnitude terms due to the u-variable error function coherently interfere and error estimation techniques fail.

The phase estimation method exploited by PCA can be used in spotlight or strip-map systems. The reintroduction of the along-track chirp smears or spreads the target energy in the along-track direction, creating a magnitude function over which reliable phase estimates can be obtained. The downside of this method is that the phase retrieval algorithm also has to account for the spatial chirp phase of the targets.

The fact that the magnitude function used in PCA during the phase estimation step is equivalent to the *range migration corrected* pulse compressed data, is an important observation for low-Q systems. If any of the Doppler parameter estimation algorithms developed for high-Q systems are used for the autofocusing of low-Q data, the range migration should first be corrected. This is necessary so that the along-track correlations of the multilook/subaperture images are correlated correctly. If range migration is left in, the data in the subaperture looks will not correlate well (if indeed at all). The comments in this paragraph also apply to the reflectivity displacement method (RDM) autofocus algorithm [111,113].

When dealing with motion errors, the error magnitude is usually quantified in terms of the carrier wavelength, λ_0, and an error of more than $\lambda_0/8$ is considered to cause image degradation. Most radar systems make the valid assumption that the shift of the target envelope can be ignored during autofocus estimation and the error function can be treated as depending only on the carrier wavelength. When the size of the along-track motion error is comparable to, or larger than, the system range resolution the effect of this error should not be treated as a phase only effect. This statement can be interpreted in terms of the system quality factor via the relative ratio of the range resolution to the carrier wavelength:

$$\frac{\delta x_{3\text{dB}}}{\lambda_0} = \frac{f_0}{2B_c} = \frac{Q}{2}. \tag{6.30}$$

In high-Q systems, motion errors on the order of λ_0 are small compared to the range resolution and hence any *shift* of the range response is not discernible in the corrupted data. Often only smearing in along-track is observed (see Fig. 6.1(a) and Fig. 6.3(b)). When dealing with low-Q systems these

peak shifts are discernible (see Fig. 6.2(a) and Fig. 6.4(c)) and another option becomes available for autofocusing; tracking the shifts in the peaks of the point-like targets.

The observations of this section do not restrict the use of PGA and PCA, as was discussed in Section 6.5.3, so long as the phase estimate is mapped back to a motion error and it is removed as a timing error, these algorithms can, in some cases, produce focused imagery in low-Q systems.

6.8   THE FAILURE OF THE PCA MODEL FOR LOW-Q SYSTEMS

The effect of the phase error on the image estimate in a wide bandwidth system is modelled by (6.3). Using a similar argument to that used in the development of the PCA algorithm above, it is assumed that the effect of a motion (timing) error can be modelled in the *range migration corrected* data (as opposed to the pulse compressed data) as

$$dd(x, y) = \delta[x - X(y)] \odot_x \left\{ \widehat{ff}(x, y) \odot_y \exp\left[-j2k_0 \left(\sqrt{x^2 + y^2} - x \right) \right] \right\}, \tag{6.31}$$

where the high-Q assumption in (6.28) that the timing error only effects the phase, has been replaced with a shift in the echo envelope.

Figure 6.4 shows images produced using the low-Q SAS system parameters given in Table 6.2. Figure 6.4(a), (b), and (c) show the pulse compressed data, the range migration corrected data, and image estimate for data that has been corrupted by a sinusoidal sway error (generated as a timing error during the image formation). Figures 6.4(d), (e), and (f) show the same images for an error free tow path. Figures 6.4(b) and (e) show the pulse compressed data after range migration correction, similar data sets could have been generated by reintroducing the along-track chirp as per (6.27) to the images in Figs. 6.3(c) and (f). The magnitude of the corrupting sinusoidal sway simulated was $\lambda_0 = 12.5$cm at a spatial frequency of 0.65 rad/m ($4\pi/19.2$ rad/m, i.e., 2 cycles along the aperture). Compare Fig. 6.4 to the images in Fig. 6.3 produced using the high-Q SAR system. The pulse compressed data in Fig. 6.4(a) clearly shows the effects of range migration and the sway error on the low-Q SAS system. The image estimate in Fig. 6.4(c) shows both range and along-track blurring as opposed to Fig. 6.3(b) which only shows along-track blurring. For the SAS system, the range resolution of 15cm is comparable to the wavelength of 12.5cm. The fact that the typical sway errors experienced in SAS are comparable to the range resolution cell size means, that any mathematical model of a low-Q SAS system can *not* be developed using the assumption that the timing errors typically experienced during data collection only effect the *phase* of the echo signal, not the location of the echo *envelope*.

To determine whether (6.31) is a valid model, Fig. 6.5 shows; (a) Fig. 6.4(b) with the known errors that were introduced during the image formation removed according to the inverse of the operation described in (6.31), and (b) Fig. 6.4(e) with the timing errors introduced into the range migration
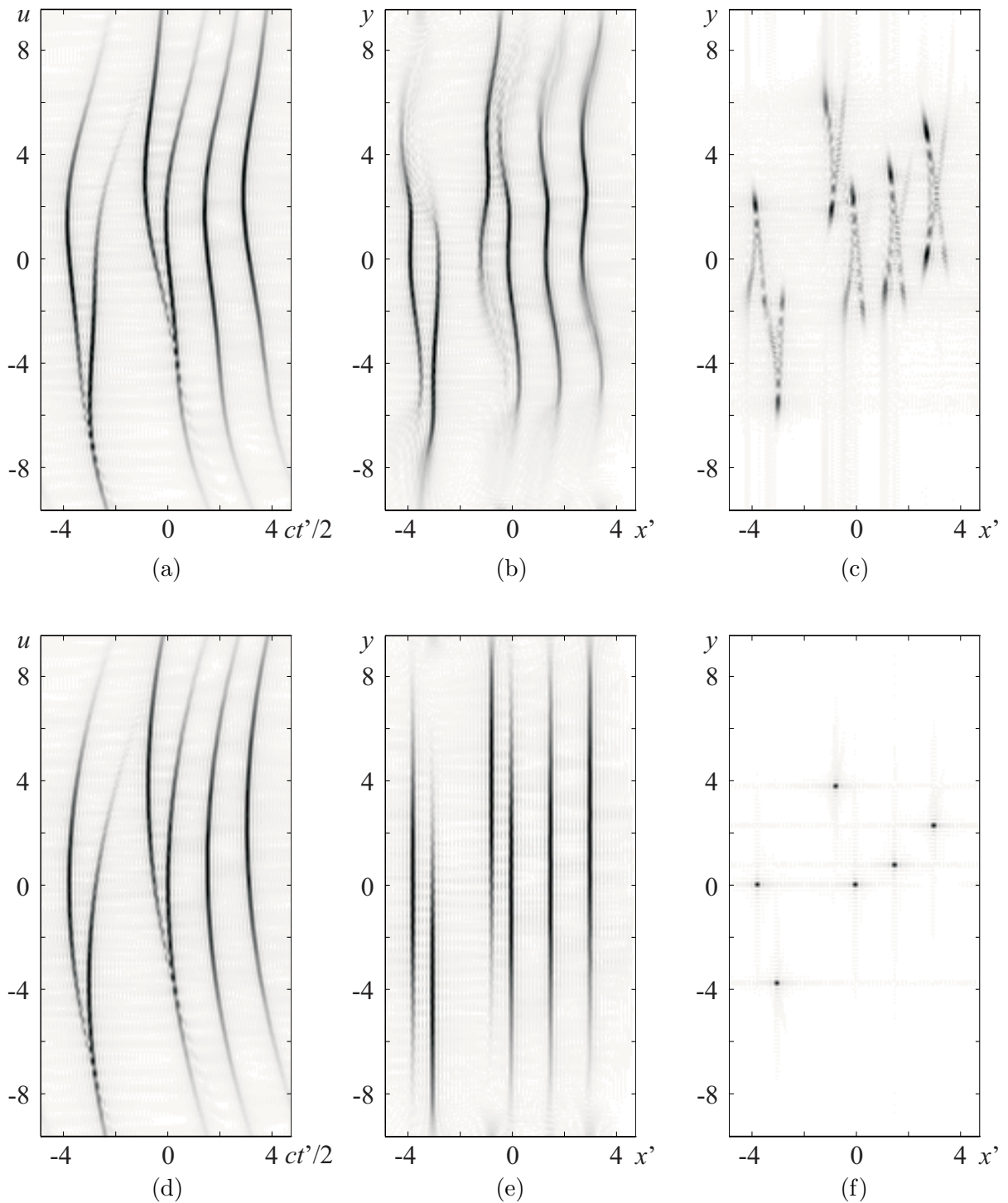
Figure 6.4   The effects of sway on low-Q, wide beam width strip-map SAS images. (a) sway corrupted pulse compressed data, (b) range migration corrected sway corrupted data (c) sway corrupted image estimate, (d) sway free pulse compressed data, (e) range migration corrected data, (e) sway free image estimate. See Table 6.2 for the system parameters and the text for more details.
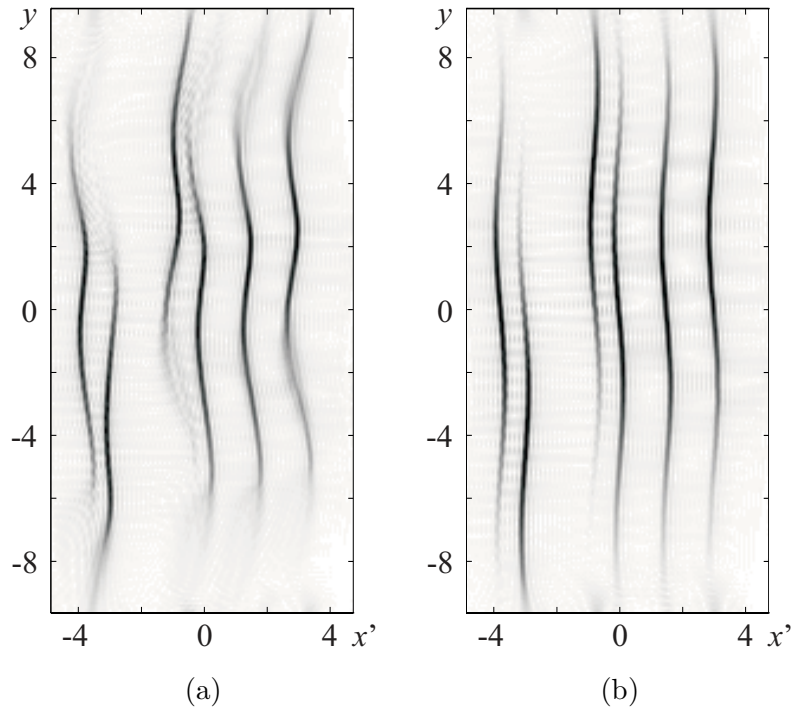
(a)   (b)

Figure 6.5   Testing the low-Q sway model. (a) the corrupted range-migration corrected data with the sway error removed via the inverse of (6.31), (b) the uncorrupted range migration corrected data with the sway error inserted as per (6.31). See the text for more details.

corrected data as per (6.31). If the model in (6.31) is accurate, then Fig. 6.5(a) should match the uncorrupted range migration corrected data in Fig. 6.4(e) and Fig. 6.5(b) should match the corrupted range migration corrected data in Fig. 6.4(b). The fact that both of these expectations are not seen in Fig. 6.5 indicates that (6.31) is *not* a good system model and an alternative model needs to be developed.

6.9   ADVANCED AUTOFOCUS TECHNIQUES

All of the autofocus procedures described in this chapter are based on the assumption that the motion error is well parameterized by a 1-D motion error, in fact, the spotlight mode autofocus algorithms are even more restrictive in that they treat this 1-D error as invariant over the whole scene. Isernia *et al*, 1996 [79, 80], present an autofocus method that is posed as a generalized phase retrieval problem. The method is capable of compensating for both 1-D and 2-D motion errors, and it has been validated on both real and simulated SAR data. Isernia *et al* claim that the algorithm compensates for both low- and high-frequency motion errors and that the algorithm does not require the presence of strong scatterers in the scene. The problem as posed by Isernia *et al* begins with the high-Q assumption that the motion error can be treated as purely a carrier phase effect, so its applicability to low-Q systems needs to be investigated.

In references [14, 53–55] researchers from Sandia National Laboratories present methods for phase

retrieval in other 1- and 2-D problems. Again, the systems discussed are high-Q, so that application to low-Q systems remains in question.

It is likely that as these advanced forms of autofocus gain wide use in the SAR community, that the SAS community will be quick to follow.

Chapter 7

RESULTS FROM THE KIWI-SAS PROTOTYPE SAS

This chapter begins with an overview of the Kiwi-SAS wide bandwidth, low-Q, synthetic aperture sonar system. Simulated images using the Kiwi-SAS parameters, and a general overview and summary of the capabilities of the synthetic aperture imaging algorithms, are given. Results from sea trials of the Kiwi-SAS system are then presented. The raw echo data obtained during these sea trials is relevant to system calibration and consists of the reflections from point-like targets placed in a harbour environment that had low background clutter. The application of autofocus procedures to the image estimates formed from this echo data produces diffraction limited imagery.

7.1 THE KIWI-SAS WIDE BANDWIDTH, LOW-Q, SYNTHETIC APERTURE SONAR

The development of the Kiwi-SAS system commenced in March, 1993, and culminated in the first sea trials in November, 1994. The system was developed 'in-house' in the University of Canterbury's Electrical and Electronic Engineering Department by this author, Peter Gough, and members of the technical staff; Art Vernon and Mike Cusdin. Figure 7.1 shows a photograph of the Kiwi-SAS towfish. Table 7.4 contains the important Kiwi-SAS system parameters, along with the parameters of the ACID/SAMI SAS and the Seasat-SAR [85] for comparison.

The towfish in Fig. 7.1 has been constructed to a blunt nosed, nose towed, neutrally buoyant design. This type of design and tow arrangement has been found to be particularly stable. The towfish is towed behind a vessel via 50m of reinforced cable. This cable is also the electronic connection to the surface vessel. The towfish housing contains; the transmitter power amplifiers, the low-noise receiver pre-amplifier, and the transmit and receive arrays. The transmitter and receiver amplifiers are based on standard design techniques (see for example Ott, 1988 [122]). Signals to and from the towfish are analog. Line drivers ensure a high common mode rejection ratio in the analog signals. Digital conversion and storage is performed at the surface. Minimal processing was performed on the received echoes, as one of
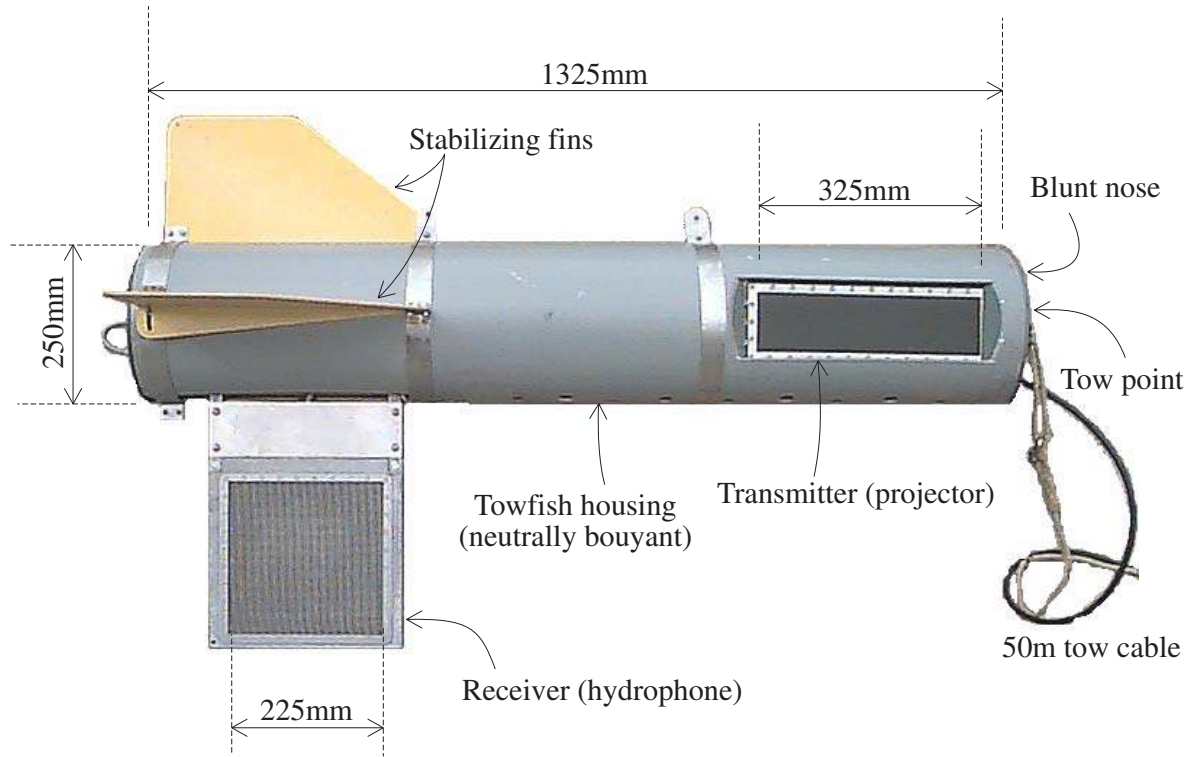
Figure 7.1 The Kiwi-SAS prototype synthetic aperture sonar. The towfish is a blunt nosed, nose towed, neutrally buoyant design. Weight in air; 38kg. The towfish houses the transmit power amplifiers, 12×3 array of Tonpilz projector elements (the black rectangle on the side of the towfish), 3×3 array of PVDF hydrophone elements (the square keel), low noise receiver pre-amp, and buoyancy material. Signals to and from the towfish are analog. Line drivers ensure a high common mode rejection ratio in the analog signals. Digital conversion and storage is performed at the surface.
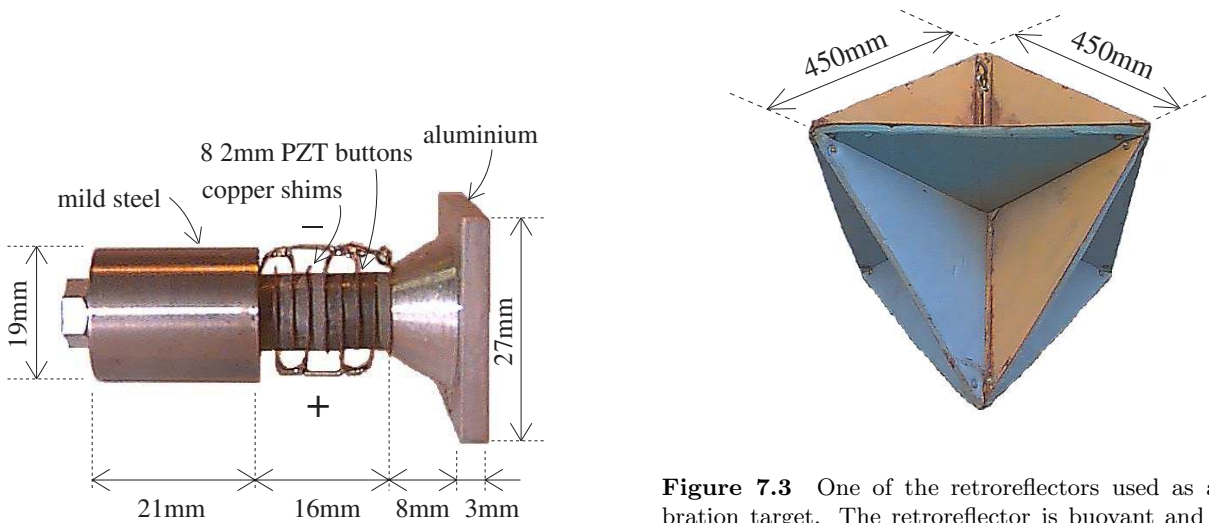
Figure 7.2 One of the wide bandwidth Tonpilz transducers that make up the projector array. The central driving stack is constructed of the piezoelectric ceramic lead-zirconate-titanate (PZT).

Figure 7.3 One of the retroreflectors used as a calibration target. The retroreflector is buoyant and is anchored to the harbour floor during testing. The target floats as it is oriented in this photo, in the 'rain-catching' mode (highest target strength). The retroreflector is constructed from three interlocking squares with sides of length 450mm.

the objectives of this thesis was to determine the optimum algorithm for processing synthetic aperture sonar data. The received echoes were simply bandpass filtered, digitized to 16 bits at a sampling rate of 100kHz, and stored.

7.1.1 Transmitter and receiver characteristics

The transmit aperture is a 12×3 array of multiresonant Tonpilz elements. The multiresonant Tonpilz design is a novel method of producing a transducer with an extremely wide bandwidth (low-Q), with high efficiency, and low pass-band ripple. Standard Tonpilz designs usually produce narrow bandwidth (high-Q) transducers. The multiresonant Tonpilz design was developed during the initial stages of the Kiwi-SAS project and is covered in [73] which is reproduced in this thesis as Appendix B. To minimize mechanical coupling between the elements in the array, the 36 transmitter elements were placed in cylindrical recesses machined into the nylatron (graphite impregnated nylon) backing material. A single Tonpilz element is shown in Fig. 7.2. The source level (the dB ratio of acoustic pressure produced by the transducer at 1m relative to 1μPa for an applied voltage of $1V_{\mathrm{rms}}$ [156]) of an element, and the overall array are shown in Fig. 7.4(a). The nominal source level per applied rms voltage of the transmit array is 160.4dB/V_{rms} re 1μPa$_{\mathrm{rms}}$@1m. The transmit array is usually operated at $76V_{p-p} = 26.9V_{\mathrm{rms}}$, thus it generates a nominal source level of SL=189.0dB re 1μPa$_{\mathrm{rms}}$@1m over the 20kHz to 40kHz bandwidth transmitted by the Kiwi-SAS.

The receive aperture is a 3×3 array of the piezoelectric polymer polyvinylidene diflouride (PVDF). Each receive element is a plate of 75mm×75mm 1500S series PVDF (three layers of 500μm thick PVDF sandwiched between two 250μm thick brass electrodes) [3]. The measured receive level (the dB ratio of voltage produced by the receiver relative to $1V_{\mathrm{rms}}$ for every 1μPa of pressure variation at 1m from the receiver) of this array is shown in Fig. 7.4(b). The nominal receive level (RL) of this array is RL=-197.6dB/1μPa$_{\mathrm{rms}}$@1m re $1V_{\mathrm{rms}}$.

Figure 7.5 shows the frequency varying response of the transmitter and receiver array when the nominal values of SL and RL are removed, i.e., the figure shows (SL-189.0dB)+(RL+197.6dB). The received signal power from an area of seafloor clutter plus the directivity should ideally match this test tank measured curve (with an irrelevant offset). A slight spurious response is seen in the received power signal. The origin of this slight notch is unknown, however, it is small and does not degrade the pulse compressed range response. Sonar calibration techniques are well covered in the book by Urick, 1975 [156].

7.1.2 Real aperture spatial-temporal response

Figure 7.6(a) is the theoretical spatial-temporal response (point spread function) of the real aperture Kiwi-SAS system (i.e., the spatial-temporal response after pulse compression, but *prior* to focusing).
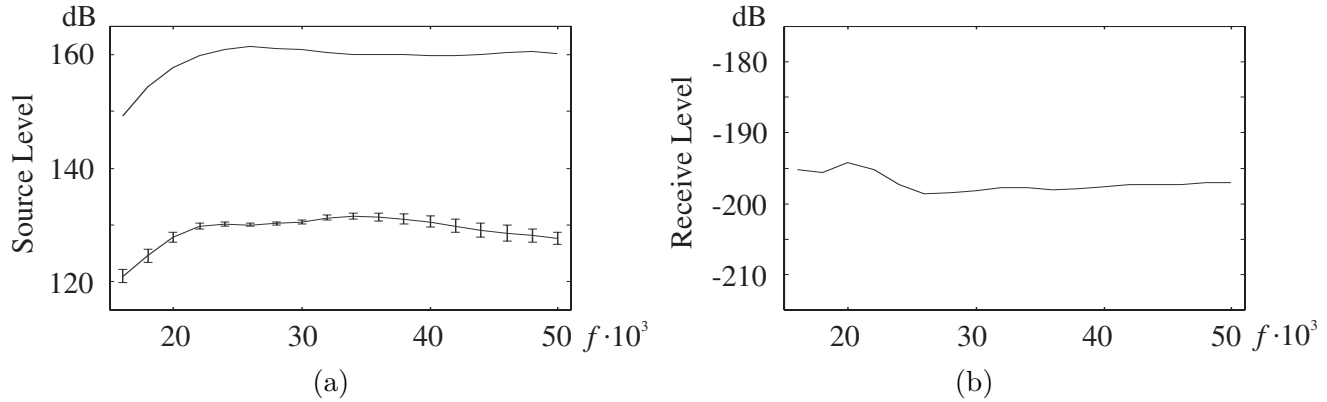
(a)   (b)

Figure 7.4 Transmitter source level per applied rms voltage and receiver receive level. (a) Transmitter $SL/V_{rms} \approx$ 160.4dB/V_{rms} re $1\mu Pa_{rms}$@1m (the dB ratio of acoustic pressure produced by the transducer at 1m relative to $1\mu Pa$ for an applied voltage of $1V_{rms}$), (b) Receiver RL ≈ -197.6dB/μPa_{rms}@1m re $1V_{rms}$ (the dB ratio of voltage produced by the receiver relative to $1V_{rms}$ for every $1\mu Pa_{rms}$ of pressure variation at 1m from the receiver).

Figures 7.6(b) and (c) show the theoretical range and along-track response, $\delta x_{3dB} = 5$cm and $\theta_{3dB} = 7°$ ($\theta_{6dB} = 9°$). The along-track response is described by the combined aperture *correlated beam pattern*, i.e., the spatial response in along-track is due to the coherent interference of the multiplicity of monochromatic radiation patterns. Also plotted on Fig. 7.6(c) is the monochromatic radiation pattern for the carrier frequency. One of the advantages of the wideband, low-Q, Kiwi-SAS system is that the -37dB sidelobe levels of the correlated (wide band) beam pattern are significantly lower than the -30dB sidelobes of the monochromatic (narrow band) pattern. If the Kiwi-SAS employed a transmitter and receiver of equivalent length then the monochromatic pattern sidelobes would be at -26dB, by employing apertures of different lengths, the nulls of one pattern suppress the sidelobes in the other. Monochromatic measurements of the radiation patterns of the transmitter and receiver arrays agreed well with theory (for example, see Fig. 8 in [73] or Appendix B).
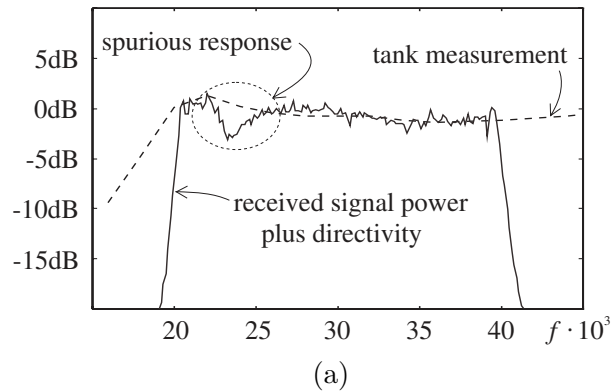
(a)

Figure 7.5   The frequency varying response of the transmitter and receiver array when the nominal values of SL and RL are removed, i.e., the figure shows (SL-160.4dB)+(RL+197.6dB). The figure also shows the received signal power from an area of seafloor clutter plus the directivity. This signal should ideally match the tank measured response (within an irrelevant offset), however, a spurious response is seen in the received power signal. Frequency dependent medium attenuation has been ignored.
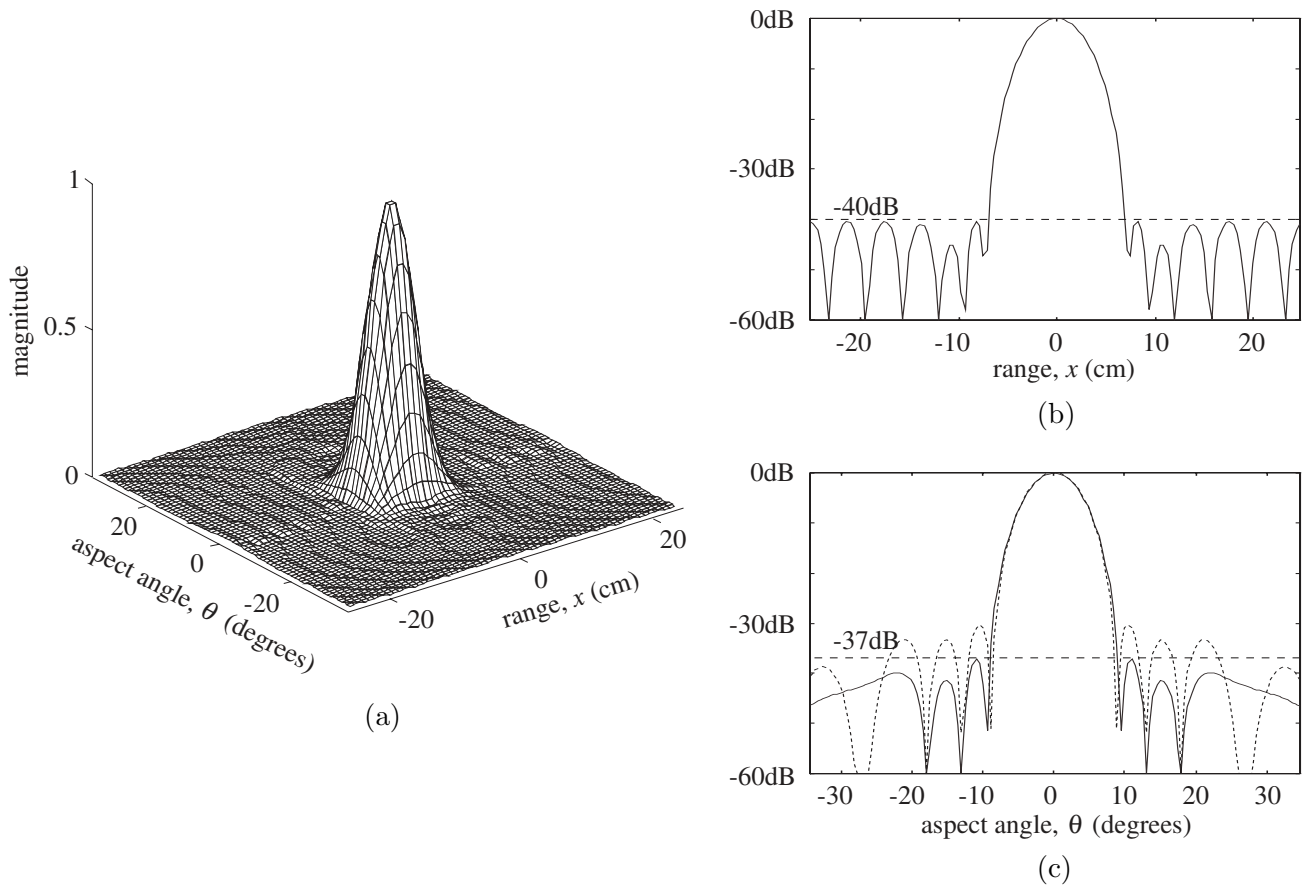
Figure 7.6   Spatial-temporal response (point spread function) of the Kiwi-SAS system (prior to focusing). (a) spatial-temporal response, (b) range response, $\delta x_{3dB} = \alpha_w \cdot c/(2B_c) = 5$cm (Hamming weighted in range), (b) angular response, $\theta_{3dB} = 7°$ ($\theta_{6dB} = 9°$). The coherent interference of the wide bandwidth radiation patterns produces a spatial response with sidelobes at -37dB. These sidelobes are 7 dB lower than the sidelobes in a monochromatic pattern and 11dB lower than the monochromatic sidelobes of an array with a transmitter and receiver of the same length.

7.2   A PICTORIAL VIEW OF SYNTHETIC APERTURE IMAGE PROCESSING

Figure 7.7 shows several of the domains of a simulated target field at various stages of the wavenumber algorithm. The objective of this figure is to demonstrate what is occurring during each step of the imaging algorithm. Figure 7.7(a) displays the baseband pulse compressed echoes, $ss_b(t', u)$, received from a target field containing 17 targets in an area 6m in range, 20m in along-track, located at $r_0 = 30$m from the towfish path. Range and range time are referenced from r_0 so that the data is in the correct format for the FFT (see Section 2.3). The orientation of targets in Fig. 7.7(a) has been chosen to emphasize the effect of the Stolt mapping process in the wavenumber domain. Figure 7.7(b) shows $SS'_b(\omega_b, k_u)$ the 2-D Fourier transform of the pulse compressed data. This data is deconvolved to give

$$GG'_b(\omega_b, k_u) = \sqrt{\frac{k}{k_0}} \cdot \exp\left[j\left(\sqrt{4k^2 - k_u^2} - 2k\right) \cdot r_0\right] \cdot SS'_b(\omega_b, k_u) \tag{7.1}$$

which is shown in Fig. 7.7(c). The effect of the Stolt mapping operation is clearly indicated in Fig. 7.7(c) and (d). Close inspection of the deconvolved signal shown in Fig. 7.7(c) reveals that the spectral lines parallel to the Doppler wavenumber axis are slightly curved. Fig. 7.7(d) shows that this curvature is removed by the Stolt mapping operation. Figure 7.7(e) shows the wavenumber domain after the processing bandwidths have been extracted and the effect of the radiation pattern has been deconvolved. Finally, Fig. 7.7(f) shows the image estimate obtained from the wavenumber estimate (Hamming weighting was applied in both range and along-track to reduce sidelobes in the final image estimate). The point targets shown in Fig. 7.7(f) are resolved to $\delta x_{3\text{dB}} \times \delta y_{3dB} = 1.30 \cdot \{c/(2B_c) \times D/2\} = 5cm\times$20cm.

Similar figures to those shown in Fig. 7.7 can be generated to explain the processing operations of the range-Doppler algorithm, the chirp scaling algorithm, and spatial-temporal domain or fast-correlation based processors. In the range-Doppler algorithm, the $\mathcal{T}\{\cdot\}$ operation in the (t, k_u)-domain produces a distortion in the (ω_b, k_u)-domain that is similar to the Stolt mapping and in the chirp scaling algorithm, the same (ω_b, k_u)-domain distortion is seen after the application of the chirp scaling multiplier, $\phi\Phi_1(t, u)$. In a fast-correlation based processor, each range line is generated by applying a frequency domain focusing filter to the spectrum of a block of data that is large enough to contain the required output range line and the range migration of any targets in that particular range line. The final image then, is built up on a range line by range line basis. A 2-D Fourier transform of the final image estimate obtained in this manner shows that the application of the range varying focusing parameters also removes the curvature in the wavenumber domain as seen in Fig. 7.7(c) (and in the process, creates a curved section of data in the wavenumber domain as seen in Fig. 7.7(d)).
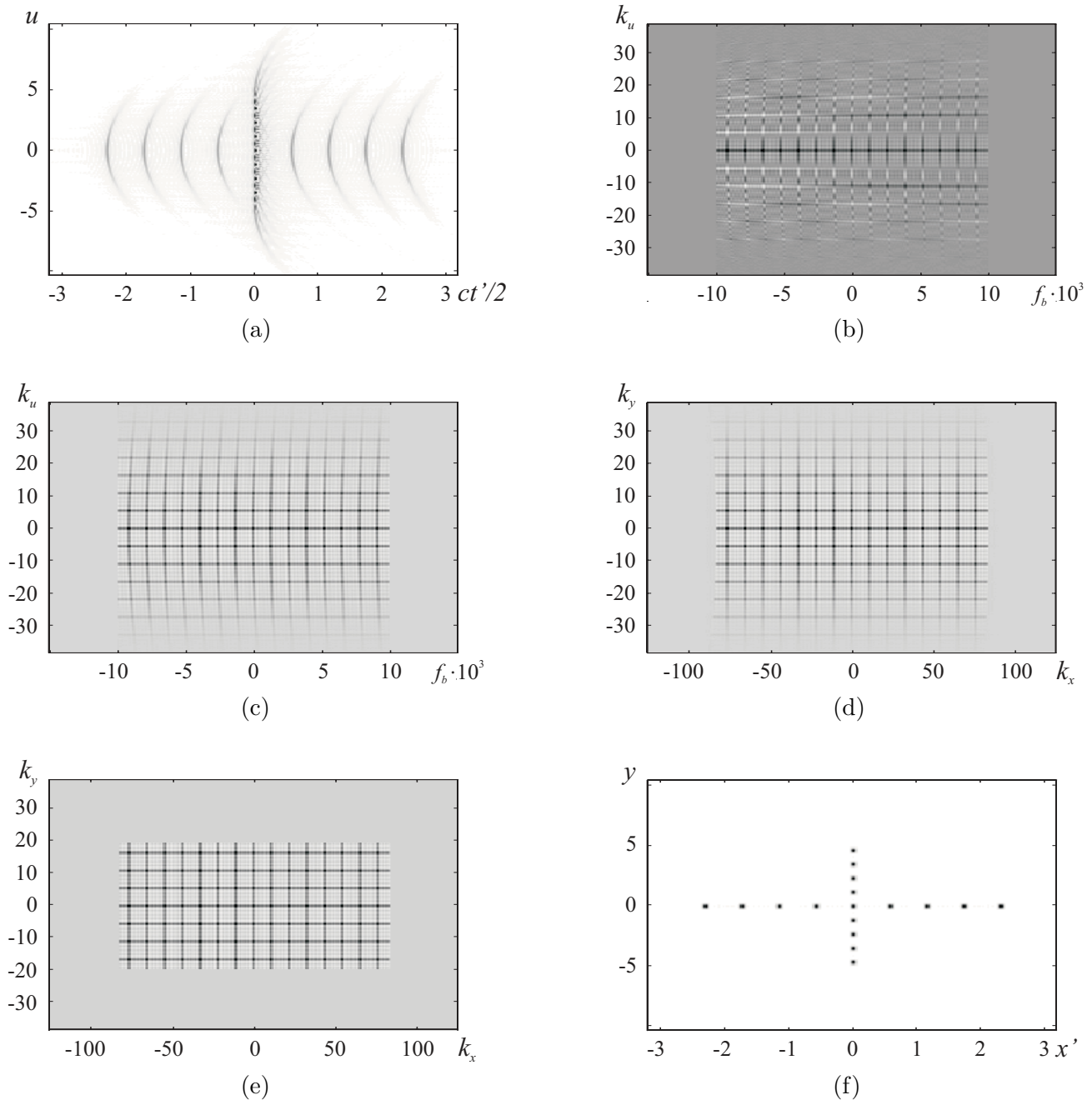
Figure 7.7 Synthetic aperture image processing domains. (a) the pulse compressed raw data $ss_b(t', u)$ (referenced to the standoff range $r_0 = 30$m), (b) the pulse compressed signal spectrum, $SS_b'(\omega_b = 2\pi f_b, k_u)$, (c) deconvolved and phase matched spectrum, (d) the Stolt mapped spectrum, i.e., the wavenumber domain estimate, (e) the extracted wavenumbers with the radiation pattern deconvolved, (f) final image estimate formed using a Hamming weighted version of the extracted wavenumber estimate. All images are linear greyscale; (a) and (f) are magnitude images while (b), (c), (d), and (e) are the real part of the signal.

7.3   "TO SRC, OR NOT TO SRC", THAT IS THE QUESTION.

The wide bandwidth wavenumber algorithm presented in Section 4.3.3 is *exact* in that the approximations used to develop it are good for any wavelength, bandwidth, beam width, or swath width. The same can also be said for the fast-correlation based processor presented in Section 4.3.1 that generates the image estimate on a range line by range line basis. The range-Doppler and chirp scaling algorithms of Section 4.3.2 and Section 4.3.4 both have inherent approximations and when the span of spatial bandwidths gets large enough or the system is highly squinted, the effect of these approximations needs to be corrected. The most well known of these correction factors is the secondary range compression (SRC).

Figure 7.8 and Table 7.1 demonstrate the applicability of the approximate range Doppler and chirp scaling algorithms, with and without secondary range compression (SRC), along with the exact fast-correlator and wavenumber algorithms. Figure 7.8 contains image domain and wavenumber domain images of a simulated point target, while Table 7.1 contains the value of the target peak and the maximum peak-to-sidelobe ratio (PSLR) calculated from the range or along-track cross-section of the target peak. The point target is slightly offset from the scene center $x' = 0$ which lies at a range of $r_0 = 30$m from the sonar track. The target is offset slightly so that its Fourier transform is a diffraction limited linear exponential representing the target shift from the x'-origin. The wavenumber domain images presented in Figure 7.8 represent the real part of the processed range and Doppler bandwidths that are passed by the processing algorithm. These images are simulated using Kiwi-SAS parameters, the along-track dimension has been sampled at $D/4$. After remapping the wavenumber data, each processor windows the wavenumber signals. The along-track bandwidth is windowed to $B_{k_y} = 4\pi/D$ and the amplitude effect of the overall aperture radiation pattern across this 3dB bandwidth is deconvolved. The range wavenumbers are windowed to $B_{k_x} \approx 4\pi B_c/c$, the range bandwidth is actually slightly less than the chirp bandwidth due to the distortion of the spatial frequency domain caused by the mapping operations. This wavenumber data is then normally Hamming weighted in both dimensions to reduce sidelobe levels in the final image estimate. The image estimates in Fig. 7.8 have not had this weighting applied so that the effects of the residual phase errors in the system are obvious.

The images displayed in Fig. 7.8 represent a progression, in terms of image quality, from worst to best. The worst algorithm being range Doppler or chirp scaling without SRC, and the best being the fast-correlation and wavenumber algorithms. The range Doppler and wavenumber algorithms used an interpolator based on a weighted sinc truncated to 8 one-sided zero crossings (this interpolator is described in Section 2.9 and Jakowatz *et al* [82]).

Figure 7.8(a) is the image produced using the range Doppler or chirp scaling algorithm *without* SRC. The image is displayed on a logarithmic greyscale with 50dB dynamic range. By looking at the real part of the wavenumber spectrum of this point target as shown in Fig. 7.8(b), it is obvious that a
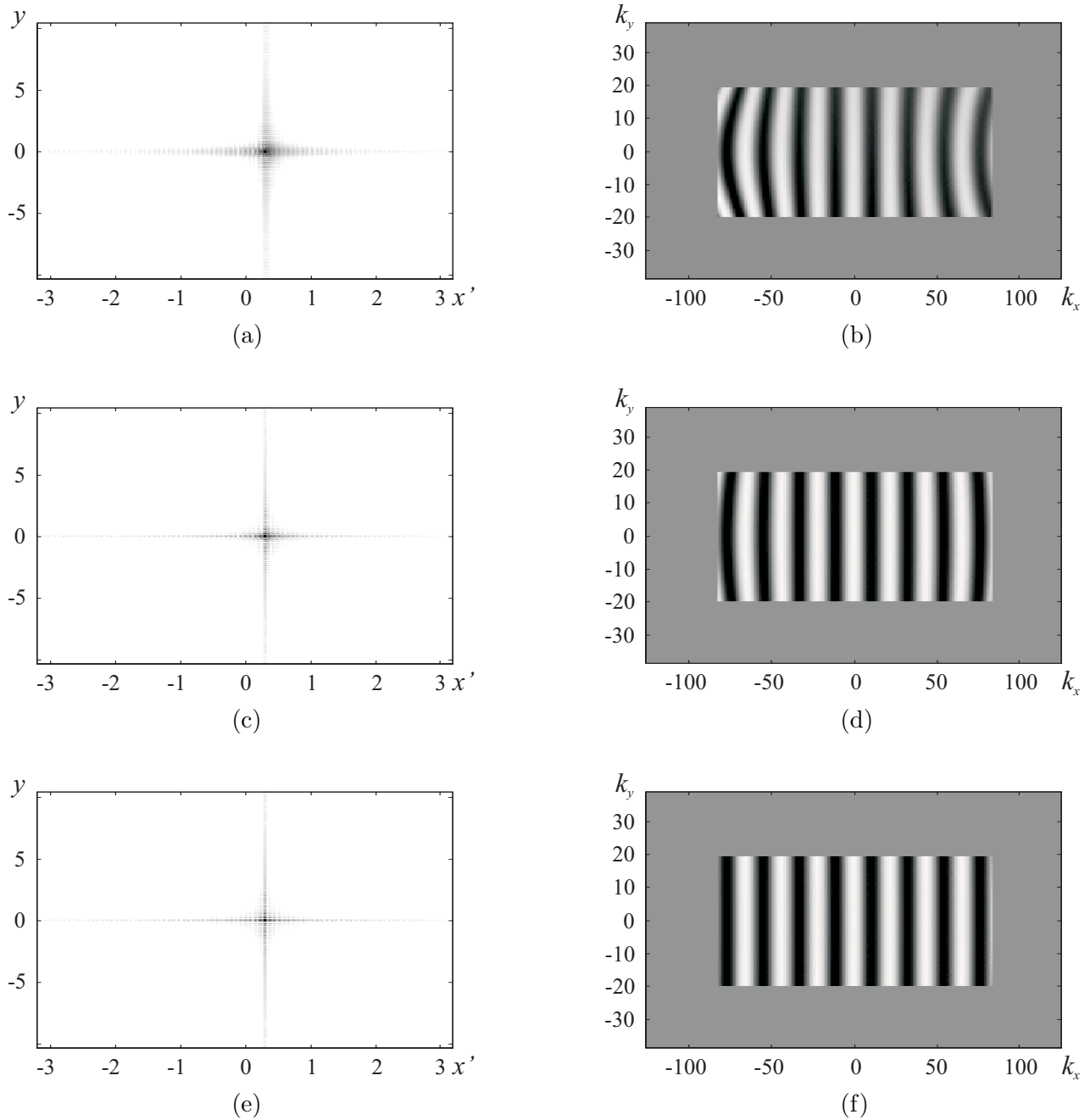
Figure 7.8 Spatial and wavenumber response of a simulated point target produced using (a) the range-Doppler or chirp scaling algorithm without secondary range compression (SRC), (b) the range-Doppler or chirp scaling algorithm with SRC, and (c) the fast-correlation or wavenumber algorithm. The spatial images are logarithmic greyscale with 50dB dynamic range, and the wavenumber images show the real part of the signal on a linear greyscale.

residual phase function still exists. Table 7.1 shows that this residual phase causes a decrease in peak height and PSLR, Fig. 7.8(a) also shows the increase in sidelobe levels. These two effects act to reduce the image dynamic range. Note that this residual phase function has a minimal impact over the central region of the k_x-dimension of the wavenumber domain, so this effect would not be observed in much narrower spatial bandwidth SAR systems. Secondary range compression is required in squinted SAR systems, it is seldom used in broadside operation. SRC removes a k_y (or k_u) dependent LFM term that is induced by the system geometry, closer inspection of Fig. 7.8(b) shows that no distortion is occurring near $k_y = 0$. Figures 7.8(c) and (d) are the resulting images after the application of SRC. Careful inspection of Fig. 7.8(d) at the edges of the k_x-domain show that a residual phase function still remains. This phase remains because the SRC term is calculated for a fixed range r_0. In SAR systems this approximation typically sets a swath limit over which the chirp scaling and range Doppler algorithms are applicable (eg. see p485 Carrara, 1995 [21]). The SRC term employed in extended chirp scaling (ECS) for application to highly squinted systems is determined using a higher order polynomial [36,110]. If the residual phase function was more of a problem for the Kiwi-SAS then this technique could be employed. However, the results of Table 7.1 indicate that the calculation of SRC at one range value is unlikely to limit the Kiwi-SAS system, in fact if Hamming weightings are applied over the Kiwi-SAS wavenumber data, Tab. 7.1 shows that the compensation of SRC may not be necessary. So the question needs to be asked; why use SRC at all as it would appear to add an unnecessary computational load?

The answer lies in the *phase* of the diffraction limited image. Normally this is of little interest as only the modulus or intensity of the image is displayed but if *any* form of autofocus is needed to correct for aperture errors or medium turbulence, this phase is critical whether we can *see* it in the uncorrected diffraction limited image or not. Bathymetric (interferometric) applications also require phase stability for accurate height estimation. It is important in both autofocus and bathymetric applications to have the residual phase errors due to the image formation algorithm as small as possible and as SRC is a deterministic phase correction, little is lost by correcting for it during image formation.

7.4   ALGORITHM SUMMARY

The wide bandwidth wavenumber algorithm represents the *exact* inversion of the synthetic aperture system model. If high order interpolators are used, then the final image is the "best" possible of all possible images. Spatial-domain or fast-correlation processing are both very inefficient. Spatial-domain processors require either interpolators or highly oversampled signals to process the data into focused images. Implementation of focusing via fast-correlation does not yield a dramatic increase in efficiency as the focusing filter has to be calculated for each range line. The range-Doppler and chirp scaling algorithms represent an approximate inversion of the system model and with the inclusion of SRC, the approximation has only a minimal impact on the final image quality. The range-Doppler and

Table 7.1 Simulated point target image metrics

Algorithm	Unweighted		Hamming weighted	
	Peak height	PSLR*	Peak Height	PSLR*
Range-Doppler without SRC	0.91	-12dB	0.99	-43dB
Range-Doppler with SRC	0.99	-13dB	1.00	-40dB
Chirp scaling without SRC	0.92	-12dB	0.99	-43dB
Chirp scaling with SRC	1.00	-13dB	1.00	-40dB
Wavenumber	1.00	-13dB	1.00	-43dB
Fast-correlation	1.00	-13dB	1.00	-43dB

* Peak-to-sidelobe ratio, maximum value for range or along-track

wavenumber algorithms perform comparably. This is because both algorithms require an interpolator to operate on the same volume of data; however, the *accelerated* chirp scaling algorithm out performs both of these algorithms. When a Hamming weighting function is applied to the extracted image wavenumber estimate, all of the algorithms produce identical images (even without SRC, the loss in peak height and the rise in sidelobe levels is minor) for the Kiwi-SAS parameters.

Similar comments regarding the phase stability and efficiency of these algorithms are found in references [6, 21, 36, 110, 128].

7.5   KIWI-SAS CALIBRATION AND SEA TRIAL RESULTS

7.5.1   Target strength estimation via one-dimensional range imaging

To systematically calibrate the Kiwi-SAS system it was necessary to determine the expected received voltage level given the measured system parameters and the available theoretical parameters. The calibration procedure is reviewed in this section using a specific example; that of a retroreflector located at 35.8m from the towfish track.

Figure 7.9 and Table 7.2 show the effect of the transmitter, target, and receiver transfer functions on a chirped waveform. The voltage and pressure decibel units in Table 7.2 are relative to $1V_{\mathrm{rms}}$ and $1\mu Pa_{\mathrm{rms}}$. The chirp rate of the transmitted waveform is negative and the signal is chirped from 40kHz to 20kHz in 50ms, so it has a time-bandwidth product of 1000 ($IF_{\mathrm{pc}} = 30$dB). The amplitude modulating effects of the transmitter are seen in Fig. 7.9(a) and (b). The uniform pulse waveform of $26.9V_{\mathrm{rms}}$ in

Fig. 7.9(a) is applied to the transmitter terminals to produce the $2.8\text{kPa}_{\text{rms}}$ pressure waveform shown in Fig. 7.9(b). The frequency dependent source level in Fig. 7.4(a) has a lower response around 40kHz, so the components at the higher frequencies of the pressure wave shown in Fig. 7.9(b), i.e., those at $t < 0$, are slightly smaller. The phase of the overall transmit transducer could not be measured due to equipment limitations, however, the transfer function of a single transducer element showed a fairly linear phase characteristic across the 20kHz to 40kHz band. Thus, it was assumed that the phase of the transmit transducer would not unduly effect the transmitted signal modulation. Since the pulse compressed real signals correctly compress to the theoretical range resolution of 5cm, this assumption is justified.

After the pressure waveform in Fig. 7.9(b) has undergone spherical spreading out to the target at 35.8m, part of the signal is scattered back toward the receiver. The amount of signal back-scattered is quantified by the targets' frequency dependent *cross-section* in units of meters squared or by the targets' *target strength* in units of decibel. The scattered and attenuated pressure waveform at a reference range of 1m from the receiver is referred to as the *echo level* (EL) (in dB re $1\mu\text{Pa}_{\text{rms}}$@1m) where

$$\text{EL} = \text{SL} - 2\text{TL} + \text{TS} \tag{7.2}$$

This equation can be considered to be frequency dependent, or alternatively nominal parameters can be used. The one-way transmission loss accounting for spherical spreading and medium attenuation is

$$\text{TL} = 10 \log_{10}\left(\frac{x^2}{r_{\text{ref}}^2}\right) + \gamma x \tag{7.3}$$

where the frequency dependent attenuation coefficient γ accounts for medium absorption [156, p103]. The target strength of a retroreflector is

$$\text{TS} = 10 \log_{10}\left(\frac{\sigma_s}{4\pi r_{\text{ref}}^2}\right) = 10 \log_{10}\left(\frac{d^4}{3\lambda^2 r_{\text{ref}}^2}\right) \tag{7.4}$$

where the target cross-section is

$$\sigma_s = \frac{4\pi d^4}{3\lambda^2} \tag{7.5}$$

The target strength of the retroreflector has been assumed to be the same as the target strength of a trihedral reflector [21, p345] [156, p276] [31, p338]. The reference range $r_{\text{ref}} = 1\text{m}$ and the length d is the length of an internal side, i.e., $d = 450\text{mm}/\sqrt{2} = 0.318\text{m}$ for the retroreflector shown in Fig. 7.3. The nominal target strength of the retroreflector is calculated at the carrier wavelength as $\text{TS} = 1.3\text{dB}$ ($\sigma_s = 17\text{m}^2$). Fig. 7.9(c) shows how the frequency dependence of the target cross-section

has imposed a further amplitude modulation on the scattered pressure waveform.

Target strengths are often positive quantities, this does not imply that signal strength is increased, it is purely a consequence of referencing the dB quantity to the surface area of a sphere of unit radius [156, p264]. In radar the equivalent dB quantity is referenced to a sphere of unit surface area giving the units dBsm [21]. The target cross-section, σ_s, indicates the ratio of the intensity of the reflected pressure wave to the intensity of the incident wave, so after scattering, the scattered waveform amplitude is reduced by $\sqrt{\sigma_s}$. In terms of the 1-D range imaging models of Chapter 2 the *target reflectivity* is

$$f(x) = \sqrt{\sigma_s(x)} \tag{7.6}$$

and for the 2-D analysis in the next section,

$$ff(x, y) = \sqrt{\sigma_s(x, y)} \tag{7.7}$$

The reflected pressure wave impinging on the receiver face is converted back to a voltage waveform according to the calibrated receive level (RL = -197.6dB/1μPa$_{\text{rms}}$@1m re $1V_{\text{rms}}$). This voltage waveform is amplified by 40dB in the low noise receiver pre-amplifier and is transmitted up the tow cable to be received by a further adjustable amplifier (set to 36dB for this example). The expected voltage waveform at the output of this surface amplifier is shown in Fig. 7.9(d). The higher sensitivity of the receiver at the lower frequencies (see Fig. 7.4(b)) causes a slight increase in the amplitudes of the lower frequency components relative to their level in the pressure waveform. The 16-bit analog-to-digital converter that sampled the received voltage waveform had a dynamic range of $2V_{\text{rms}}$, this level is indicated in Fig. 7.9(d). Slight clipping of the received waveform occurs. Pulse compression of this clipped waveform produces a signal with a peak of 35.8dB re $1V_{\text{rms}}$ (clipping reduces the peak magnitude by 0.8dB).

Figure 7.10(a) shows the signal received and sampled by the Kiwi-SAS system from a target located 38m from the towfish path. This sequence of data consists of three reflected pulses obtained during the collection of synthetic aperture data. These three pulses are located at the range of closest approach of the target locus shown in Fig. 7.11. This signal is used so that the target strength estimated in this section can be compared to the target strength estimated from the focused target.

The pulse compressed version of the received signal is shown in Figure 7.10(b). The largest peak of the three pulses is 29.9dB re $1V_{\text{rms}}$. Using the formulae in Table 7.2, the estimated target strength is

$$\begin{aligned} \text{TS}_{\text{est}}^{\text{1D}} &= \text{V}_{\text{pc}}\text{dB - (SL - 2TL + G + IF}_{\text{pc}} + \text{RL)} \\ &= -5.2\text{dB} \end{aligned} \tag{7.8}$$

where the transmission loss was estimated at TL = 31.1dB and the amplifier gain was G = 76dB. This estimated target strength differs by 6.6dB to the nominal target strength predicted by (7.4), i.e., the

estimated target reflectivity is different by a factor that is very close to 2. This result seems surprising given that the received voltage waveform in Fig. 7.10 looks very similar to that predicted in Fig.7.9. The following energy considerations show that the difference is not due to an error during processing: The energy in a real valued voltage signal $h_r(t)$ is

$$E = \int_t |h_r(t)|^2 dt \approx \Delta_t \sum_n |h_r[n]|^2 \tag{7.9}$$

where the real samples $h_r[n] = h_r(t = n\Delta_t)$, $\Delta_t = 1/f_s$, n is the sample indice, and the energy is calculated for power relative to a 1 ohm reference impedance. The energy in a signal represented by its complex form $h_c(t)$ is

$$E = \frac{1}{2} \int_t |h_c(t)|^2 dt \approx \frac{\Delta_t}{2} \sum_n |h_c[n]|^2 \tag{7.10}$$

where the complex samples $h_c[n] = h_c(t = n\Delta_t)$. The energy measured in the simulated pulse and the compressed version is 0.19J (the measurement is made over the chirp length τ_c). Given that the range of the target is known from the pulse compressed data, the energy for time τ_c about the target is calculated as 0.11J. The same temporal samples in the raw data yield 0.12J. The raw data signal has slightly more energy due mainly to the cross-talk signal. No factor of two shows up in the energy considerations, however, the received signal energy is less than predicted by 2.4dB. Indicating that although the received and simulated voltage levels look similar, the underlying energy content is different. (From Fig. 7.10(b) it is obvious that the energy estimate for the pulse compressed data also includes the sea floor clutter about the target location).

In summary, the estimated target strength of -5.2dB differs from the predicted level of 1.4dB due to an unknown source and not processing errors. The most likely cause of the measured difference is poor target orientation, or alternatively, this particular target *has* a lower target strength than that predicted by theory. The estimated target strength of a second retroreflector target in the same data record was -4.1dB.

7.5.1.1   Cross-talk attenuation factor

Direct cross-talk of the transmitted signal to the receiver limits the dynamic range of pulsed and CTFM sonar systems. As wide bandwidth signals are often employed, it is difficult to baffle this direct path. Figure 7.10(b) can be used to estimate the level of cross-talk attenuation in the Kiwi-SAS system. From Fig. 7.10(b), the measured pulse compressed cross-talk level is CT = 27.4dB. If the transmitter and receiver were to face each other at the reference distance of 1m (this is close to the distance from the center of the receiver to the transmitter; see Fig. 7.1), then the signal after pulse compression would be
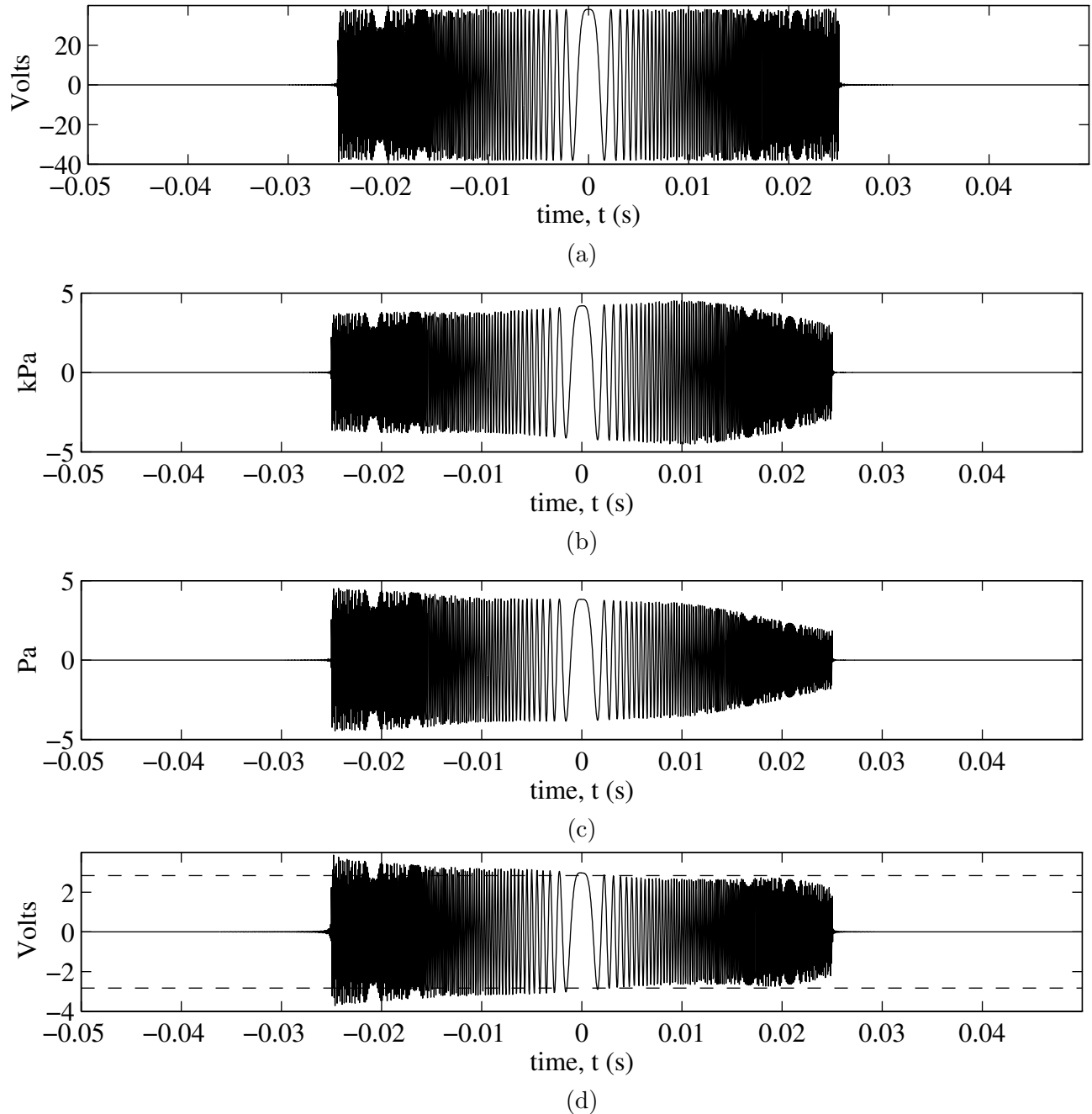
Figure 7.9 The expected voltages and pressures of a simulated pulse that has reflected off a retroreflector at a range of 35.8m. (a) transmitted voltage waveform, (b) pressure wave 1m from the transmitter, (c) pressure wave 1m from the receiver due to target backscatter, and (d) the voltage at the output of the surface amplifier and the dashed line shows the sampled dynamic range (some signal clipping occurs).
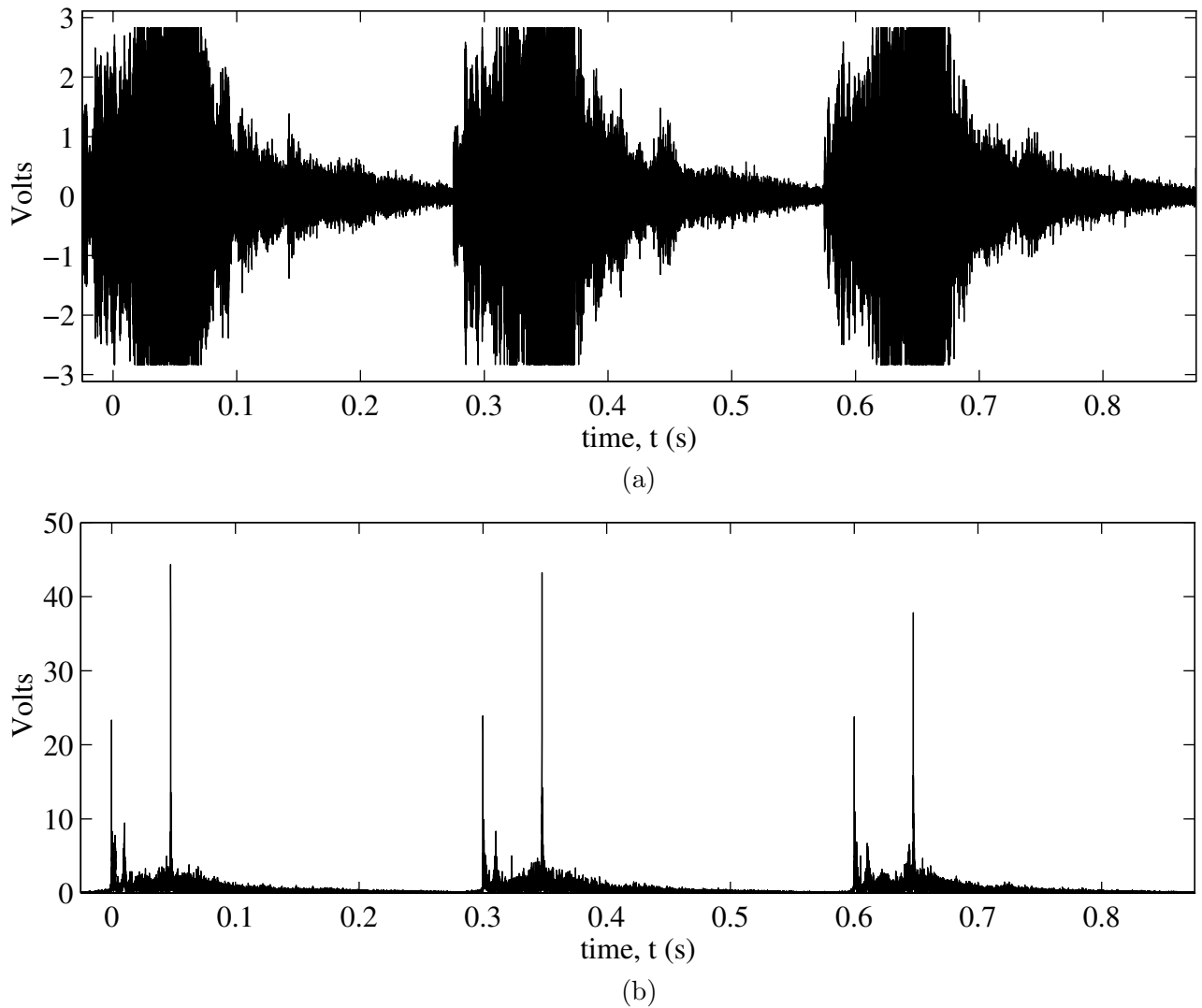
Figure 7.10 Received backscatter signal from a retroreflector at a range of 35.8m. (a) sampled voltage waveform (clipping has occurred), (b) pulse compressed waveform; small peak is the cross-talk, large peak is the target. Time t is referenced to the phase center of the transmitted pulse, the chirp length $\tau_c = 50$ms, the repetition period $\tau_{\mathrm{rep}} = 300$ms, and the chirp bandwidth $B_c = 20$kHz. No Hamming weighting is applied in range (the addition of Hamming weighting did not improve the TS estimate).

Table 7.2   Nominal, simulated, and measured signal levels from a retroreflector at a range of 35.8m.

	decibel symbol/formula	nominal/ simulated		measured	
Applied voltage	$V_{app}dB$	$26.9V_{rms}$	28.6dB	$26.9V_{rms}$	28.6dB
Transmit pressure@1m	$SL=V_{app}dB+160.4dB$	$2.8kPa_{rms}$	189dB	$2.8kPa_{rms}$	189dB
Receive pressure@1m	$EL = SL-2TL+TS$	$2.6Pa_{rms}$	128.2dB	-	-
Surface amp. output voltage	$V_{out}dB=EL+RL+G$	$2.1V_{rms}$	6.6dB	-	-
Pulse compressed voltage (clipped signal)	$V_{pc}dB=V_{out}dB+IF_{pc}$	$95.7V_p/$ $87.1V_p$	36.6dB/ 35.8dB	$44.3V_p$	29.9dB

$SL + RL + G + IF_{pc}$. The *cross-talk attenuation* is then,

$$CA = (SL + RL + G + IF_{pc}) - CT$$
$$= 70dB$$

(7.11)

Therefore the radiation patterns of the transmitter and receiver, and the orientation of the receiver on the towfish are such that the cross-talk level is considerably smaller than the maximum possible level.

7.5.2   Target strength estimation via synthetic aperture processing

For an ideal reflecting target, synthetic aperture processing increases the target peak by the synthetic aperture improvement factor, IF_{sa}, given in (4.51). Figure 7.11 shows the pulse compressed locus and image estimate of a retroreflector located at 35.8m from the towfish path. The peak of the pulse compressed locus was used in the 1-D range imaging section and is at 29.9dB re $1V_{rms}$. The synthetic aperture improvement factor for a target located at 35.8m is 15.3dB. Ideally then, the peak of Fig. 7.11(b) will be at 45.2dB re $1V_{rms}$. Processing the 3dB bandwidth, $B_p = 4\pi/D$, yielded a target strength estimate of -9dB, while processing only $B_p = \pi/D$ yielded a target strength estimate of -5dB. The result from the smaller processed bandwidth is close to that estimated purely from range imaging. The origin of the disparity between the two measurements is obvious when the locus in Fig. 7.11(a) is plotted as a 3-D surface. Where a simulated locus is smooth, the recorded locus shows dips and spurious peaks; specular-like reflections have occurred. The recorded locus is smoothest along the range of closest approach of the sonar to the target, so the target strength estimate obtained from the reduced processing bandwidth is more accurate.
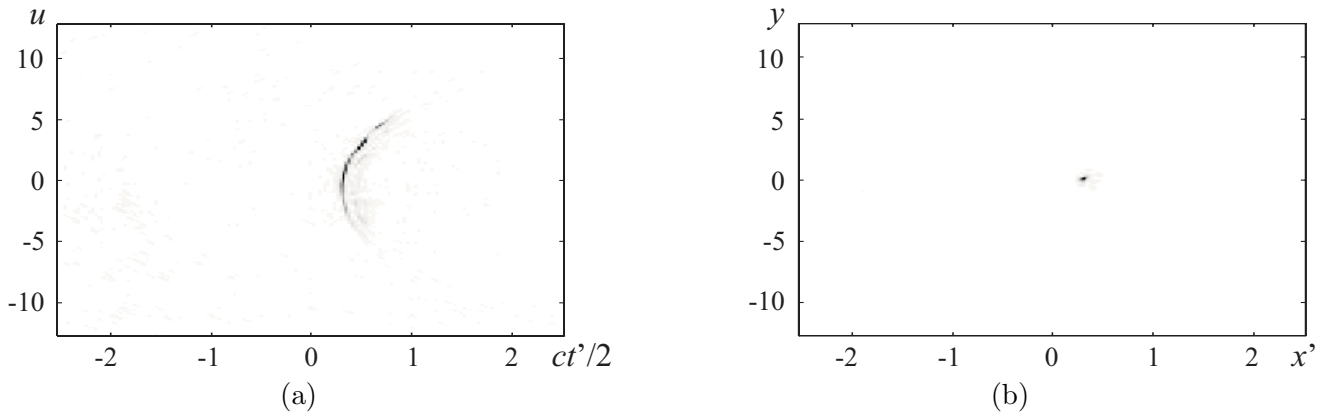
Figure 7.11   Synthetic aperture processing of a retroreflector located at a range of 35.8m. (a) Pulse compressed signal and (b) image estimate obtained from a Hamming weighted version of the wavenumber domain estimate. Both images represent backscatter cross-section, i.e., the figures are $|ss_b(t', u)|^2$ and $\sigma_s(x', y) = |\widehat{ff}_b(x', y)|^2$. Both images are linear greyscale.

As the two target strength estimates obtained from this particular target are consistent between the two methods of estimation, it is likely that the difference between the predicted target strength and the measured target strength is due to the scattering properties of the target used for calibration. Target strength estimates obtained from other echo records and from other similar retroreflectors also yielded estimates in the range -9dB to -5dB. Curlander and McDonough [31, p339] indicate quite a large variability in the measured *radar* cross-section of this form of target (a trihedral-type reflector).

The data shown in Figure 7.11 was processed using the fast-correlation, the range-Doppler, the wavenumber, and the chirp scaling algorithm. Each algorithm produced the same target strength estimate.

7.5.3   Geometric resolution

The results of the geometric calibration of Fig. 7.11(b) (which is shown as a 3-D surface plot in Fig. 7.14(a)) are displayed in Table 7.3. Although near diffraction limited results have been obtained, the application of autofocus improves the initial image estimate.

7.6   WIDE BANDWIDTH, LOW-Q, AUTOFOCUS RESULTS

The windowing of the wavenumber domain before the production of the image estimate in Fig. 7.11(b) has masked an important observation. What appears to be a well focused point, was in fact two closely spaced peaks. These two peaks are clearly shown in the unweighted image estimate shown in

Fig. 7.13(a). This image estimate is plotted as a 3-D surface to clearly show the peak splitting. The improvement that occurs when the data is autofocused is seen in Fig. 7.13(b). The reason for the split in the main peak is explained by way of Fig. 7.12.

Figures 7.12(a) and (b) show the range-Doppler domain and the *phase* of wavenumber domain estimate of a simulated point target at the same location as the retroreflector in Fig. 7.11. The sinc-like shape of the unweighted image estimate causes the range-Doppler estimate to be sinc-like in x and rect-like in k_y. The phase of the wavenumber domain is plotted to emphasize the effects of motion in the real data. The phase is used in the wavenumber domain images, and not the real part as was done for Fig. 7.8, as the non-uniform spectral magnitude of the real data masks the underlying effects of the motion. The phase tilt seen in Fig. 7.12(b) is due to the target offset from both the x and y origins. Fig. 7.12(d) shows how this expected phase tilt has been corrupted by the motion error. The motion error is clearly seen by comparing Figs. 7.12(a) and (c); the simulated target response lies parallel to the k_y-axis, whereas the recorded data has a (mostly) linear slant. When an inverse Fourier transform is performed on the along-track dimension for an *unweighted* estimate, the slanted response collapses to two peaks. When the inverse transform is preceded by a weighting operation, the components that cause the peak to split are suppressed and a single peak forms. The slanting effect of the error seen in Fig. 7.12(c) is similar to the *range walk* experienced by spaceborne SARs. It was most likely caused by lateral towfish movement during aperture formation (probably due to cross-currents). Regardless of the origin of this motion, the autofocusing of this data via a prominent point processing-type autofocus algorithm removes the range slant and produces the well focused point target shown in Fig. 7.13(b). The straightened range-Doppler response is seen in Fig. 7.12(e) and the phase of the wavenumber domain in Fig. 7.12(f) is now predominantly composed of a slant due to target offset from the ordinate origins. The effect of the autofocus on the phase of the real target is seen in Fig. 7.12(d) and (f); relative to the phase in Fig. 7.12(b), the phase of Fig. 7.12(d) starts out slanted too much, then progresses to being almost straight, however, after autofocusing, the phase of Fig. 7.12(f) is similar to that of the simulated target.

7.6.1   Target strength estimation and geometric resolution of the autofocused data

A comparison of Fig. 7.13(a) to Fig. 7.13(b) shows that not only has autofocus improved the geometric resolution of the image, it has improved the estimate of the scattering cross-section (both images are normalized to the estimated cross-section of Fig. 7.13(b)). Figures 7.14(a) and (b) are image estimates formed from a weighted version of the wavenumber domain estimate that formed the images in Fig. 7.13. Though the use of weighting has reduced the image geometric resolution slightly, it has improved the scattering estimate (the images in Fig. 7.14 are normalized to the estimated cross-section of Fig. 7.14(b)). Wavenumber domain weighting acts to suppress image artifacts caused by the spectral discontinuities

Table 7.3   Geometric and scattering results from a retroreflector located at a range of 35.8m.

	units	Initial image estimate		Autofocused image estimate	
		no weighting Fig. 7.13(a)	weighting Fig. 7.14(a) Fig. 7.11(b)	no weighting Fig. 7.14(a)	weighting Fig. 7.14(b)
Range resolution, $\delta x_{3\mathrm{dB}}$	cm	6.4	5.6	3.8	4.9
Along-track res., $\delta y_{3\mathrm{dB}}$	cm	16.6	28.0	16.3	21.1
Range PSLR	dB	0	-23	-13	-22
Along-track PSLR	dB	0	-15	-9	-18
Target strength, TS	dB	-13.1	-8.8	-8.7	-6.9
Target cross-section, σ_s	m^2	0.6	1.7	1.7	2.6

at the edges of the useful along-track and range wavenumber bandwidths. Artifacts due to spectral discontinuities degrade both the geometric resolution (due to high sidelobes) and the scattering estimates in calibrated synthetic aperture images.

Table 7.3 shows the geometric and scattering parameters estimated from the images shown in Fig. 7.11(b), Fig. 7.13(a) and (b), and Fig. 7.14(a) and (b). Though the initial image estimate produced *near* diffraction limited imagery, the autofocused image *is* diffraction limited. For the Kiwi-SAS system, the finest theoretical resolution possible in an unweighted image estimate is $\delta x_{3\mathrm{dB}} \times \delta y_{3dB} = 3.8\mathrm{cm} \times 16.3\mathrm{cm}$ and for a Hamming weighted estimate this becomes $\delta x_{3\mathrm{dB}} \times \delta y_{3dB} = 4.9\mathrm{cm} \times 21.1\mathrm{cm}$. Both of these limits are achieved in the autofocused images. The scattering parameters obtained from the well focused synthetic aperture images are consistent with each other, and they are consistent with the estimates obtained from the 1-D range imaging method.

In summary, this section has demonstrated the requirement for autofocus and spectral weighting in a fully calibrated synthetic aperture system. When these processes are included in the synthetic aperture processing suite, diffraction limited imaging is possible.
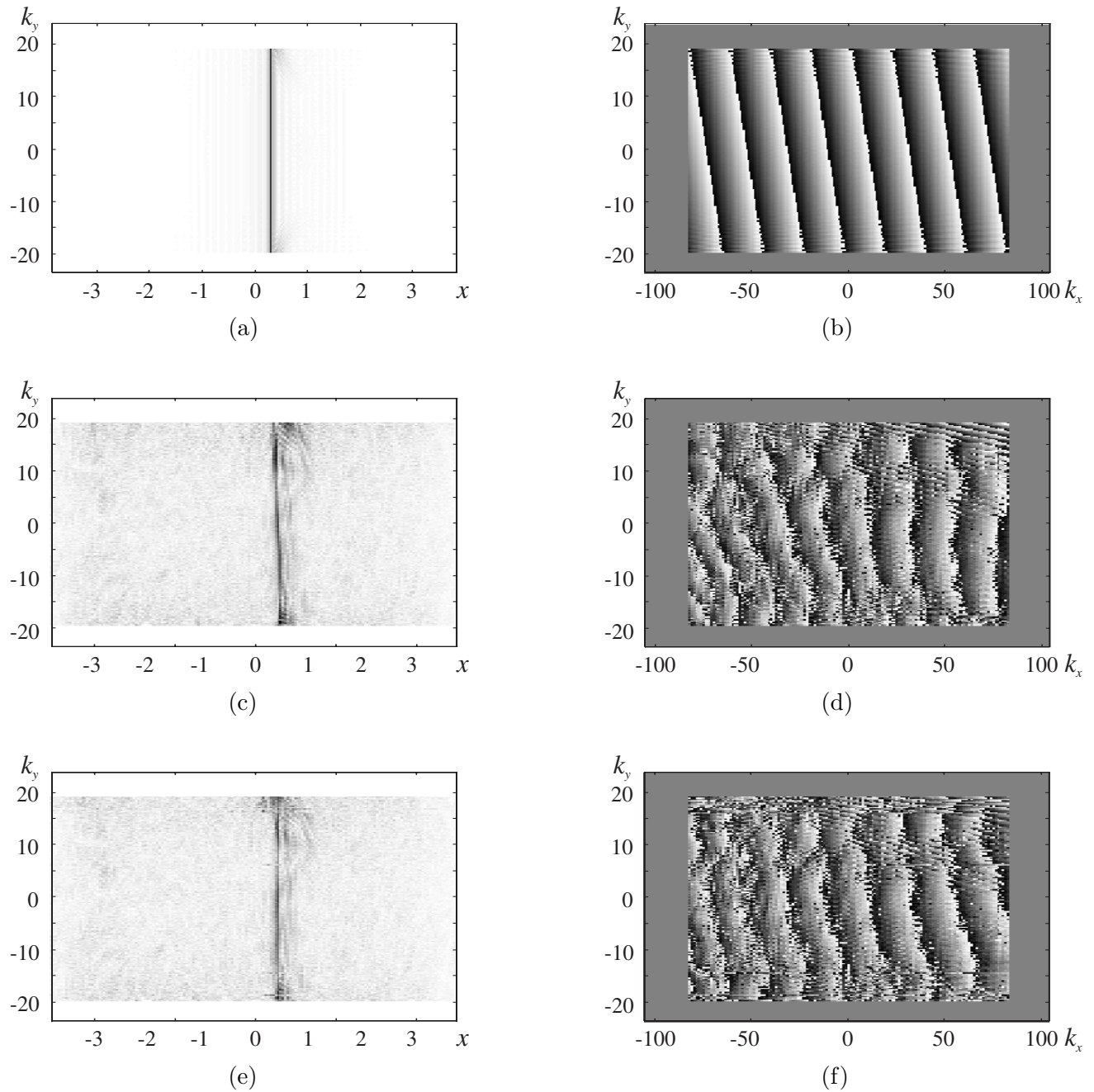
Figure 7.12   Autofocusing via prominent point processing of a retroreflector located at a range 35.8m. Range-Doppler and wavenumber domain estimates for; (a) and (b) a simulated target, (c) and (d) for the initial image estimate, and (e) and (f) for the image estimate after autofocus. The range-Doppler images are linear greyscale, and the wavenumber images show the phase of the signal on a linear greyscale. The scene x-origin lies at a range of 35.5m from the towfish path.

(a)

(b)
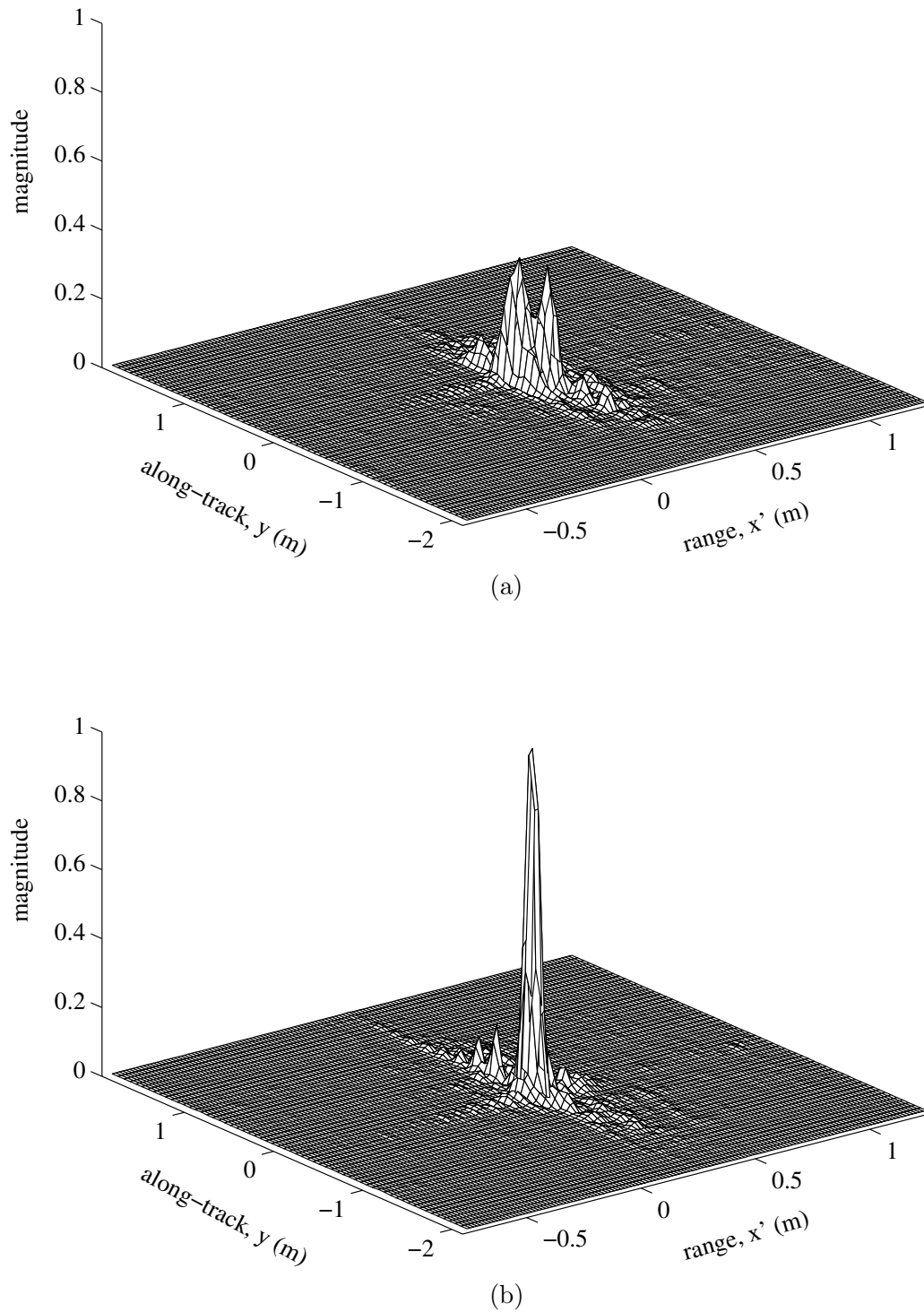
Figure 7.13 Synthetic aperture image estimates. (a) initial image estimate, (b) image estimate after autofocusing. No weighting was applied to the wavenumber domain. Geometric and scattering parameters are shown in Table 7.3. Both images are normalized to the cross-section of (b) (see Table 7.3).
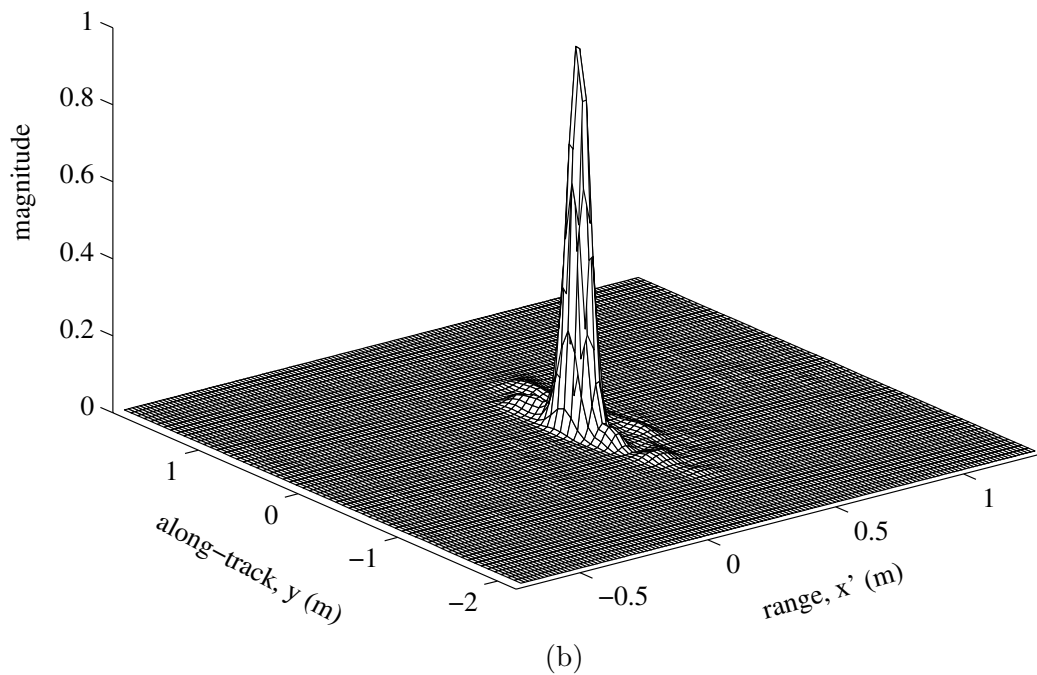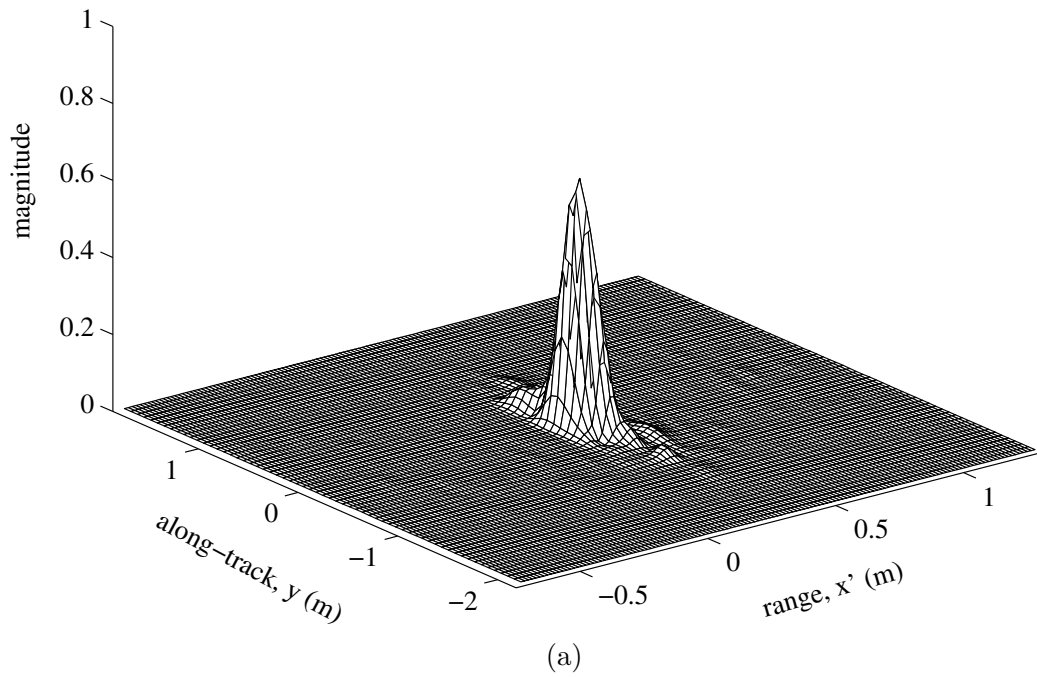
Figure 7.14 Synthetic aperture image estimates. (a) initial image estimate, (b) image estimate after autofocusing. Hamming weighting was applied to the wavenumber domain. Geometric and scattering parameters are shown in Table 7.3. Both images are normalized to the cross-section of (b) (see Table 7.3).

Table 7.4   System and nominal operating parameters for the Kiwi-SAS, ACID/SAMI SAS, and the Seasat-SAR [85].

Parameter	Symbol	Units	Kiwi-SAS	ACID/SAMI	Seasat
Wave propagation speed	c	m/s	$1.5 \cdot 10^3$	$1.5 \cdot 10^3$	$3.0 \cdot 10^8$
Platform speed	v	m/s	0.4-0.9	1-1.6	7400
Carrier frequency	f_0	Hz	$30 \cdot 10^3$	$8 \cdot 10^3$	$1.3 \cdot 10^9$
Carrier wavelength	λ_0	cm	5	18.7	23.5
Bandwidth	B_c	Hz	$20 \cdot 10^3$	$6 \cdot 10^3$	$20 \cdot 10^6$
Pulse repetition period	τ_{rep}	s	0.15-0.30	0.4-0.8	$650 \cdot 10^{-6}$
Swath width	X_s	m	100-200	200-400	$100 \cdot 10^3$
Pulse length	τ_c	s	0.05	0.02	$33 \cdot 10^{-6}$
Time-bandwidth product	$\tau_c B_c$	-	1000	120	630
Effective aperture length	D	m	0.3	2.0	10.7
Beamwidth at f_0	θ_{3dB}	degrees	9.5	5.4	1.2
Grazing angle	ϕ_g	degrees	5	30-50	80
Standoff range	r_0	m	50-100	100-200	$850 \cdot 10^3$
Ground-plane resolution	$\delta x_{g\mathrm{3dB}} \times \delta y_{\mathrm{3dB}}$	m	0.05×0.20	0.16×1	25×25
Number of looks	-	-	1	1	4

Chapter 8

CONCLUSIONS AND DISCUSSION

This thesis has presented the complete end-to-end simulation, development, implementation and calibration of the wide bandwidth, low-Q, Kiwi-SAS synthetic aperture sonar. Through the use of a very stable towfish, a new novel wide bandwidth transducer design, and autofocus procedures, high-resolution diffraction limited imagery was produced. As a complete system calibration was performed, this diffraction limited imagery is not only geometrically calibrated, it is also calibrated for target cross-section or target strength estimation. Is is important to note that the diffraction limited images were formed *without* access to any form of inertial measurement information.

The synthetic aperture processing algorithms presented in this thesis are unique in that they are presented in their *wide bandwidth forms*. No published SAR or SAS system uses the same spatial bandwidth as the Kiwi-SAS, so the derivation of the wide bandwidth versions of these algorithms has not previously been necessary. These wide bandwidth versions differ from the previous derivations in that deconvolution filters are necessary to remove wide bandwidth amplitude modulating terms. The compensation of these terms allows for the proper calibration of the final image estimate. This wide bandwidth treatment also introduced the relatively new field of *Fourier array imaging* and showed that the approximations made in classical array theory are not necessary in situations where the spatial information across the face of the array is available. Specifically, classical array theory should not be applied to phased or synthetic arrays when imaging targets in the far-field of the elements that make up each array. The wide bandwidth radiation patterns obtained via Fourier array theory showed up the $D/2$ along-track sampling myth that has been propagated through the SAR and SAS community. The correct interpretation of grating lobe suppression indicates that $D/4$ is the limit imposed. However, in a trade-off with along-track ambiguity to signal ratio (AASR), a relaxed along-track sample spacing of $D/2$ is often used in conjunction with a reduction in the processed along-track wavenumber bandwidth. A thorough investigation of the wide bandwidth synthetic aperture imaging algorithms has shown that the *phase* estimates produced by the algorithms described in this thesis are suitable for bathymetric (interferometric) applications. This investigation points out that secondary range compression (SRC)

is necessary even in unsquinted geometries when the spatial bandwidths used by the system become large. The use of a unified mathematical notation and the explicit statement of any change of variables via mapping operators assists in the clear presentation of the available processing algorithms and their direct implementation via digital processing.

This thesis has also increased the efficiency of the chirp scaling algorithm. The *accelerated* chirp scaling algorithm is elegant in that it minimizes the processing overhead required to focus synthetic aperture data, yet the whole algorithm consists of simple multiplies and Fourier transform operations, both of which are easily implemented on general purpose hardware. These features will make it the processor of choice in future SAR and SAS systems.

The holographic properties of synthetic aperture data were discussed and multilook processing was reviewed. The ease with which sonar systems can produce wide wavenumber bandwidths in range due to the low speed of sound indicated that SAS can use multilook processing in range to trade-off range resolution for speckle reduction. In fact, if the more non-linear multilook techniques are employed, minimal loss of image resolution occurs.

An indepth analysis of the published methods of mapping rate improvement led to the conclusion that any synthetic aperture sonar of commercial value will have to employ a single transmitter in conjunction with multiple receive apertures in a vernier array arrangement, i.e., separately channeled and unphased. The ACID/SAMI SAS, the UCSB-SAS, the Alliant SAS and the MUDSS SAS are already employing this type of arrangement. Combining multiple apertures with the wide bandwidth transmit transducer of the Kiwi-SAS will produce an extremely high resolution sea floor imaging system that can satisfy the operational requirements of the end-user as well as satisfy the sampling requirements of the synthetic aperture. An analysis of frequency coded waveforms also pointed out a little known fact; that coded pulses must have identical carrier/center frequencies. If this requirement is not adhered to, no gain in along-track sampling is possible. When codes with equivalent carrier frequencies are used, the dynamic range of the final image estimate is limited by the cross-correlation properties of the transmitted coded pulses. Wide bandwidth waveforms were analysed and their grating lobe smearing properties were accurately quantified by the peak-to-grating lobe ratio (PGLR). The use of wide bandwidth waveforms to relax the along-track sampling requirement is not recommended. The final image suffers a loss of dynamic range due to the smeared grating lobes in the undersampled image.

The diffraction limited imagery produced in the results chapter of this thesis emphasizes the fact that autofocus and wavenumber domain weighting are an integral part of the calibration of image geometric and scattering properties. The analysis of many of the SAR autofocus algorithms provided a new form of phase gradient autofocus (PGA)—modified phase gradient autofocus is applicable to low-Q spotlight systems that do not satisfy the restrictions of the tomographic formulation of the spotlight mode. The strip map version of PGA, phase curvature autofocus (PCA), was investigated for its application to low-Q systems and it was shown how the algorithm failed for these systems. This thesis also used

a low-Q autofocus procedure, based on prominent point processing, that produced diffraction limited imagery for the Kiwi-SAS system.

In summary, this thesis contains a comprehensive review and development of synthetic aperture processing and post-processing for mapping applications using wide bandwidth systems; radar or sonar.

8.1   FUTURE RESEARCH

Further calibration sea trials need to be performed. These new sea trails would ideally map a scene containing a variety of man-made targets. In a bland scene it is often difficult to quantify image improvement, if targets are placed in the scene, then image improvement is obvious. At the same time as this data is collected, a high resolution, real aperture, side-looking sonar (SLS) system should be used to image the same scene. SAR has an advantage in that it can compare aerial photographs to the final image product. Sonar does not have this option; another sonar needs to be used. An image comparison of a synthetic aperture image to the image produced from the side-looking sonar will emphasize the image clarity available in a synthetic aperture image.

The development of autofocus algorithms for low-Q SAS requires further investigation. The phase retrieval algorithms of Isernia *et al* [79, 80] are creating a stir in the SAR community and should be investigated for their application to SAS.

8.1.1   Kiwi-SAS II

Hardware limitations of this project restricted the Kiwi-SAS to having a single receiver channel. In the second prototype, multiple receivers should be implemented. Each of the PVDF elements in the current receiver are slightly less than one-quarter the length of the transmitter. If the nine elements are constructed into a nine channel receiver in the along-track direction, then the along-track sampling requirement would be easier to satisfied for practical towing speeds. For example, a 100m swath could be mapped at a tow speed of 5m/s (10 knots) with the sample spacings of the synthetic aperture at $D_T/4$ producing well sampled data. The far-field of the receiver array would begin at a range of 18m, so no modification to the inversion scheme is required to account for near-field effects (this form of modification was necessary for the high-Q UCSB-SAS).

In terms of data acquisition, digital hardware could convert the echoes received on each channel into data packets that could be piped to the surface via a single channel using standard networking protocols. This would be an ideal system, as any number of computers could be attached to the receiving "network". One computer could be assigned to storage, while others produced real-time SAS, or at least a preliminary side-looking sonar image.

Appendix A

THE PRINCIPLE OF STATIONARY PHASE

For integrands having wide phase variation and with an envelope $g_0(u)$ [28, p39],

$$
\int_u g_0(u) \exp\{j\phi(u)\} du \approx \left(-\frac{2\pi}{\phi''(u^*)}\right)^{1/2} \cdot \exp\left\{-j\frac{\pi}{4}\right\} \cdot g_0(u^*) \exp\{j\phi(u^*)\}
$$
$$
= \sqrt{\frac{j2\pi}{\phi''(u^*)}} \cdot g_0(u^*) \exp\{j\phi(u^*)\}
$$

(A.1)

where the *stationary point*, u^*, is the solution of,

$$
\phi'(u^*) = 0
$$

(A.2)

under the assumption that the derivative with respect to u of the phase is single valued, or has only one value of physical significance.

A.1   DERIVATION OF THE WAVENUMBER FOURIER PAIR

The proof of the Fourier pair,

$$
\mathcal{F}_u\left\{\exp\left(-j2k\sqrt{x^2+(y-u)^2}\right)\right\} \approx \sqrt{\frac{\pi x}{jk}} \cdot \exp\left(-j\sqrt{4k^2-k_u^2}\cdot x - jk_u y\right)
$$
$$
= \boldsymbol{A}_1 \cdot \exp[j\psi(k_u)]
$$

(A.3)

is as follows;

$$
\mathcal{F}_u\left\{\exp\left(-j2k\sqrt{x^2+(y-u)^2}\right)\right\} = \int_u \exp\left(-j2k\sqrt{x^2+(y-u)^2} - jk_u u\right) du
$$

(A.4)

The phase and derivatives of (A.4) are,

$$\phi(u) = -2k\sqrt{x^2 + (y-u)^2} - k_u u$$

$$\phi'(u) = \frac{2k(y-u)}{\sqrt{x^2 + (y-u)^2}} - k_u \tag{A.5}$$

$$\phi''(u) = -\frac{2kx^2}{[x^2 + (y-u)^2]^{3/2}}$$

Solving $\phi'(u)$ for the stationary point gives,

$$u^* = y - \frac{k_u x}{\sqrt{4k^2 - k_u^2}} \tag{A.6}$$

Substituting this back into $\phi(u)$ gives the phase of the transform as,

$$\psi(k_u) \equiv \phi(u^*) = -\sqrt{4k^2 - k_u^2} \cdot x - k_u y \tag{A.7}$$

which is the phase term on the RHS of (A.3).

The complex amplitude function is given by,

$$\begin{aligned}
\boldsymbol{A}_1 &= \sqrt{\frac{j2\pi}{\phi''(u^*)}} \\
&= \sqrt{\frac{8\pi k^2 x}{j(4k^2 - k_u^2)^{3/2}}} \\
&\approx \sqrt{\frac{\pi x}{jk}}
\end{aligned} \tag{A.8}$$

where the approximation is valid for wide bandwidth and low-Q system parameters.

A.2   RANEY'S RANGE-DOPPLER SIGNAL

This section contains those 'easy to show' steps that Raney and others skip when developing the range-Doppler model that is used as the basis of the chirp scaling algorithm.

For the transmission of a LFM waveform of chirp rate K_c (Hz/s) with a uniform envelope over

$t \in [-\tau_c/2, \tau_c]$, the appropriate baseband return from a point target at range x, is given by [126–128],

$$
\begin{aligned}
ee_\delta(t, u) = & a\left(t - \frac{2}{c}\sqrt{x^2 + u^2}, x, -u\right) \odot_t \text{rect}\left(\frac{t - \frac{2}{c}\sqrt{x^2 + u^2}}{\tau_c}\right) \\
& \cdot \exp\left\{j\pi K_c\left(t - \frac{2}{c}\sqrt{x^2 + u^2}\right)^2\right\} \cdot \exp\left\{-j2k_0\sqrt{x^2 + u^2}\right\}
\end{aligned}
\tag{A.9}
$$

where it is important to note that Raney's derivation uses a LFM waveform that is modulated by $-K_c$, not the $+K_c$ shown here.

The temporal Fourier transform of (A.9) is calculated using the following phase and derivatives,

$$
\begin{aligned}
\phi(t) &= \pi K_c\left(t - \frac{2}{c}\sqrt{x^2 + u^2}\right)^2 - \omega_b t \\
\phi'(t) &= 2\pi K_c\left(t - \frac{2}{c}\sqrt{x^2 + u^2}\right) - \omega_b \\
\phi''(t) &= 2\pi K_c
\end{aligned}
\tag{A.10}
$$

the stationary point is at

$$
t^* = \frac{\omega_b}{2\pi K_c} + \frac{2}{c}\sqrt{x^2 + u^2}
\tag{A.11}
$$

giving the transform phase as,

$$
\psi(\omega_b) \equiv \phi(t^*) = -\frac{\omega_b^2}{4\pi K_c} - 2k_b\sqrt{x^2 + u^2}
\tag{A.12}
$$

The complex constant in the temporal Fourier transform is due to the transformation of the chirp,

$$
\boldsymbol{A}_1 = \sqrt{\frac{j}{K_c}}
\tag{A.13}
$$

The temporal transform of (A.9) is then [126, 127]

$$
Ee_\delta(\omega_b, u) = \sqrt{\frac{j}{K_c}} \cdot A(\omega, x, -u) \cdot \text{rect}\left(\frac{\omega_b}{2\pi K_c \tau_c}\right) \cdot \exp\left\{-j\frac{\omega_b^2}{4\pi K_c} - j2k\sqrt{x^2 + u^2}\right\}
\tag{A.14}
$$

where the transform contains both baseband and modulated quantities.

The spatial Fourier transform of (A.14) is obtained using (A.3),

$$
EE_\delta(\omega_b, k_u) = \sqrt{\frac{\pi x}{k K_c}} \cdot A(k_u) \cdot \text{rect}\left(\frac{\omega_b}{2\pi K_c \tau_c}\right) \cdot \exp\left\{-j\frac{\omega_b^2}{4\pi K_c} - j\sqrt{4k^2 - k_u^2} \cdot x\right\}
\tag{A.15}
$$

The required range-Doppler signal is obtained from this 2-D Fourier signal by performing an inverse temporal Fourier transform. The following expansion of the phase term of (A.15) is necessary before it is possible to perform this inverse transform [126, 127]. The terms within the square root in the phase of (A.15) may be written as

$$
\begin{aligned}
\sqrt{4k^2 - k_u^2} &= \sqrt{4(k_0^2 + 2k_0 k_b + k_b^2) - k_u^2} \\
&= 2k_0 \left[\left(1 - \frac{k_u^2}{4k_0^2}\right) + \left(\frac{2k_b}{k_0} + \frac{k_b^2}{k_0^2}\right) \right]^{1/2} \\
&= 2k_0 \left(1 - \frac{k_u^2}{4k_0^2}\right)^{1/2} \left[1 + \frac{\left(\frac{2k_b}{k_0} + \frac{k_b^2}{k_0^2}\right)}{\left(1 - \frac{k_u^2}{4k_0^2}\right)} \right]^{1/2}
\end{aligned}
\tag{A.16}
$$

Letting,

$$
\beta = \sqrt{1 - \frac{k_u^2}{4k_0^2}}
\tag{A.17}
$$

and expanding the second square root term of (A.16) using $(1+\epsilon)^{1/2} = 1 + \epsilon/2 - \epsilon^2/8$ and keeping only terms up to k_b^2 we get

$$
\begin{aligned}
\sqrt{4k^2 - k_u^2} &\approx 2k_0 \beta \left[1 + \frac{1}{2\beta^2}\left(\frac{2k_b}{k_0} + \frac{k_b^2}{k_0^2}\right) - \frac{1}{8\beta^4}\left(\frac{2k_b}{k_0}\right)^2 \right] \\
&= 2k_0 \beta + \frac{2k_b}{\beta} + \frac{k_b^2}{k_0 \beta} - \frac{k_b^2}{k_0 \beta^3} \\
&= \frac{2\omega_0 \beta}{c} + \frac{2\omega_b}{c\beta} + \frac{\omega_b^2}{\omega_0 c\beta} - \frac{\omega_b^2}{\omega_0 c\beta^3}
\end{aligned}
\tag{A.18}
$$

The phase function required to determine $eE_\delta(t, k_u)$ is

$$
\begin{aligned}
\phi(\omega_b) &= -\frac{\omega_b^2}{4\pi K_c} - \sqrt{4k^2 - k_u^2} \cdot x + \omega_b t \\
&\approx -\frac{\omega_b^2}{4\pi K_c} - x\left(\frac{2\omega_0 \beta}{c} + \frac{2\omega_b}{c\beta} + \frac{\omega_b^2}{\omega_0 c\beta} - \frac{\omega_b^2}{\omega_0 c\beta^3}\right) + \omega_b t \\
&= \omega_b^2 \left(-\frac{1}{4\pi K_c} - \frac{x}{\omega_0 c\beta} + \frac{x}{\omega_0 c\beta^3}\right) + \omega_b\left(t - \frac{2x}{c\beta}\right) - \frac{2\omega_0 x\beta}{c}
\end{aligned}
\tag{A.19}
$$

The expansion contained within this phase function, (A.18), corresponds to a quadratic approximation to the Stolt mapping of the wavenumber algorithm. The multiplication of this expansion by the range parameter x eventually produces the range dependent secondary range compression term. In SAR

systems employing high squint, or in cases where the range dependence of the SRC term becomes significant, a higher order expansion is necessary [36, 110, 128]

The stationary point of the phase function in (A.19), i.e., $\phi'(\omega_b^*) = 0$, is,

$$\omega_b^* = \frac{\left(t - \frac{2x}{c\beta}\right)}{2\left[\frac{1}{4\pi K_c} + \frac{x}{\omega_0 c}\left(\frac{\beta^2 - 1}{\beta^3}\right)\right]} \tag{A.20}$$

so that the range-Doppler phase is given as,

$$\psi(t) \equiv \phi(\omega_b^*)$$

$$= \frac{1}{4\left[\frac{1}{4\pi K_c} + \frac{x}{\omega_0 c}\left(\frac{\beta^2 - 1}{\beta^3}\right)\right]} \cdot \left(t - \frac{2x}{c\beta}\right)^2 - \frac{2\omega_0 x\beta}{c} \tag{A.21}$$

$$= \pi K_s(k_u; x) \cdot \left[t - \frac{2}{c}R_s(k_u; x)\right]^2 - \sqrt{4k_0^2 - k_u^2} \cdot x$$

where the range migration through (t, k_u)-space is given by [110, 126–128]

$$R_s(k_u; x) = \frac{x}{\sqrt{1 - \left(\frac{k_u}{2k_0}\right)^2}} \tag{A.22}$$

$$= x[1 + C_s(k_u)]$$

where the *curvature factor*,

$$C_s(k_u) = \frac{1}{\sqrt{1 - \left(\frac{k_u}{2k_0}\right)^2}} - 1 \tag{A.23}$$

describes the Doppler wavenumber dependent part of the signal trajectory [128]. The new k_u dependent chirp rate is given by [110, 127, 128]

$$K_s(k_u; x) = \frac{1}{1/K_c - K_{src}(k_u; x)} \tag{A.24}$$

where the additional range dependent chirp term,

$$K_{src}(k_u; x) = \frac{8\pi x}{c^2} \cdot \frac{k_u^2}{(4k_0^2 - k_u^2)^{3/2}} \tag{A.25}$$

is compensated during processing by *secondary range compression* (SRC) [110,128]. Note that this range distortion is not a function of the FM rate: rather, it is a function of the geometry. Range distortion

is a direct consequence only of the lack of orthogonality between "range" and "along-track" for signal components away from zero Doppler (i.e. $k_u = 0$), and applies to any form of range modulation, not just to LFM [128].

The amplitude functions of $eE_\delta(t, k_u)$ are now evaluated. Using,

$$\phi''(\omega_b^*) = -2 \left(\frac{1}{4\pi K_c} + \frac{x}{\omega_0 c\beta} - \frac{x}{\omega_0 c\beta^3} \right) \tag{A.26}$$

gives,

$$\boldsymbol{A}_3 = \frac{1}{2\pi} \cdot \sqrt{\frac{-j2\pi}{2\left(\frac{1}{4\pi K_c} + \frac{x}{\omega_0 c\beta} - \frac{x}{\omega_0 c\beta^3} \right)}} \approx \sqrt{\frac{K_c}{j}} \tag{A.27}$$

The $1/(2\pi)$ is from the definition of an inverse Fourier transform with respect to ω (i.e., it is the Jacobian of the transform between the variables f and $\omega = 2\pi f$). Inside the envelope of the $\text{rect}(\cdot)$ in (A.15), the stationary point is approximated as,

$$\omega_b^* \approx \frac{t - \frac{2}{c}R_s(k_u; x)}{\frac{1}{2\pi K_c}} \tag{A.28}$$

Raney's range-Doppler signal is then given as,

$$\begin{aligned}
eE_\delta(t, k_u) \approx &\sqrt{\frac{\pi x}{jk_0}} \cdot A(k_u) \cdot \text{rect}\left(\frac{t - \frac{2}{c}R_s(k_u; x)}{\tau_c} \right) \\
&\cdot \exp\left\{ j\pi K_s(k_u; x) \cdot \left[t - \frac{2}{c}R_s(k_u; x) \right]^2 \right\} \\
&\cdot \exp\left\{ -j\sqrt{4k_0^2 - k_u^2} \cdot x \right\}
\end{aligned} \tag{A.29}$$

where the amplitude function is only given approximately by its carrier wavenumber value in the range-Doppler domain, i.e., $\sqrt{\pi x/(jk)}$ transforms approximately to $\sqrt{\pi x/(jk_0)}$. A deconvolution operation during processing in the frequency domain can correct for this approximation.

A.3   THE CHIRP SCALING ALGORITHM

The following subsections derive the phase multiplies required by the chirp scaling algorithm. These phase multiplies are developed using the range-Doppler point target model developed above.

A.3.1 The chirp scaling phase multiply

The first step of the chirp scaling algorithm is to transform the raw data into the range-Doppler domain. This data is then multiplied by a phase function which equalizes the range migration to that of the *reference range*, r_0. The time locus of the reference range in the range-Doppler domain is,

$$t_0(k_u) = \frac{2}{c} \cdot r_0 \cdot [1 + C_s(k_u)] \tag{A.30}$$

The required chirp scaling multiplier is then given by [128]

$$\phi\Phi_1(t, k_u) = \exp\left\{ j\pi K_s(k_u; r_0) \cdot C_s(k_u) \cdot [t - t_0(k_u)]^2 \right\} \tag{A.31}$$

where the *chirp scaling* parameter $C_s(k_u)$ is the curvature factor of (A.23). The SRC modified chirp rate has been calculated for the reference range r_0. Chapter 7 shows that this represents an approximation of minor consequence.

Multiplication of the range-Doppler data by $\phi\Phi_1(t, k_u)$ causes a small, Doppler wavenumber dependent, deformation of the phase structure within each range chirp. This deformation shifts the phase centers (but not the envelopes) of all the range chirps such that they all follow the same reference curvature trajectory [128]:

$$t(k_u) = \frac{2}{c} \cdot [x + r_0 \cdot C_s(k_u)] \tag{A.32}$$

A.3.2 The range Fourier transform

The chirp scaled range-Doppler signal for the point target is given by,

$$mM_\delta^1(t, k_u) = \phi\Phi_1(t, k_u) \cdot eE_\delta(t, k_u) \tag{A.33}$$

The principle of stationary phase is used to perform the temporal Fourier transform of this data. The appropriate time dependent phase function is,

$$\begin{aligned}
\phi(t) =& \pi K_s(k_u; r_0) \cdot \left(t - \frac{2}{c} x[1 + C_s(k_u)] \right)^2 + \pi K_s(k_u; r_0) C_s(k_u) \left(t - \frac{2}{c} r_0[1 + C_s(k_u)] \right)^2 - \omega_b t \\
=& + \left\{ \pi K_s(k_u; r_0)[1 + C_s(k_u)] \right\} \cdot t^2 \\
& - \left\{ \frac{4\pi}{c} K_s(k_u; r_0)[1 + C_s(k_u)][x + r_0 C_s(k_u)] + \omega_b \right\} \cdot t \\
& + \left\{ \frac{4\pi}{c^2} K_s(k_u; r_0)[1 + C_s(k_u)]^2 [x^2 + r_0^2 C_s(k_u)] \right\}
\end{aligned} \tag{A.34}$$

where only the t dependent phase terms of (A.29) have been used and the value of the effective FM rate has been fixed to the value of the reference range r_0, i.e. $K_s(k_u; x)$ has been replaced with $K_s(k_u; r_0)$.

The stationary point of (A.34) is given by,

$$t^* = \frac{2}{c}[x + r_0 C_s(k_u)] + \frac{\omega_b}{2\pi K_s(k_u; r_0)[1 + C_s(k_u)]} \tag{A.35}$$

substituting this back into (A.34) gives the phase of the transform,

$$
\begin{aligned}
\psi(\omega_b) \equiv\ & \phi(t^*) \\
=\ & -\frac{\left[\frac{4\pi}{c} K_s(k_u; r_0)[1 + C_s(k_u)][x + r_0 C_s(k_u)] + \omega_b\right]^2}{4\pi K_s(k_u; r_0)[1 + C_s(k_u)]} \\
& + \frac{4\pi}{c^2} K_s(k_u; r_0)[1 + C_s(k_u)]^2 [x^2 + r_0^2 C_s(k_u)] \\
=\ & +\frac{4\pi}{c^2} K_s(k_u; r_0) C_s(k_u)[1 + C_s(k_u)](x - r_0)^2 \\
& - \frac{\omega_b^2}{4\pi K_s(k_u; r_0)[1 + C_s(k_u)]} \\
& - \frac{2}{c}\omega_b[x + r_0 C_s(k_u)]
\end{aligned}
\tag{A.36}
$$

The overall two-dimensional frequency domain signal is then,

$$
\begin{aligned}
MM_\delta^1(\omega_b, k_u) \approx & \sqrt{\frac{\pi x}{k K_c}} \cdot A(k_u) \cdot \operatorname{rect}\left(\frac{\omega_b}{2\pi K_s(k_u; r_0)[1 + C_s(k_u)]\tau_c}\right) \\
& \cdot \exp\left\{-j\sqrt{4k_0^2 - k_u^2} \cdot x\right\} \\
& \cdot \exp\left\{+j\frac{4\pi}{c^2} K_s(k_u; r_0) C_s(k_u)[1 + C_s(k_u)](x - r_0)^2\right\} \\
& \cdot \exp\left\{-j\frac{\omega_b^2}{4\pi K_s(k_u; r_0)[1 + C_s(k_u)]}\right\} \\
& \cdot \exp\left\{-j2k_b[x + r_0 C_s(k_u)]\right\}
\end{aligned}
\tag{A.37}
$$

Each of the phase terms has a direct interpretation; the first two phase functions are the azimuth modulation terms, independent of ω_b. They include a parametric dependence on range x that is matched in the range-Doppler domain by the third phase multiplier $\psi\Psi_3(t, k_u)$ that is discussed shortly. The third phase term, quadratic in ω_b, is the effective range chirp modulation whose initial rate K_c has been modified by the range curvature, and modified further still by the Doppler dependent chirp rate of the chirp scaling multiply. The effective chirp rate in the frequency domain (s/Hz), may be re-written in

the form [128],

$$\frac{1}{K_s(k_u; r_0)[1 + C_s(k_u)]} = \frac{1}{K_c[1 + C_s(k_u)]} - \frac{K_{src}(k_u; r_0)}{1 + C_s(k_u)} \tag{A.38}$$

showing that the chirp scaled chirp rate (the first term) and the geometric phase distortion (the second term) are separable and additive. The modulation in temporal frequency that is matched by SRC is the second term. The fourth phase term of (A.37), linear in ω_b, carries the correct range position x of each scatterer as well as its range curvature. This range curvature is the same for all ranges due to the chirp scaling multiplier, $\phi\Phi_1(t, k_u)$.

A.3.3   Bulk range migration correction, pulse compression, and SRC multiply

Bulk range migration correction, pulse compression, secondary range compression and deconvolution of the frequency dependent amplitude term are performed with the following multiply,

$$\Theta\Theta_2(\omega_b, k_u) = \sqrt{\frac{k}{k_0}} \cdot \exp\left\{j\frac{\omega_b^2}{4\pi K_s(k_u; r_0)[1 + C_s(k_u)]}\right\} \cdot \exp\left\{j2k_b r_0 C_s(k_u)\right\} \tag{A.39}$$

The phase terms in (A.39) are the conjugates of the phase terms in (A.37) that are dependent only on ω_b and k_u. The first phase factor achieves range focus (pulse compression) including SRC. Range migration correction is performed by the second phase factor, and since it corrects the dominant range curvature effects (the chirp scaling multiplier corrected a minor amount of the curvature), it is known as *bulk range migration correction.*

The wavenumber signal after the multiplication of (A.37) and (A.39) is given by,

$$\begin{aligned} MM_\delta^{12}(\omega_b, k_u) \approx & \sqrt{\frac{\pi x}{k_0 K_c}} \cdot A(k_u) \cdot \text{rect}\left(\frac{\omega_b}{2\pi K_s(k_u; r_0)[1 + C_s(k_u)]\tau_c}\right) \\ & \cdot \exp\left\{-j\sqrt{4k_0^2 - k_u^2} \cdot x\right\} \\ & \cdot \exp\left\{j\frac{4\pi}{c^2} K_s(k_u; r_0) C_s(k_u)[1 + C_s(k_u)](x - r_0)^2\right\} \\ & \cdot \exp(-j2k_b x) \end{aligned} \tag{A.40}$$

Though it has not been indicated explicitly in the envelope of the rect($\cdot$) function, the chirp scaling operations have distorted the wavenumber spectrum to a similar curved shape as produced by the Stolt mapping operator of the wavenumber algorithm. This distorted spectrum is windowed in a similar way to the wavenumber algorithm; the effect of the radiation pattern is deconvolved and a weighting function

is applied across the processing bandwidths (in this case, just a rect function). The resulting signal is,

$$
\begin{aligned}
NN_\delta^{12}(\omega_b, k_u) = {} & \cdot \sqrt{\frac{\pi x}{k_0 K_c}} \cdot \operatorname{rect}\left(\frac{k_u}{B_p}\right) \cdot \operatorname{rect}\left(\frac{\omega_b}{2\pi B_{\mathrm{eff}}}\right) \\
& \cdot \exp\left\{-j\sqrt{4k_0^2 - k_u^2} \cdot x\right\} \\
& \cdot \exp\left\{j\frac{4\pi}{c^2} K_s(k_u; r_0) C_s(k_u)[1 + C_s(k_u)](x - r_0)^2\right\} \\
& \cdot \exp(-j2k_b x)
\end{aligned}
\tag{A.41}
$$

A.3.4 The range inverse Fourier transform

With near perfect phase compensation of all range modulation, the inverse temporal Fourier transform of the windowed version (A.41) collapses the range dimension of the point target function to a compressed sinc envelope at the correct range position $2x/c$, leaving only azimuth phase terms,

$$
\begin{aligned}
nN_\delta^{12}(t, k_u) = {} & \sqrt{\frac{\pi x}{k_0}} \cdot \frac{B_{\mathrm{eff}}}{\sqrt{K_c}} \cdot \operatorname{rect}\left(\frac{k_u}{B_p}\right) \cdot \operatorname{sinc}\left[B_{\mathrm{eff}} \cdot \left(t - \frac{2x}{c}\right)\right] \\
& \cdot \exp\left\{-j\sqrt{4k_0^2 - k_u^2} \cdot x\right\} \\
& \cdot \exp\left\{+j\frac{4\pi}{c^2} K_s(k_u; r_0) C_s(k_u)[1 + C_s(k_u)](x - r_0)^2\right\} \\
= {} & \sqrt{\frac{\pi x}{k_0}} \cdot \frac{B_{\mathrm{eff}}}{\sqrt{K_c}} \cdot \operatorname{rect}\left(\frac{k_u}{B_p}\right) \cdot \operatorname{sinc}\left[B_{\mathrm{eff}} \cdot \left(t - \frac{2x}{c}\right)\right] \\
& \cdot \exp\left\{-j\left(\sqrt{4k_0^2 - k_u^2} - 2k_0\right) \cdot x\right\} \\
& \cdot \exp\left\{+j\frac{4\pi}{c^2} K_s(k_u; r_0) C_s(k_u)[1 + C_s(k_u)](x - r_0)^2\right\} \\
& \cdot \exp\left\{-j2k_0 x\right\}.
\end{aligned}
\tag{A.42}
$$

Now that the rows and columns have been decoupled by range migration correction, x and $ct/2$ can be used interchangeably. The last line of (A.42) splits up the along-track modulation term from the carrier (range modulation) term.

A.3.5   Azimuth compression and phase residual removal

The phase multiplier needed to complete the algorithm is given by,

$$\psi\Psi_3(t, k_u) = \exp\left\{ j\left(\sqrt{4k_0^2 - k_u^2} - 2k_0 \right) \cdot x \right\} \cdot \exp\left\{ -j\frac{4\pi}{c^2} K_s(k_u; r_0) C_s(k_u) \left[1 + C_s(k_u)\right] (x - r_0)^2 \right\}$$

(A.43)

The multiplication of (A.42) and (A.43) gives,

$$nN_\delta^{123}(t, k_u) = \sqrt{\frac{\pi x}{k_0}} \cdot \text{rect}\left(\frac{k_u}{B_p} \right) \cdot \frac{B_{\text{eff}}}{\sqrt{K_c}} \cdot \text{sinc}\left[B_{\text{eff}} \cdot \left(t - \frac{2x}{c} \right) \right] \cdot \exp(-j2k_0 x)$$

(A.44)

Inverse spatial transformation gives the final, bandlimited, point target estimate,

$$nn_\delta^{123}(t, u) = \sqrt{\frac{4\pi x}{k_0 D^2}} \cdot \frac{B_{\text{eff}}}{\sqrt{K_c}} \cdot \text{sinc}\left(\frac{2}{D} \cdot u \right) \cdot \text{sinc}\left[B_{\text{eff}} \cdot \left(t - \frac{2x}{c} \right) \right] \cdot \exp(-j2k_0 x)$$

(A.45)

To show that the phase term is as expected, consider the following 1-D argument; the demodulated return of a modulated pulse, $p_m(t) = \text{sinc}(B_{\text{eff}} t) \cdot \exp(j\omega_0 t)$, from range x is,

$$\begin{aligned}
e_b(t) &= p_m\left(t - \frac{2x}{c} \right) \cdot \exp(-j\omega_0 t) \\
&= \text{sinc}\left[B_{\text{eff}} \cdot \left(t - \frac{2x}{c} \right) \right] \cdot \exp(-2k_0 x)
\end{aligned}$$

(A.46)

Appendix B

MULTIRESONANCE DESIGN OF A TONPILZ TRANSDUCER USING THE FINITE ELEMENT METHOD

This appendix is a reproduction of the publication "Multiresonance design of a Tonpilz transducer using the Finite Element Method", by David W. Hawkins and Peter T. Gough, it appeared in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 43, No. 5, September 1996.

Multi-resonance design of a Tonpilz transducer using the Finite Element Method

D. W. Hawkins and P. T. Gough

Abstract— **The design and characterization of a wide band-width Tonpilz transducer is carried out using the Finite Element Method. This wide bandwidth has been achieved by introducing a symmetric flexural resonance (sometimes called a "flapping" resonance) in the head-piece of the Tonpilz transducer. This flexural resonance is exploited by lip-mounting of the transducer as opposed to the more traditional nodal mount. Each transducer is characterized by high-power handling, high-electroacoustic efficiency, broad-bandwidth (low-Q), and high-electromechanical coupling. These are characteristics which are usually associated with designs employing more complicated electrical or mechanical matching techniques. An array of these transducers was constructed and it displays low-ripple (<3dB) transmission of one octave 20kHz-40kHz signals. A comparison with a recent broad-band design by Inoue *et al* [1] which uses a matching plate is also made to illustrate the inherent simplicity of the flexural resonance approach.**

Keywords— **sonar, broadband transducer design, Tonpilz.**

I. Introduction

Active underwater acoustic transducers often transmit large time-bandwidth (TB) or wide bandwidth waveforms. These waveforms are necessary in applications where high-resolution or high bit-rates are required, such as Data Transmission [1, 2], Synthetic Aperture Sonar [3], Marine Sediment Analysis [4] and Underwater Communication [5]. The ability to encode such waveforms allows discrimination of the transmitted signal from uncorrelated background sources. These transducers are usually characterized by high-power handling, high-electroacoustic efficiency, low-Q and high-electromechanical coupling—design criteria which are generally difficult to obtain in practice.

Historically transducer design for sonar projectors has been largely based on one-dimensional approximations [6–10]. These one-dimensional models are limited to characterizing narrow-band transducers where the input parameters for the model are taken from experimental data [11], with the characteristics being valid only around the fundamental longitudinal resonance. When designing a broad-band transducer, the shape of the transducer becomes more complex and other vibrational modes begin to appear and interact with the fundamental mode [11, 12]. These additional modes are generally difficult to model in one-dimension [11], however, a simple one-dimensional model has been developed for a Tonpilz transducer ac-

The authors are with the Department of Electrical and Electronic Engineering, Canterbury University, Christchurch, New Zealand (e-mail: gough@elec.canterbury.ac.nz).

counting for these modes [13].

The Finite Element Method (FEM) [14] uses a three-dimensional representation of the transducers governing equations and so has none of the limitations of one-dimensional modelling. It can be used for any transducer configuration and material property [15, 16]. The FEM allows full characterization of a transducers vibrational modes (i.e. resonance and anti-resonance frequencies, coupling coefficients and admittance). When used in conjunction with the Boundary Element Method [17], it can also characterize water loaded conditions (i.e. beam-width, efficiency and source-level). The flexibility of the FEM has led to its near-exclusive use in the design of arbitrarily shaped transducers such as the Tonpilz.

Impedance matching is often used to increase the bandwidth of transducers. Impedance matching can be done in two ways, electrically or mechanically. Matching electrical networks can be placed on the input to a transducer [18] or matching mechanical impedances (plates) can be placed on the radiating face of the transducer [1, 19]. Further compliant sections can also be incorporated with mechanical matching to achieve broad-banding [2]. A rarely used alternative to impedance matching is the use of multi-resonance in a device. Previous investigations of the Tonpilz design [12, 20] have indicated the presence of a flexural "flapping" resonance mode well above that of the fundamental longitudinal resonance. Butler *et al* [13] discuss the advantages of using a "flapping" mode to increase bandwidth. Coates [2] also investigated this method of broad-banding; both authors found the distribution of velocities across the face of the transducer had a detrimental effect on the transducer pass-band and polar response.

This paper shows how this "flapping" mode can be exploited by lip-mounting of the transducer, as opposed to the more traditional method of nodal mounting. Lip-mounting ensures that piston-like motion will occur at both resonance frequencies, avoiding adverse effects in the transmission characteristics. Many existing Tonpilz designs use cone-shaped head-pieces with slow tapers to avoid this "flapping" mode. However, we have modified the design by ending the taper with first a circular-, then a square-plate and by changing the taper and the plate dimensions, we have introduced a symmetric flexural resonance about one octave above the fundamental resonance.

Initially we investigate the flexural mode produced in the head-piece of the Tonpilz transducer using two finite-element models, one axisymmetric and one non-axisymmetric. Then we examine the interaction of this flexural mode with the fundamental under water-loaded conditions. An array of these transducers was constructed

which displays low-ripple transmission and a normal polar response. The results obtained from this analysis are then compared to a recent design by Inoue *et al* [1] which uses a matching section on the face of a Tonpilz transducer to achieve a wide bandwidth. This comparison shows the advantages of a multi-resonant design.

II. SUMMARY OF THE FINITE ELEMENT METHOD

As the use of the Finite Element Method (FEM) is well covered in [14, 21, 22] and the its application to piezoelectric transducers is described in [11, 12, 15, 16, 20, 23, 24] we proceed by stating the classical set of equations in matrix form for an undamped structure containing piezoelectric and non-piezoelectric materials with no external mechanical forces ($\boldsymbol{F} = 0$)

$$
\begin{pmatrix} (\boldsymbol{K}_{uu} - \omega^2 \boldsymbol{M}) & \boldsymbol{K}_{u\phi} \\ \boldsymbol{K}_{u\phi}^t & -\boldsymbol{K}_{\phi\phi} \end{pmatrix} \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{\Phi} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{Q} \end{pmatrix} \quad (1)
$$

where $\boldsymbol{U}$, $\boldsymbol{\Phi}$, and $\boldsymbol{Q}$ are the vectors of nodal displacements, electrical potentials, and electrical charges for the complete structure. The matrices $\boldsymbol{K}_{uu}$, $\boldsymbol{K}_{u\phi}$, $\boldsymbol{K}_{\phi\phi}$, and $\boldsymbol{M}$ are respectively the conventional stiffness, piezoelectric stiffness, dielectric stiffness and consistent mass matrices (superscript t indicates transposed). The radian frequency ω represents the harmonic driving frequency. The negative in front of the dielectric stiffness matrix is there to avoid a negative sign in the equation for the dielectric energy, many references absorb this sign into the definition of $\boldsymbol{K}_{\phi\phi}$.

Any practical transducer has a number of electrodes, these electrodes represent equipotential surfaces on each of which the following holds true:

$$
\begin{array}{cccccccc} \phi_0 & = & \phi_1 & = & \phi_2 & = & \dots & = & \phi_n \\ Q_0 & = & Q_1 & + & Q_2 & + & \dots & + & Q_n \end{array} \quad (2)
$$

where n is the number of nodes on the equipotential surface. The potential degrees of freedom of the nodes on the 'hot' electrodes charged to potential ϕ_0 can be *reduced* to only one degree of freedom corresponding with the charge Q_0. The potential degrees of freedom associated with the nodes on the grounded electrodes can be *eliminated* from the system equations entirely as they are equal to zero [11, 20]. The remaining internal potential degrees of freedom can be eliminated from the system equations by *condensation*, using a method suggested by Guyan [25]. The result of this condensation is that (1) becomes

$$
\begin{pmatrix} (\overline{\boldsymbol{K}}_{uu} - \omega^2 \boldsymbol{M}) & \overline{\boldsymbol{K}}_{u\phi} \\ \overline{\boldsymbol{K}}_{u\phi}^t & -\overline{K}_{\phi\phi} \end{pmatrix} \begin{pmatrix} \boldsymbol{U} \\ \phi_0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ \frac{I}{j\omega} \end{pmatrix} \quad (3)
$$

where the potential and the dielectric stiffness matrix are now scalar quantities, charge has been replaced with $I/(j\omega)$ where I is the input current, and where the quantities with bars include the effects of the condensed potentials [20]. This condensation process transforms the *total formulation* in (1) into what can be called the *surface formulation* [26], due to its dependence now on the surface potential only. The surface formulation will become relevant

when dealing with the predicted and experimentally measurable coupling factors. Note that the negative in front of the condensed dielectric stiffness matrix in (3) is now consistent with Boucher *et al* [11].

Solving (3) gives us a direct expression for the terminal admittance Y in terms of the various electroelastic stiffness properties [20] (note that in [20] the negative sign is absorbed into the definition of $\overline{K}_{\phi\phi}$).

$$
Y = \frac{I}{\phi_0} = j\omega(-\overline{K}_{\phi\phi} - \overline{\boldsymbol{K}}_{u\phi}^t (\overline{\boldsymbol{K}}_{uu} - \omega^2 \boldsymbol{M})^{-1} \overline{\boldsymbol{K}}_{u\phi}) \quad (4)
$$

At resonance $Y \to \infty$ so $\phi_0 \to 0$ for $I \neq 0$ and regardless of $\overline{\boldsymbol{K}}_{u\phi}$ and $\overline{K}_{\phi\phi}$

$$
\left(\overline{\boldsymbol{K}}_{uu} - \omega^2 \boldsymbol{M} \right) \boldsymbol{U} = 0 \quad (5)
$$

thus $\det\left(\overline{\boldsymbol{K}}_{uu} - \omega^2 \boldsymbol{M}\right) = 0$ will give the eigenvalues ω_i and corresponding eigenvectors $\boldsymbol{U}_i$ of the ith resonance mode.

At anti-resonance $Y \to 0$ and so $I \to 0$ for $\phi_0 \neq 0$ leaving

$$
\left(\overline{\boldsymbol{K}}_{uu} + \frac{\overline{\boldsymbol{K}}_{u\phi} \cdot \overline{\boldsymbol{K}}_{u\phi}^t}{\overline{K}_{\phi\phi}} - \omega^2 \boldsymbol{M} \right) \boldsymbol{U} = 0 \quad (6)
$$

where this time the determinant will give the eigenvalues ω_i and corresponding eigenvector $\boldsymbol{U}_i$ of the ith anti-resonance mode.

The ANSYS FE package [16] used to model the transducer solves (5) and (6) by not only condensing out the internal potential degrees of freedom, but also by condensing out a number of displacement degrees of freedom. The condensed displacement degrees of freedom are referred to as *slaves* while the remaining displacement degrees of freedom are known as the *masters*. This particular condensation process preserves the stiffness matrix but approximates the mass. By selecting the master degrees of freedom based on a minimum stiffness to mass ratio, the effect of the use of the approximate mass has little effect on the lower eigenvalues and eigenvectors [16].

Characterization of a transducer also requires a knowledge of the coupling of the modes, this coupling is given by the electromechanical coupling coefficient k [11, 15]

$$
k \equiv \left(\frac{E_m^2}{E_e E_d} \right)^{\frac{1}{2}} \quad (7)
$$

where E_m is the electromechanical coupling energy, E_e is the elastic energy, and E_d is dielectric energy.

The stored (potential) energy U in the piezoelectric structure can be written [15, 26]

$$
U = \frac{1}{2} \int (\boldsymbol{S}^t \boldsymbol{T} - \boldsymbol{D}^t \boldsymbol{E}) dV \quad (8)
$$

where $\boldsymbol{S}$, $\boldsymbol{T}$, $\boldsymbol{D}$, and $\boldsymbol{E}$ are respectively the vectors of mechanical strain, mechanical stress, electric displacement, and electric field. If U is distributed so that

$$
U \equiv E_e + 2E_m - E_d \quad (9)
$$

then in terms of the assembled finite element matrices

$$U = \frac{1}{2}\boldsymbol{U}^t \boldsymbol{K}_{uu}\boldsymbol{U} + \frac{1}{2}(\boldsymbol{U}^t \boldsymbol{K}_{u\phi}\boldsymbol{\Phi} + \boldsymbol{\Phi}^t \boldsymbol{K}_{u\phi}^t \boldsymbol{U}) \\ - \frac{1}{2}\boldsymbol{\Phi}^t \boldsymbol{K}_{\phi\phi}\boldsymbol{\Phi} \tag{10}$$

If the total formulation matrices in (10) are replaced with the partitioned matrices used in the condensation of (1) to (3), then the *surface formulation* of the internal energy is

$$U = \frac{1}{2}\boldsymbol{U}^t \overline{\boldsymbol{K}}_{uu}\boldsymbol{U} + \frac{1}{2}(\boldsymbol{U}^t \overline{\boldsymbol{K}}_{u\phi}\phi_0 + \phi_0 \overline{\boldsymbol{K}}_{u\phi}^t \boldsymbol{U}) \\ - \frac{1}{2}\phi_0 \overline{\boldsymbol{K}}_{\phi\phi}\phi_0 \tag{11}$$
$$= E_{e,stiff} + 2E_{m,surf} - E_{d,surf}$$

where the condensation of the piezoelectric potential degrees of freedom results in a stiffened value of the elastic energy and surface formulations for the coupled and dielectric energy.

This leads us to two possible forms for the electromechanical coupling factor, (7) based on the total formulation and a coupling factor based on the surface formulation

$$k_{surf} \equiv \left(\frac{E_{m,surf}^2}{E_{e,stiff}E_{d,surf}} \right)^{\frac{1}{2}} \tag{12}$$

As it is not possible to measure the elastic, dielectric and electromechanical energy in a transducer, an approximation which uses the distance between the resonance (5) and anti-resonance (6) frequencies is used to calculate an effective electromechanical coupling coefficient [6,15]

$$k \approx k_{eff} = \left(\frac{f_a^2 - f_r^2}{f_r^2} \right)^{\frac{1}{2}} \tag{13}$$

where f_r and f_a are the resonance (maxima) and anti-resonance (minima) frequencies of the admittance.

Using the resonance and anti-resonance conditions in (5) and (6) and assuming that the displacements $\boldsymbol{U}$ are the same for each mode, it can be shown that

$$k_{surf} = \left(\frac{f_a^2 - f_r^2}{f_r^2} \right)^{\frac{1}{2}} \tag{14}$$

which shows that the experimentally measurable effective coupling coefficient can be predicted using the surface formulation of the system equations.

The magnitude of the electromechanical coupling of a vibrational mode represents the significance of that particular mode compared to other modes [15]. If the coupling of a certain mode is of the order of 50% or higher, that mode will be strongly excited [15]. The coupling of any particular mode can never exceed the coupling coefficient of the working mode of the ceramic (k_{33} in our case). This is logical since the coupling coefficient k is calculated in the whole structure and includes non-piezoelectric materials [11].

TABLE I
CIRCULAR-FACED TRANSDUCER RESONANT
FREQUENCIES AND EFFECTIVE COUPLING

mode	Predicted			Measured		
	f_r kHz	f_a kHz	k_{eff}	f_r kHz	f_a kHz	k_{eff}
1	27.2	32.7	0.67	26.1	30.7	0.60
2	62.8	64.0	0.20	67.5	68.4	0.16
3	85.2	85.6	0.10	85.0	85.5	0.11

III. An Application

A. The Tonpilz Design

One of the most common composite transducer designs is the Tonpilz; this word originating from the German words "sound" and "mushroom" where the mushroom refers to the Tonpilz's distinctive shape. The design consists of an acoustically active ceramic stack, sandwiched between a head and tail piece. Since the specific impedance of the ceramic is higher than that of water, a better impedance match would be achieved if the area of the radiating face is significantly larger than that of the piezoelectric materials. This leads to a design with a tapered ("mushroom") head-piece. The material used in the head-piece is chosen to have an acoustic impedance between that of the ceramic and water. The tail section has the highest impedance. This is so that the overall transducer will act like a piston with most of the displacement occurring at the head-piece. The tail-piece usually radiates into air, so the large impedance mismatch helps to reflect energy back to the head section, improving its radiating performance. Ceramics are weak under expansion and when high voltages are used, it is possible to shatter a ceramic. So to complete the design, the whole structure is placed under compressive prestress by placing a bolt through the middle—the prestress afforded by this bolt allows the ceramic to withstand these high voltages as well as ensuring intimate mechanical contact between the components.

References [2,12,13,20,27] show clearly the operation of several classical Tonpilz designs.

B. Circular-faced Tonpilz Design

The initial transducer design by Knight [28] was axisymmetric, with the tapered cone of the head-piece ending in a circular plate. This circular plate gave the transducer a circular radiating face. Using axisymmetric finite elements a two-dimensional finite element (FE) model was constructed. The head (aluminium) and tail (mild steel) sections were modeled using axisymmetric two-dimensional elements (PLANE42, PLANE82 [16]). While the ceramic (PZT-4) was modeled using a coupled field 2-D solid (PLANE13 [16]) with potential as the extra degree of freedom.

A modal analysis was performed (in air) on the transducer, this determined the resonance and anti-resonance frequencies of the device. The FEM correctly predicted
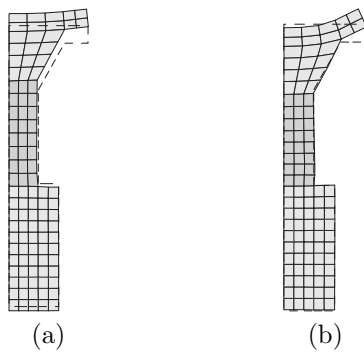
Fig. 1. Free-body resonance mode shapes for the circular-faced trans-
ducer at (a) 27.2kHz and (b) 62.8kHz (dashed line is the undis-
placed shape). Axisymmetric FE model.

Fig. 2. Circular-faced transducer admittance.

the fundamental resonance between 20kHz and 30kHz as
Knight [28] predicted and observed, as well as higher or-
der resonances (see Table I). The FEM has several other
advantages; it will also predict the mode shape and the
electromechanical coupling of the mode. The results of
this analysis are shown in Fig. 1, with Table I showing the
modes and effective coupling factors.

The first mode (Fig. 1(a)) is obviously a longitudinal
resonance with displacement occurring predominantly in
the head-piece. The second mode also has most of the
displacement occurring in the head-piece, this motion being
largely flexural in nature. These diagrams are free-body, in
that no damping or boundary conditions due to mounting
have been included.

Table I shows a comparison of the predicted finite ele-
ment resonance and anti-resonance frequencies and effec-
tive electromechanical coupling coefficients to those ob-
served for the circular-faced transducer. The correspon-
dence between measurement and prediction is not per-
fect but it is a substantial improvement over the one-
dimensional equivalent circuit results where no resonant
or anti-resonant frequencies were predicted near those ob-
served. Exact correspondence between the predicted and
measured resonances is difficult to achieve as the dis-
cretized FE model is only an approximation to the contin-
uous electro-elastic equations and the exact material prop-
erties of the materials used in the transducers construction
are seldom available. To investigate the effect of discretiza-
tion several other models were generated; a finer meshed
axisymmetric model and two quarter-symmetry solid mod-
els. The predicted resonance frequencies from these models
were negligibly different for the first resonance, and mesh
refinement caused the second resonance to move downward,
away from the actual measured second resonance. A simi-
lar effect was observed when we added structural damping
to the models. The material properties used in the FE
models were taken from references [6, 11, 29], but previous
experience has shown that these properties can vary quite
substantially [30]. The difference between the measured
resonances and those predicted by the FE models can be
attributed to material property variations.

Next a harmonic analysis was performed. This consisted
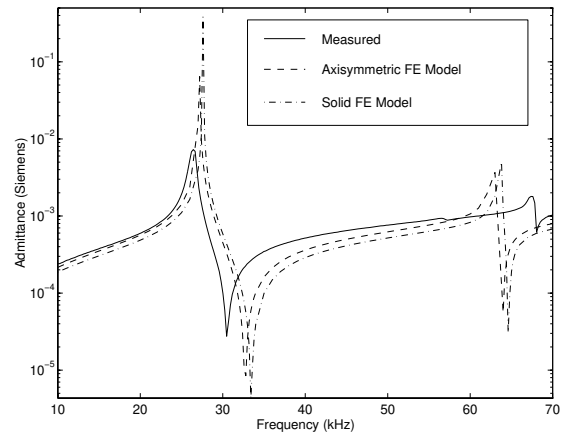of a time varying potential of 1 volt applied to the 'hot'

terminals of the transducer, over the frequency range 10-
70 kHz. Information obtained from this type of analysis
can be used to determine the admittance of the transducer
as shown in (4). Fig. 2 shows the FE admittance (in air),
along with measured results. The result for a one-quarter
solid FE model of the circular-faced transducer is also in-
cluded to show the slight disagreement between the two
models. The difference between the admittance predictions
of the two FE models is due to the lower mesh density in
the solid model. The discrepancy between the admittance
predictions of the two FE models and the actual measured
admittance can again be attributed to material property
variability and not mesh density. An analysis of the mesh
parameters, memory demands, and processing times for the
FE models is given in Appendix I.

C. Square-faced Tonpilz Design

Once the power of the FEM had been determined, it
was of interest to see what would occur as the design was
changed so that the transducer head-piece ended with a
square plate. Fig. 3 shows the shift of resonance and anti-
resonance frequencies with changing head shape. The head
geometry can be described by the formula $x^n + y^n = r^n$,
where x and y are directions, and r is the radius of the
circular-faced transducer, or half-width of the square-faced
transducer. With $n = 2$, the head-piece has a circular
plate, as $n \rightarrow \infty$ the plate becomes square. As stated
before, the distance between resonance and anti-resonance
frequencies gives a measure of the electromechanical cou-
pling. In Fig. 3 you can see that the coupling of the first
mode decreases as the head shape becomes square, while
the second resonance moves down into our band of inter-
est, with increasing coupling. As long as this coupling is
around 50%, it is still considered useful.

The square-plate allows lip-mounting instead of the tra-
ditional nodal mount. Because we were intending to exploit
the "flapping" mode, it was unlikely that we would find a
nodal mount appropriate for this mode as well as the fun-
damental mode. The square-face also has the advantage
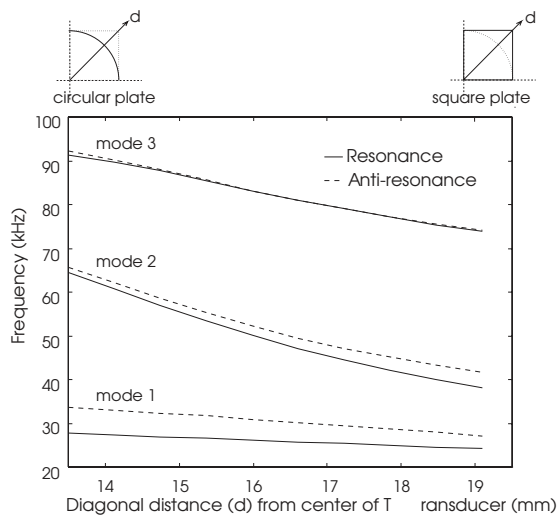that when it is used in a close packed array it maximizes

Fig. 3.   Progression of modes with changing head shape.

TABLE II
SQUARE-FACED TRANSDUCER RESONANT
FREQUENCIES AND EFFECTIVE COUPLING

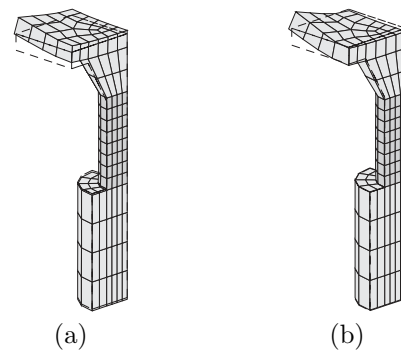| mode	Predicted			Measured		
	f_r kHz	f_a kHz	k_{eff}	f_r kHz	f_a kHz	k_{eff}
1	24.1	26.9	0.50	24.0	26.8	0.50
2	37.7	41.4	0.45	38.3	40.1	0.31
3	42.7	42.7	0	N/A	N/A	N/A*
						*see text

Fig. 4.   Free-body resonance mode shapes for the square-faced transducer at (a) 24.1kHz and (b) 37.7kHz (dashed line is the undisplaced shape). Solid FE model.
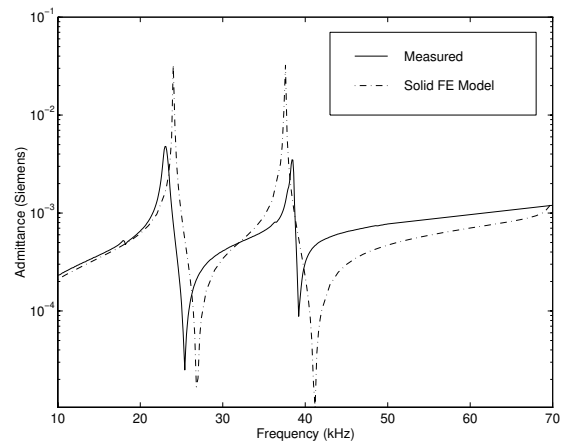

Fig. 5.   Square-faced transducer admittance.

the radiating area coupled to the acoustic medium.

The square-faced transducer has quarter-symmetry so only needs to have one-quarter of the transducer modelled, along with appropriate symmetry conditions. This still necessitates the use of solid elements (SOLID5, SOLID45 [16]), but reduces the computation time significantly. The first two mode shapes of this model are shown in Fig. 4. The frequency of both these modes are within our band of interest, but are uncoupled from each other in air. The admittance curve shown in Fig. 5 matches well with that experimentally recorded. Table II shows the resonant frequencies and effective coupling of the square-faced model. Appendix I discusses the FE model processing demands.

The effect of the piezoelectric boundary conditions can be seen in Table II, the voltage-dependent strain conditions split the mechanical resonance that would have occurred had the material not been piezoelectric, into a lower electrical resonance and an upper electrical anti-resonance. Since mode 3 resonance and anti-resonance frequencies are the same, it is a purely mechanical resonance and will not be observed in the electrical admittance shown in Fig. 5.

The free-body displacement of the flexural mode shown in Fig. 4(b) shows that there is a phase difference across the face of the transducer. This phase difference can have a number of detrimental effects; it can cause a null in the polar response [2], dips in the transmission response [13], or at least increase the sidelobe level [31]. We avoid these effects by lip-mounting from the corners of the plate on the head-piece. This forces the vibration to be in phase across the whole face, with less movement occurring at the corners.

To show that lip-mounting avoids the detrimental effects, the transducer was mounted on compliant supports. These supports were constructed from short lengths of O-ring recessed into the mounting material (see Fig. 6), and source-level (the dB ratio of acoustic pressure produced by the transducer at 1m relative to 1μPa for an applied voltage of 1V$_{RMS}$) measurements were made. Generally source-level measurements are taken in the far-field and referenced back to 1m; for a single element of aperture $D = 27$mm the far-field is at range $R \geq 2D^2/\lambda$ [32], so for $f = 40$kHz the far-field lies at range $R \approx 40$cm. As we took measurements at $R = 1$m, far-field can be assumed. The single transducer generated a source level of 130 dB/V re 1μPa@1m, over the band 19-50kHz (see Fig. 7). The mounts were then altered to non-compliant brass pins to see what effect this had; this change had very little effect (measurements were made at 1m depth).

Other mechanical aspects of the device were modelled to see how they affected the operation of the device. As a steel stress rod exists through the center of the device, a prestressed analysis was performed on the transducer, followed by a modal analysis. The resulting mode shapes

TABLE III
TONPILZ TRANSDUCER COMPARISON

| Characteristic	Transducer Design	
	Inoue *et al* [1]	Hawkins and Gough
Broad-banding Method	single matching plate	multi-resonance
Support	nodal-mount	lip-mount
Dimensions:		
• face	25mm × 25mm	27mm × 27mm
• length	81mm (23mm matching plate)	48mm
Efficiency, η		
• without mounting losses	85%	97% @ 24kHz
		87% @ 38kHz
• with mounting losses	not given	75% @ 24kHz
		47% @ 38kHz
Available 3dB BW	20 - 35kHz	20 - >50kHz
Source-level	130 dB/V re 1μPa@1m	130 dB/V re 1μPa@1m
Frequencies used	20 - 30kHz	20 - 40kHz
Central frequency	25kHz	30kHz
Transducer Q	2.5	1.5
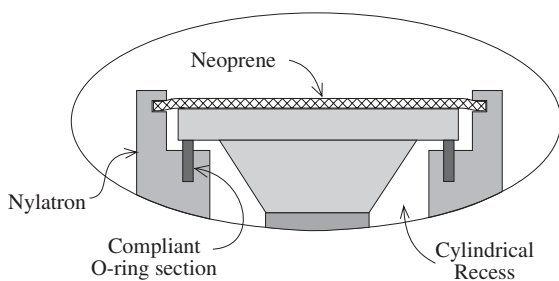
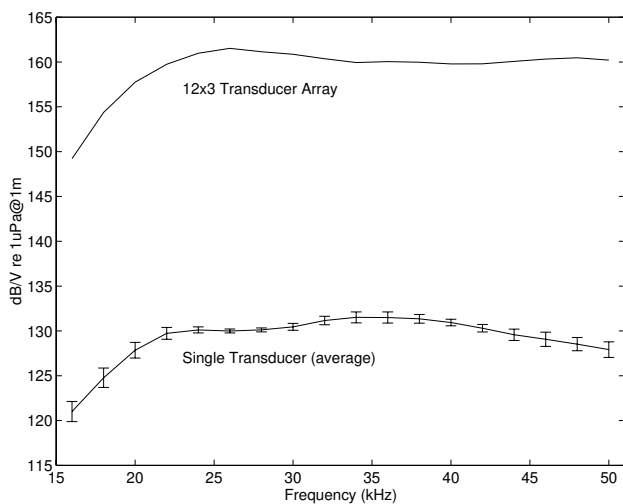Fig. 6.   Single transducer testing configuration.

Fig. 7.   Source-Level measurement of a single transducer and of the array. Compare this to Fig. 10 of Inoue *et al* [1]. Error bars on the single transducer measurement indicate manufacturing spread not measurement spread.

differed slightly, with the resonant frequencies being predicted slightly higher. Overall, the prestressed bolt had very little impact on the modes. The effect of the thin copper electrodes is negligible, so they were also ignored. The only aspect of the design that alters the displacements and resonances of the device is the loading introduced by external water pressure acting on the transducer face. To model the effects of the water load it would be necessary to perform a Boundary Element Analysis [17], or model the water as a volume of acoustic elements connected to infinite acoustic elements in the far-field of the transducer [17, 33]. The water load will cause damping due to sound emission as well as down-shifting of the resonances due to mass loading. These effects cause the fundamental and "flapping" mode to couple together and results in a transducer with a low-ripple, wide-bandwidth characteristic.

Another useful figure used for characterizing an electroacoustic transducer is the electroacoustic efficiency, η, of the device. This is defined as the ratio of the acoustic power radiated into the water to the total input electrical power [34]. This is usually calculated from admittance loci measurements. Unfortunately the technique is not appropriate for wide bandwidth or multi-resonant transducers as the loci are no longer circular [34]. An alternative method which we have used fits an equivalent circuit to the measured admittance. The circuit consists of a capacitance representing the bulk capacitance of the piezoelectric in parallel with two motional RLC arms representing the two resonance modes. Measurements are first taken in air in the housing to estimate internal electrical losses and radiative losses due to the housing, then measurements are made in the water. The increase in resistance in each motional arm can be attributed to radiative losses in the water. In this way the electroacoustic efficiency for the ith resonance can be defined as,

$$\eta_i = \frac{R_w - R_a}{R_w} \qquad (15)$$

where R_a is the resistance in air due to radiative and electrical losses, while R_w is the resistance representing the combined losses in water. Thus the difference $R_w - R_a$ is the resistance due to radiation into the water. These values can also be obtained approximately from the inverse of (maximum) conductance at each resonance since the two resonances are about an octave apart.

For the longitudinal mode this gave $\eta_1 \approx 77\%$ and for the "flapping" mode $\eta_2 \approx 48\%$. Indicating a reasonable efficiency at each resonance mode. Even though the second resonance is of lower efficiency, the directivity of the transducer is higher because it occurs at a higher frequency, and these factors combine to produce a transducer with a low-ripple source-level (see Fig. 7) and high-efficiency.

IV. CONSTRUCTION OF AN ARRAY

The projector is made up of 36 Tonpilz transducers in a 12×3 array. To mount each transducer cylindrical holes just large enough to accommodate the tail-piece and ceramic section are drilled into the backing material. The transducer is lip-mounted by sitting the corners of the square-faced head-piece on four sections of O-ring drilled into the backing material. The separation of each transducer is approximately $\lambda/2$, thus avoiding grating lobes in the beam pattern. To hold the transducers in place and to water-proof the array, the faces of the transducers are glued to a neoprene cover which is attached to the backing material around the edge of the array.

The source-level (SL) for this array can be determined approximately from the SL from a single transducer. For every doubling of the number of transducers, there is a doubling of the transmitted power and a halving of the beamwidth, resulting in approximately a 6dB rise in the SL. For the 36 element array this means a $\log_2(36){\times}6\text{dB} \approx 31\text{dB}$ increase in SL over that of a single element, i.e. we would expect to measure a SL of about 161 dB/V re $1\mu\text{Pa@1m}$. As can be seen from Fig. 7, this estimate is verified by direct measurement. As with the SL measurement for a single transducer, projector SL measurements should be made in the far-field. For our projector the far-field is at range $R \geq 2D^2/\lambda$ [32], where D is the larger dimension of the array. So for $f = 40\text{kHz}$ the far-field lies at range $R \approx 6\text{m}$. As we did not have access to a large enough body of water (the SL measurements were taken in a tank of 3m diameter), the SL measurement were taken in the near-field. However analysis by Walter (page 56 [32]) indicates that the amplitude difference between the near- and far-field occurs mainly around the beam pattern nulls.

The beam pattern of a rectangular array along either of its main axes can be represented approximately as that of a linear array [35]. Fig. 8 shows the horizontal beam pattern at 30kHz. From Fig. 8 it can be seen that the nulls in the sidelobe pattern are not apparent and the level of the first sidelobe is higher than expected. These observations
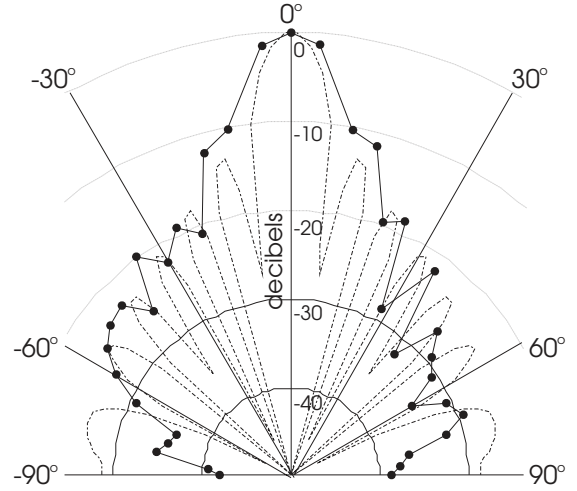

Fig. 8. Typical measured (near-field) horizontal beam pattern (30kHz) for the 12×3 array. Dashed line is the theoretical (far-field) beam pattern for a linear array of the same length.

are the result of measuring the SL in the near-field (page 56 [32]). Similar results were obtained for SL measurements at 20kHz and 40kHz.

V. A COMPARISON OF TWO TONPILZ TRANSDUCERS

Recently a wide bandwidth Tonpilz-style transducer was described by Inoue *et al* [1] for use in underwater data transmission. Their transducer possesses many similarities to ours. However, Inoue *et al* [1] produced a wide-bandwidth transducer by using a single matching plate on the face of their transducer. The data for the comparison made in Table III is taken from the text of Inoue's paper with some parameters being estimated from Fig. 10 of their paper.

Inoue *et al* [1] designed their transducer using one-dimensional techniques. This was possible as the head-piece in their design consisted of a 23mm matching plate attached to an aluminium head mass, no flexural motion will occur anywhere near the fundamental frequency. To use a one-dimensional model, Inoue had to experimentally estimate *effective* material properties for the bare transducer. Once these effective properties were estimated, the matching sections were modelled. This is a common way to use the equivalent circuit method; unfortunately, it still means a prototype transducer must be built and parameters estimated before any design optimization can proceed.

Table III shows a summary of the two transducer designs. Some comments on the numbers in the table are needed before a sensible comparison can be made. Inoue's measurement of electroacoustic efficiency, η, needs clarifying. Their efficiency is calculated from admittance loci comparisons of the bare transducer in air to the bare transducer with its face dipped into water. Thus the quoted efficiency does not account for housing or mounting losses. A similar measurement made for our transducer gave an electroacoustic efficiency, η, of 97% for the longitudinal mode, and 87% for the "flapping" mode. The lower efficiency of Inoue can be attributed to the fact that there are quite high losses

in the (large) matching plate. A comparison of Fig. 10 in Inoue's paper with Fig. 7 of this paper, shows that the multi-resonant transducer has a much smoother pass-band response than the matching-plate transducer. Inoue's results also show a large notch in the upper frequency range of their transducer due to a higher-order longitudinal resonance. The quoted source-level for a single matched-plate transducer was estimated from the measured source-level in Fig. 10 of Inoue's paper. Inoue's measurement is for an array of 9 elements, following the argument in Section IV the SL for one element will be about 149dB - $\log_2(9) \times 6$dB ≈ 130dB/V re 1μPa@1m. In all aspects the multi-resonant design equals or exceeds Inoue's single matching plate design.

VI. Conclusion

In this paper we have shown how to exploit symmetric "flapping" modes in the head-piece of a Tonpilz transducer. Attempts to model this transducer using one-dimensional models has shown the inability of these models to handle flexural modes other than by introducing assumptions of multi-piston resonance [13]. Contrary to the one-dimensional model results, the Finite Element Method correctly predicts the performance characteristics of the transducer.

The "flapping" mode of the transducer was exploited by lip-mounting the transducer in its housing, as opposed to the commonly used nodal mount. The lip-mounted transducer did not show any of the detrimental effects of previous designs. A comparison with a recent design shows that the multi-resonant design can produce a transducer with characteristics that exceed more conventional one-dimensional design methods.

The final Tonpilz design had a broad-bandwidth (low-Q), low-ripple, high-electroacoustic efficiency, high-electromechanical coupling, and is able to be driven at high-power levels. These characteristics are rarely seen in previous Tonpilz designs.

VII. Acknowledgments

Appendix

I. FE model processing demands

Table IV shows the various FE model mesh parameters, memory demands, and processing times. The modal analysis of each model used matrix reduction with 200 master degrees of freedom. The processing time indicated for each modal analysis is the time taken for the extraction of the 5 lowest vibrational modes for both resonance and anti-resonance frequencies and also includes the time taken to write the displacement results as shown in Figs. 1 and 4. The calculation of the transducer admittance required a harmonic solution of the full system equations at stepped frequency locations. These steps were 1kHz apart from 10-70kHz with finer steps about each resonance peak. The larger times for the harmonic analysis relative to the modal solution reflect this multiple solution time. To determine the effects of mesh refinement several axisymmetric models were analyzed. Model A1 is that shown in Fig. 1, model A2 has the same geometry as A1 but has a finer mesh. The models C1 and C2 were solid-model versions with the same geometry as A1/2. The model C2m used for the modal analysis is smaller than the model C2y used to calculate the admittance. In each of the mesh refinement cases the difference in the resonance frequencies and admittance curves was insignificant. Model S1 used for the square-faced transducer in Fig. 4 was considered a reasonable trade-off between mesh density and processing time.

References

[1] T. Inoue, T. Nada, T. Tsuchiya, T. Nakanishi, T. Miyama, and M. Konno, "Tonpilz piezoelectric transducers with acoustic matching plates for underwater color image transmission", *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 40, no. 2, pp. 121–129, Mar. 1993.

[2] R. F. W. Coates, "The design of transducers and arrays for underwater data transmission", *IEEE Journal of Oceanic Engineering*, vol. 16, no. 1, pp. 123–135, Jan. 1991.

[3] M. P. Hayes and P. T. Gough, "Broad-band synthetic aperture sonar", *IEEE Journal of Oceanic Engineering*, vol. 17, no. 1, pp. 80–94, Jan. 1992.

[4] S. G. Schock, L. R. Leblanc, and S. Panda, "Spatial and temporal pulse design considerations for a marine sediment classification sonar", *IEEE Journal of Oceanic Engineering*, vol. 19, no. 3, pp. 406–415, July 1994.

[5] R. Coates, "Underwater acoustic communication", *Sea Technology*, vol. 35, no. 7, pp. 41–47, July 1994.

[6] D. A. Berlincourt, D. R. Curran, and H. Jaffe, "Chapter 3: Piezoelectric and piezomagnetic materials and their function in transducers", *Physical Acoustics*, vol. 1A, pp. 169–270, 1964.

[7] F.V. Hunt, *Electroacoustics: The Analysis of Transduction and its Historical Background*, American Institute of Physics, New York, 1982.

[8] W. P. Mason, *Electromechanical Transducers and Wave Filters*, Van Nostrand, 2nd edition, 1948.

[9] G. E. Martin, "Vibrations of longitudinally polarized ferroelectric cylindrical tubes", *Journal of the Acoustical Society of America*, vol. 35, no. 4, pp. 510–520, Apr. 1964.

[10] O. B. Wilson, *An Introduction to the Theory and Design of Sonar Transducers*, Naval Postgraduate School, Monterey, Ca. 93943, 1985.

[11] D. Boucher, M. Lagier, and C. Maerfeld, "Computation of the vibrational modes for piezoelectric array transducers using a mixed finite element-perturbation method", *IEEE Transactions on Sonics and Ultrasonics*, vol. 28, no. 5, pp. 318–330, Sept. 1981.

[12] J. N. Decarpigny, J. C. Debus, B. Tocquet, and D. Boucher, "In-air analysis of piezoelectric Tonpilz transducers in a wide frequency band using a mixed finite element-plane wave method", *Journal of the Acoustical Society of America*, vol. 78, no. 5, pp. 1499–1507, Nov. 1985.

[13] J. L. Butler, J. R. Cipolla, and W. D. Brown, "Radiating head flexure and its effect on transducer performance", *Journal of the Acoustical Society of America*, vol. 70, no. 2, pp. 500–503, Aug. 1981.

[14] O. C. Zienkiewicz and R. L. Taylor, *The Finite Element Method: Volumes I and II*, McGraw-Hill Book Co., London, 4th edition, 1989.

[15] R. Lerch, "Simulation of piezoelectric devices by two- and three-dimensional finite elements", *IEEE Transactions on Ultrason-*

TABLE IV

FE MODEL MESH PARAMETERS, MEMORY DEMANDS, AND PROCESSING TIMES

Model[a]	Analysis Type	Element[b] Types	Nodes	Elements	Active DOF	Mem. (MB)	Disk (MB)	Proc.[c] Time
A1	Modal	13,42,82	159	101	318	17.5	2.8	4 mins
A1	Harmonic	”	159	101	318	16.6	0.7	30 mins
A2	Modal	”	288	202	576	17.5	3.8	5 mins
A2	Harmonic	”	288	202	576	16.6	1.3	63 mins
C1	Modal	5,45	384	224	1302	17.6	12.4	12 mins
C1	Harmonic	”	384	224	1302	16.8	8.7	5.7 hrs
C2m	Modal	”	456	272	1482	17.6	13.9	13 mins
C2y	Harmonic	”	1614	1200	5488	17.7	48.3	2.2 days
S1	Modal	5,45	386	224	1312	17.6	12.3	11 mins
S1	Harmonic	”	386	224	1312	16.8	8.8	5.8 hrs

[a]A = axisymmetric circular-faced model, C = solid circular-faced model, S = solid square-faced model

[b]see ANSYS manuals [16]

[c]Processed on a DX4/100 PC with 16MB RAM

ics, Ferroelectrics and Frequency Control, vol. 37, no. 2, pp. 233–247, May 1990.

[16] G. J. Desalvo and J. A. Swanson, ANSYS-Finite Element Analysis, Swanson Analysis Systems, Inc., Houston, PA., 1985.

[17] R. Lerch, H. Landes, W. Friedrich, R. Hebel, A. Höβ, and H. Kaarmann, "Modelling of acoustic antennas with a combined finite-element-boundary-element-method", IEEE Ultrasonics Symposium, vol. 1, pp. 581–584, 1992.

[18] R. Coates and R. F. Mathams, "Design of matching networks for acoustic transducers", Ultrasonics, vol. 26, pp. 59–64, Mar. 1988.

[19] M. van Crombrugge and W. Thompson, "Optimization of the transmitting characteristics of a Tonpilz-type transducer by proper choice of impedance matching layers", Journal of the Acoustical Society of America, vol. 77, no. 2, pp. 747–752, Feb. 1985.

[20] H. Allik, K. M. Webman, and J. T. Hunt, "Vibrational response of sonar transducers using piezoelectric finite elements", Journal of the Acoustical Society of America, vol. 56, no. 6, pp. 1782–1791, Dec. 1974.

[21] D. S. Burnett, Finite Element Analysis, Addison-Wesley Publishing Co., 1988.

[22] T. R. Chandrupatla and A. D. Belegundu, Introduction to Finite Elements in Engineering, Prentice-Hall, Inc., 1991.

[23] H. Allik and T. J. R. Hughes, "Finite element method for piezoelectric vibration", International Journal for Numerical Methods in Engineering, vol. 2, pp. 151–157, 1970.

[24] D. F. Ostergaard and T. P. Pawlik, "Three-dimensional finite elements for analyzing piezoelectric structures", IEEE Ultrasonics Symposium, pp. 639–644, 1986.

[25] R. J. Guyan, "Reduction of stiffness and mass matrices", AIAA Journal, vol. 3, no. 2, pp. 380, Feb. 1965.

[26] M. Naillon, F. Besnier, and R. H. Coursant, "Finite element analysis of narrow piezoelectric parallelepiped vibrations energetical coupling modeling", IEEE Ultrasonics Symposium, pp. 773–777, 1983.

[27] E. A. Neppiras, "The pre-stressed piezoelectric sandwich transducer", in Ultrasonics International Conference Proceedings, 1973, pp. 295–302.

[28] J. S. Knight, "Ultrasonic transducer development for a CTFM synthetic aperture sonar", Master's thesis, University of Canterbury, 1987.

[29] L. E. Kinsler and A. R. Frey, Fundamentals of Acoustics, John Wiley and Sons, Inc., 1962.

[30] P. T. Gough and J. S. Knight, "Wide bandwidth, constant beamwidth acoustic projectors: a simplified design procedure", Ultrasonics, vol. 27, pp. 234–238, July 1989.

[31] B. Koyuncu, "A study of PZT-4 transducer surface motion related to their underwater far field beam patterns", Acoustics Letters, vol. 10, no. 4, pp. 51–58, 1986.

[32] C. H. Walter, Travelling Wave Antennas, McGraw-Hill Book Company, 1965.

[33] R. Lerch, "Finite element analysis of piezoelectric transducers", IEEE Ultrasonics Symposium, vol. 2, pp. 643–54, 1988.

[34] B. V. Smith and B. K. Gazey, "High-frequency sonar transducers: a review of current practice", IEE Proceedings, vol. 131, no. 3, pp. 285–297, June 1984, Part F.

[35] W. S. Burdic, Underwater Acoustic System Analysis, Prentice-Hall, Inc., 1984.

David W. Hawkins was born in Waipukurau, New Zealand, in 1970. He received the B.Sc. degree (physics) in 1991 from Victoria University, Wellington and is currently studying for a Ph.D. in the Electrical and Electronic Engineering Department of Canterbury University, Christchurch.

Peter T. Gough was born in Dunedin, New Zealand in 1947 and received his B.E.(1st class honours) in 1970 from the University of Canterbury. He was awarded a Ph.D. from the same institution in 1974 after which he spent some 5 years as a post-doctoral fellow and then as an Assistant Professor at the Institute of Optics, University of Rochester, Rochester, New York and at the University of Manitoba, Winnipeg, Canada. After working as a consultant in California, he returned to the Electrical and Electronic Engineering Department, University of Canterbury, in 1980 where he is now the Head of Department. He is a Fellow of the Institute of Professional Engineers of New Zealand.

REFERENCES

[1] A. E. Adams, O. R. Hinton, M. A. Lawlor, B. S. Sharif, and V. S. Riyait. A synthetic aperture sonar image processing system. *IEE Acoustic Sensing and Imaging Conference*, pages 109–113, March 1993.

[2] A. E. Adams, M. A. Lawlor, V. S. Riyait, O. R. Hinton, and B. S. Sharif. Real-time synthetic aperture sonar processing system. *IEEE Proceedings on Radar, Sonar, and Navigation*, 143(3):169–176, June 1996.

[3] AMP Incorporated, P.O. Box 799, Valley Forge, PA 19482. *Piezoelectric copolymer hydrophone tile (hydrostatic mode)*, 1993. Application specification 114-1082 (01 FEB 94 Rev B).

[4] D.K. Anthony, C.F.N. Cowan, H.D. Griffiths, Z. Meng, T.A. Rafik, and H. Shafeeu. 3-D high resolution imaging using interferometric synthetic aperture sonar. *Proceedings of the Institute of Acoustics*, 17(8):11–20, December 1995.

[5] D.A. Ausherman, A. Kozma, J.L. Walker, H.M. Jones, and E.C. Poggio. Developments in radar imaging. *IEEE Transactions on Aerospace and Electronic Systems*, 20(4):363–399, July 1984.

[6] R. Bamler. A comparison of range-Doppler and wavenumber domain SAR focussing algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 30(4):706–713, July 1992.

[7] B. C. Barber. Theory of digital imaging from orbital synthetic aperture data. *International Journal of Remote Sensing*, 6(7):1009–1057, 1985.

[8] R. H. T. Bates, V. A. Smith, and R. D. Murch. Manageable multidimensional inverse scattering theory. *Physics Reports*, 201(4):187–277, April 1991.

[9] D. Blacknell, A. Freeman, S. Quegan, I. A. Ward, I. P. Finley, C. J. Oliver, R. G. White, and J. W. Wood. Geometric accuracy in airborne SAR images. *IEEE Transactions on Aerospace and Electronic Systems*, 25(2):241–258, March 1989.

[10] D. Blacknell, A. Freeman, R. G. White, and J. W. Wood. The prediction of geometric distortions in airbourne synthetic aperture radar imagery from autofocus measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 25(6):775–781, November 1987.

[11] D. Blacknell and S. Quegan. Motion compensation of airborne synthetic aperture radars using autofocus. *GEC Journal of Research*, 7(3):168–182, 1990.

[12] R. N. Bracewell. *The Fourier transform and its applications*. McGraw-Hill Book Company, 1986.

[13] E. Brookner, editor. *Radar Technology*. Artech House, Inc, 1978.

[14] W. D. Brown and D. C. Ghiglia. Some methods for reducing propagation-induced phase errors in coherent imaging systems. i. formalism. *Journal of the Optical Society of America*, 5(6):924–941, June 1988.

[15] W. M. Brown and R. J. Fredricks. Range-Doppler imaging with motion through resolution cells. *IEEE Transactions on Aerospace and Electronic Systems*, 5(1):98–102, January 1969.

[16] W. M. Brown, G. G. Houser, and R. E. Jenkins. Synthetic aperture processing with limited storage and presumming. *IEEE Transactions on Aerospace and Electronic Systems*, 9(2):166–176, 1972.

[17] W. M. Brown and L. J. Porcello. An introduction to synthetic-aperture radar. *IEEE Spectrum*, pages 52–62, September 1969.

[18] M. P. Bruce. A processing requirement and resolution capability comparison of side-scan and synthetic aperture sonars. *IEEE Journal of Oceanic Engineering*, 17(1):106–117, January 1992.

[19] C. B. Burckhardt, P. Grandchamp, and H. Hoffman. An experimental 2MHz synthetic aperture sonar system intended for medical use. *IEEE Transactions on Sonics and Ultrasonics*, 21(1):1–6, January 1974.

[20] C. Cafforio, C. Pratti, and F. Rocca. SAR data focussing using seismic migration techniques. *IEEE Transactions on Aerospace and Electronic Systems*, 27(2):194–207, March 1991.

[21] W. G. Carrara, R. S. Goodman, and R. M. Majewski. *Spotlight synthetic aperture radar: signal processing algorithms*. Artech House, 1995.

[22] J. Chatillon, M. E. Bouhier, and M. E. Zakharia. Synthetic aperture sonar: wide band vs narrow band. *Proceedings U.D.T. Conference, Paris*, pages 1220–1225, 1991.

[23] J. Chatillon, M. E. Bouhier, and M. E. Zakharia. Synthetic aperture sonar for seabed imaging: relative merits of narrow-band and wide-band approaches. *IEEE Journal of Oceanic Engineering*, 17(1):95–105, January 1992.

[24] D. G. Checketts and B. V. Smith. Analysis of the effects of platform motion errors upon synthetic aperture sonar. *Proceedings of the Institute of Acoustics*, 8(3):135–143, 1988.

[25] D. A. Christensen. *Ultrasonic bioinstrumentation*. Wiley, 1988.

[26] J.T. Christoff, C.D. Loggins, and E.L. Pipkin. Measurement of the temporal phase stability of the medium. *Journal of the Acoustical Society of America*, 71(6):1606–1607, June 1982.

[27] Personal communication with Jim Christoff. Coastal Systems Station Dahlgran Division, Naval Surface Warfare Center, 6703 West Highway 98, Panama City, FL 32407-9701, June 1996.

[28] C. E. Cook and M. Bernfeld. *Radar signals: an introduction to theory and application*. Academic Press Inc., 1967.

[29] R. Crochiere and L. R. Rabiner. *Multirate digital signal processing*. Prentice-Hall, 1993.

[30] I. Cumming, F. Wong, and R. K. Raney. A SAR processing algorithm with no interpolation. *International Geoscience and Remote Sensing Symposium*, 1:376–379, 1992.

[31] J. C. Curlander and R. N. McDonough. *Synthetic Aperture Radar: systems and signal processing*. John Wiley & Sons, Inc., 1991.

[32] L. J. Cutrona. Comparison of sonar system performance achievable using synthetic-aperture techniques with the performance achievable by more conventional means. *Journal of the Acoustical Society of America*, 58(2):336–348, August 1975.

[33] L. J. Cutrona. Additional characteristics of synthetic-aperture sonar systems and a further comparison with nonsynthetic-aperture sonar systems. *Journal of the Acoustical Society of America*, 61(5):1213–1217, May 1977.

[34] L.J. Cutrona, E.N. Leith, L.J. Porcello, and W.E. Vivian. On the application of coherent optical processing techniques to synthetic-aperture radar. *Proceedings of the IEEE*, 54(8):1026–1032, August 1966.

[35] L.J. Cutrona, W.E. Vivian, E.N. Leith, and G.O. Hall. A high-resolution radar combat-surveillance system. *IRE Transactions on Military Electronics*, 5:127–131, April 1961.

[36] G. W. Davidson, I. G. Cumming, and M. R. Ito. A chirp scaling approach for processing squint mode SAR data. *IEEE Transactions on Aerospace and Electronic Systems*, 32(1):121–133, January 1996.

[37] P. de Heering. A synthetic aperture sonar study. Technical report, Huntech-Lapp Systems Limited, Scarborough, Ontario, MIR 3A6, August 1982.

[38] P. de Heering. Alternative schemes in synthetic aperture sonar processing. *IEEE Journal of Oceanic Engineering*, 9(4):277–280, October 1984.

[39] A. de Roos, J. J. Sinton, P. T. Gough, W. K. Kennedy, and M. J. Cusdin. The detection and classification of objects lying on the seafloor. *Journal of the Acoustical Society of America*, 84(4):1456–1477, October 1988.

[40] B. L. Douglas and H. Lee. Synthetic aperture active sonar imaging. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, III:III/37–40, 1992.

[41] B. L. Douglas and H. Lee. Motion compensation for improved sidescan sonar imaging. *IEEE Oceans '93 Conference Proceedings*, I:I/378–383, 1993.

[42] B. L. Douglas and H. Lee. Synthetic-aperture sonar imaging with a multiple-element receiver array. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 5:445–448, April 1993.

[43] B.L. Douglas. *Signal processing for synthetic aperture sonar imaging*. PhD thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1993.

[44] P. H. Eichel, D. C. Ghiglia, and C. V. Jakowatz. Speckle processing method for synthetic-aperture-radar phase correction. *Optics Letters*, 14(1):1–3, January 1989.

[45] P. H. Eichel, D. C. Ghiglia, and C. V. Jakowatz. Phase correction system for automatic focusing of synthetic aperture radar. U.S. Patent 4,924,229, May 1990.

[46] P. H. Eichel and C. V. Jakowatz. Phase-gradient algorithm as an optimal estimator of the phase derivative. *Optics Letters*, 14(20):1101–1103, October 1989.

[47] C. Elachi. *Spaceborne Radar Remote Sensing: Applications and Techniques*. IEEE Press, 1988.

[48] C. Elachi, T. Bicknell, R. L. Jordan, and C. Wu. Spaceborne synthetic-aperture imaging radars: applications, techniques, and technology. *Proceedings of the IEEE*, 70(10):1174–1209, October 1982.

[49] C. Elachi and D. D. Evans. Effects of random phase changes on the formation of synthetic aperture radar imagery. *IEEE Transactions on Antennas and propagation*, pages 149–153, January 1977.

[50] A. B. E. Ellis. The processing of synthetic aperture radar signals. *GEC Journal of Research*, 2(3):169–177, 1984.

[51] J. Franca, A. Petraglia, and S. K. Mitra. Multirate analog-digital systems for signal processing and conversion. *Proceedings of the IEEE*, 85(2):242–262, February 1997.

[52] A. Freeman. SAR calibration: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, 30(6):1107–1121, November 1992.

[53] D. C. Ghiglia and W. D. Brown. Some methods for reducing propagation-induced phase errors in coherent imaging systems. ii. numerical results. *Journal of the Optical Society of America*, 5(6):942–957, June 1988.

[54] D. C. Ghiglia and G. A. Mastin. Two-dimensional phase correction of synthetic-aperture-radar imagery. *Optics Letters*, 14(20):1104–1106, October 1989.

[55] D. C. Ghiglia and L. A. Romero. Direct phase estimation from phase differences using fast elliptic partial differential equation solvers. *Optics Letters*, 14(20):1107–1109, October 1989.

[56] G. A. Gilmour. Synthetic aperture side-looking sonar system. *Journal of the Acoustical Society of America*, 65(2):557, May 1978. Review of U.S. Patent 4,088,978.

[57] R. C. Gonzalez and P. Wintz. *Digital Image Processing*. Addison-Wesley Publishing Co., Inc., 2nd edition, 1987.

[58] J. W. Goodman. *Introduction to Fourier optics*. McGraw-Hill Book Company., 1968.

[59] P. T. Gough. A synthetic aperture sonar system capable of operating at high speed and in turbulent media. *IEEE Journal of Oceanic Engineering*, 11(2):333–339, April 1986.

[60] P. T. Gough, A. de Roos, and M. J. Cusdin. Continuous transmission FM sonar with one octave bandwidth and no blind time. *IEE Proceedings, Part F*, 131(3):270–274, June 1984.

[61] P. T. Gough and J. S. Knight. Wide bandwidth, constant beamwidth acoustic projectors: a simplified design procedure. *Ultrasonics*, 27:234–238, July 1989.

[62] P.T. Gough and M.P. Hayes. Measurements of the acoustic phase stability in Loch Linnhe, Scotland. *Journal of the Acoustical Society of America*, 86(2):837–839, August 1989.

[63] P.T. Gough and M.P. Hayes. Test results using a prototype synthetic aperture sonar. *Journal of the Acoustical Society of America*, 86(6):2328–2333, December 1989.

[64] L.C. Graham. Synthetic interferometric radar for topographic mapping. *Proceedings of the IEEE*, 62(6):763–768, June 1974.

[65] H. D. Griffiths. New ideas in FM radar. *Electronics and Communication Engineering Journal*, 2(5):185–194, October 1990.

[66] H. D. Griffiths, J. W. R. Griffiths, Z. Meng, C. F. N. Cowan, T. A. Rafik, and H. Shafeeu. Interferometric synthetic aperture sonar for high-resolution 3-D imaging. *Proceedings of the Institute of Acoustics*, 16(6):151–159, December 1994.

[67] H. D. Griffiths and L. Vinagre. Design of low-sidelobe pulse compression waveforms. *Electronics Letters*, 30(12):1004–1005, June 1994.

[68] S. Guyonic. Experiments on a sonar with a synthetic aperture array moving on a rail. *IEEE Oceans '94 Conference Proceedings*, 3:571–576, 1994.

[69] R. O. Harger. *Synthetic aperture radar systems: theory and design*. Academic Press, 1970.

[70] B. Harris and S. A. Kramer. Asymptotic evaluation of the ambiguity functions of high-gain FM matched filter sonar systems. *Proceedings of the IEEE*, 56(12):2149–2157, December 1968.

[71] F. J. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66(1):51–83, January 1978.

[72] D. W. Hawkins and P. T. Gough. Recent sea trials of a synthetic aperture sonar. *Proceedings of the Institute of Acoustics*, 17(8):1–10, December 1995.

[73] D. W. Hawkins and P. T. Gough. Multiresonance design of a Tonpilz transducer using the finite element method. *IEEE Ultrasonics, Ferroelectrics, and Frequency Control*, 43(5):????–?????, September 1996.

[74] M. P. Hayes and P. T. Gough. Broad-band synthetic aperture sonar. *IEEE Journal of Oceanic Engineering*, 17(1):80–94, January 1992.

[75] M.P. Hayes. *A CTFM synthetic aperture sonar*. PhD thesis, Department of Electrical and Electronics Engineering, University of Canterbury, Christchurch, New Zealand, September 1989.

[76] S. Haykin. *Communication Systems*. John Wiley and Sons, Inc, 3rd edition, 1994.

[77] J. Hermand and W. I. Roderick. Delay-doppler resolution performance of large time-bandwidth-product linear FM signals in a multipath ocean environment. *Journal of the Acoustical Society of America*, 84(5):1709–1727, November 1988.

[78] B. D. Huxtable and E. M. Geyer. Motion compensation feasibility for high resolution synthetic aperture sonar. *IEEE Oceans '93 Conference Proceedings*, 1:125–131, October 1993.

[79] T. Isernia, V. Pascazio, R. Pierri, and G. Schirinzi. Image reconstruction from fourier transform magnitude with applications to synthetic aperture radar imaging. *Journal of the Optical Society of America*, 13(5):922–934, May 1996.

[80] T. Isernia, V. Pascazio, R. Pierri, and G. Schirinzi. Synthetic aperture radar imaging from phase corrupted data. *IEE Proceedings on Radar, Sonar, and Navigation*, 143(4):268–274, August 1996.

[81] C. V. Jakowatz and D. E. Wahl. Eigenvector method for maximum-likelihood estimation of phase errors in synthetic-aperture-radar imagery. *Journal of the Optical Society of America*, 10(2):2539–2546, December 1993.

[82] C. V. Jakowatz, D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and P. A. Thompson. *Spotlight-mode synthetic aperture radar: A signal processing approach.* Kluwer Academic Publishers, Boston, 1996.

[83] C.V. Jakowatz and P.A. Thompson. A new look at spotlight mode synthetic aperture radar as tomography: imaging 3-D targets. *IEEE Transactions on Image Processing*, 4(5):699–703, May 1995.

[84] M. Y. Jin and C. Wu. A SAR correlation algorithm which accommodates large-range migration. *IEEE Transactions on Geoscience and Remote Sensing*, 22(6):592–597, November 1984.

[85] R. L. Jordan. The Seasat-A synthetic aperture radar system. *IEEE Journal of Oceanic Engineering*, 5(2):154–164, April 1980.

[86] M. Karaman, P. Li, and M. O'Donnell. Synthetic aperture imaging for small scale systems. *IEEE Ultrasonics, Ferroelectrics, and Frequency Control*, 42(3):429–442, May 1995.

[87] J. C. Kirk. A discussion of digital processing in synthetic aperture radar. *IEEE Transactions on Aerospace and Electronic Systems*, 11(3):326–337, May 1975.

[88] J. C. Kirk. Motion compensation for synthetic aperture radar. *IEEE Transactions on Aerospace and Electronic Systems*, 11(3):338–348, May 1975.

[89] W. E. Kock. Extending the maximum range of synthetic aperture (hologram) systems. *Proceedings of the IEEE*, 60:1459–1460, November 1972.

[90] S. A. Kramer. Doppler and acceleration tolerances of high-gain, wideband linear FM correlation sonars. *Proceedings of the IEEE*, 55(5):627–636, May 1967.

[91] J. J. Kroszczynski. Pulse compression by means of linear-period modulation. *Proceedings of the IEEE*, 57(7):1260–66, July 1969.

[92] R. W. Larson, P. L. Jackson, and E. S. Kasischke. A digital calibration method for synthetic aperture radar systems. *IEEE Transactions on Geoscience and Remote Sensing*, 26(6):753–763, November 1988.

[93] M. A. Lawlor, A. E. Adams, O. R. Hinton, V. S. Riyait, and B. Sharif. Methods for increasing the azimuth resolution and mapping rate of a synthetic aperture sonar. *IEEE Oceans '94 Conference Proceedings*, 3:565–570, 1994.

[94] H. E. Lee. Extension of synthetic aperture radar (SAR) techniques to undersea applications. *IEEE Journal of Oceanic Engineering*, 4(2):60–63, April 1979.

[95] A. Li. Algorithms for the implementation of Stolt interpolation in SAR processing. *International Geoscience and Remote Sensing Symposium*, 1:360–362, 1992.

[96] F. K. Li, C. Croft, and D. N. Held. Comparison of several techniques to obtain multiple-look SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 21(3):370–375, July 1983.

[97] F. K. Li, D. N. Herd, J. C. Curlander, and C. Wu. Doppler parameter estimation for spaceborne synthetic-aperture radars. *IEEE Transactions on Geoscience and Remote Sensing*, 23(1):47–56, January 1985.

[98] F. K. Li and W. T. K. Johnson. Ambiguities in spaceborne synthetic aperture radar systems. *IEEE Transactions on Aerospace and Electronic Systems*, 19(3):389–397, May 1983.

[99] Z. Lin. Wideband ambiguity function of broadband signals. *Journal of the Acoustical Society of America*, 83(6):2108–2116, June 1988.

[100] A. P. Luscombe. Auxiliary data networks for satellite synthetic aperture radar. *The Marconi Review*, 45(225):84–105, Second Quarter 1982.

[101] The Mathworks, Inc., 24 Prime Park Way, Natick, MAss. 01760-1500. *MATLAB*.

[102] R. N. McDonough, B. E. Raff, and J. L. Kerr. Image formation from spaceborne synthetic aperture radar signals. *Johns Hopkins APL Technical Digest*, 6(4):300–312, 1987.

[103] J. G. Mehlis. Synthetic aperture radar range-azimuth ambiguity design and constraints. *IEEE International Radar Conference*, pages 143–152, 1980.

[104] Z. Meng. *A study on synthetic aperture sonar*. PhD thesis, Loughborough University of Technology, Loughborough, England, January 1995.

[105] D.L. Mensa. *High resolution radar imaging*. Artech House, Inc., 1981.

[106] J. H. Mims and J. F. Farrell. Synthetic aperture imaging with maneuvers. *IEEE Transactions on Aerospace and Electronic Systems*, 8(4):410–418, July 1972.

[107] A. Moreira. Improved multilook techniques applied to SAR and SCANSAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 29(4):529–534, July 1991.

[108] A. Moreira. Removing the azimuth ambiguities of point targets in synthetic aperture radar images. *International Geoscience and Remote Sensing Symposium*, 1:614–616, 1992.

[109] A. Moreira. Supressing the azimuth ambiguities in synthetic aperture radar images. *IEEE Transactions on Geoscience and Remote Sensing*, 31(4):885–895, July 1993.

[110] A. Moreira. Airborne SAR processing of highly squinted data using a chirp scaling approach with integrated motion compensation. *IEEE Transactions on Geoscience and Remote Sensing*, 32(5):1029–1040, September 1994.

[111] A. Moreira. Method of image generation by means of two-dimensional data processing in connection with radar with synthetic aperture. Canadian Patent Application 2,155,502, 1995.

[112] A. Moreira and T. Misra. On the use of the ideal filter concept for improving SAR image quality. *Journal of Electromagnetic Waves and Applications*, 9(3):407–420, 1995.

[113] J. R. Moreira. A new method of aircraft motion error extraction from radar raw data for real time motion compensation. *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):620–626, July 1990.

[114] D. C. Munson, J. D. O'Brien, and W. K. Jenkins. A tomographic formulation of spotlight-mode synthetic aperture radar. *Proceedings of the IEEE*, 71(8):917–925, August 1983.

[115] D. C. Munson and J. L. C. Sanz. Image reconstruction from frequency-offset Fourier data. *Proceedings of the IEEE*, 72:661–669, June 1984.

[116] D. C. Munson and R. L. Visentin. A signal processing view of strip-mapping synthetic aperture radar. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(12):2131–2147, December 1989.

[117] J. C. Nelander, A. C. Kenton, and J. A. Wright. A vertical beamforming design approach for increased area coverage rate for synthetic aperture sonar. *IEEE Proceedings of Southeastcon*, pages 42–47, April 1989.

[118] M. Neudorfer, T. Luk, D. Garrood, N. Lehtomaki, and M. Rognstad. Acoustic detection and classification of buried unexploded ordinance (uxo) using synthetic aperture sonar. *PACON '96*, 1996.

[119] L. F. Nock and G. E. Trahey. Synthetic receive aperture imaging with phase correction for motion and tissue inhomogeneities–Part I: basic principles. *IEEE Ultrasonics, Ferroelectrics, and Frequency Control*, 39(4):489–495, July 1992.

[120] C. J. Oliver. Synthetic-aperture radar imaging. *Journal of Physics D: Applied Physics*, 22:871–890, 1989.

[121] C. J. Oliver. Information from SAR images. *Journal of Physics D: Applied Physics*, 24:1493–1514, 1991.

[122] H. W. Ott. *Noise reduction in electronic systems.* John Wiley and Sons, 2nd edition, 1988.

[123] A. Papoulis. *Systems and transforms with applications in optics.* McGraw Hill, 1968.

[124] M. Pollakowski and H. Ermert. Chirp signal matching and signal power optimisation in pulse-echo mode ultrasonic non-destructive testing. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 41(5):655–659, September 1995.

[125] M. Pollakowski, H. Ermert, L. von Bernus, and T. Schmeidl. The optimum bandwidth of chirp signals in ultrasonic applications. *Ultrasonics*, 31(6):417–420, 1993.

[126] R. K. Raney. An exact wide field digital imaging algorithm. *International Journal of Remote Sensing*, 13(5):991–998, 1992.

[127] R. K. Raney. A new and fundamental Fourier transform pair. *International Geoscience and Remote Sensing Symposium*, 1:106–107, 1992.

[128] R. K. Raney, H. Runge, R. Bamler, I. G. Cumming, and F. H. Wong. Precision SAR processing using chirp scaling. *IEEE Transactions on Geoscience and Remote Sensing*, 32(4):786–799, July 1994.

[129] A. W. Rihaczek. *Principles of high resolution radar.* McGraw Hill, Inc, 1969.

[130] V. S. Riyait, M. A. Lawlor, A. E. Adams, O. R. Hinton, and B. S. Sharif. Comparison of the mapping resolution of the ACID synthetic aperture sonar with existing sidescan sonar systems. *IEEE Oceans '94 Conference Proceedings*, 3:???–???, 1994.

[131] V.S. Riyait, M.A. Lawlor, A.E. Adams, O. Hinton, and B. Sharif. Real-time synthetic aperture sonar imaging using a parallel architecture. *IEEE Transactions on Image Processing*, 4(7):1010–1019, July 1995.

[132] K. D. Rolt and H. Schmidt. Azimuthal ambiguities in synthetic aperture sonar and synthetic aperture radar imagery. *IEEE Journal of Oceanic Engineering*, 17(1):73–79, January 1992.

[133] K.D. Rolt. Ocean, platform, and signal processing effects on synthetic aperture sonar performance. Scientiae Magister Thesis, Department of Ocean Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, February 1991.

[134] H. Runge and R. Bamler. A novel high precision SAR focussing algorithm based on chirp scaling. *International Geoscience and Remote Sensing Symposium*, 1:372–375, 1992.

[135] T. Sato and O. Ikeda. Sequential synthetic aperture sonar system–a prototype of a synthetic aperture sonar system. *IEEE Transactions on Sonics and Ultrasonics*, 24(4):253–259, July 1977.

[136] T. Sato and O. Ikeda. Super-resolution ultrasonic imaging by combined spectral and aperture synthesis. *Journal of the Acoustical Society of America*, 62(2):341–345, August 1977.

[137] T. Sato, M. Ueda, and S. Fukuda. Synthetic aperture sonar. *Journal of the Acoustical Society of America*, 54(3):799–802, 1973.

[138] R.W. Sheriff. Synthetic aperture beamforming with automatic phase compensation for high frequency sonars. *IEEE Proceedings of the 1992 Symposium on Autonomous Underwater Vehicle Technology*, pages 236–245, June 1992.

[139] C.W. Sherwin, J.P. Ruina, and R.D. Rawcliffe. Some early developments in synthetic aperture radar systems. *IRE Transactions on Military Electronics*, 6:111–115, April 1962.

[140] G. Shippey and T. Nordkvist. Phased array acoustic imaging in ground co-ordinates, with extension to synthetic aperture processing. *IEE Proceedings on Radar, Sonar, and Navigation*, 143(3):131–139, June 1996.

[141] J. M. Silkaitis, B. L. Douglas, and H. Lee. A cascade algorithm for estimating and compensating motion error for synthetic aperture sonar imaging. *IEEE Proceedings ICIP*, 1:905–909, November 1994.

[142] M. Skolnik, editor. *Radar Handbook.* McGraw Hill, Inc., 2nd edition, 1990.

[143] M. L. Somers and A. R. Stubbs. Sidescan sonar. *IEE Proceedings, Part F*, 131(3):243–256, June 1984.

[144] M. Soumekh. A system model and inversion for synthetic aperture radar imaging. *IEEE Transactions on Image Processing*, 1(1):64–76, January 1992.

[145] M. Soumekh. Digital spotlighting and coherent subaperture image formation for stripmap synthetic aperture systems. *IEEE Proceedings of the International Conference on Image Processing*, 1:476–480, November 1994.

[146] M. Soumekh. *Fourier Array Imaging.* Prentice Hall, Englewood Cliffs, NJ, 1994.

[147] M. Soumekh. Reconnaissance with ultra wideband UHF synthetic aperture radar. *IEEE Signal Processing Magazine*, 12(4):21–40, July 1995.

[148] M. Soumekh. Reconnaissance with slant plane circular SAR imaging. *IEEE Transactions on Image Processing*, 5(8):1252–1265, August 1996.

[149] H. C. Stankwitz, R. J. Dallaire, and J. R. Feinup. Nonlinear apodization for sidelobe control in SAR imagery. *IEEE Transactions on Aerospace and Electronic Systems*, 31(1):267–279, January 1995.

[150] B. D. Steinberg. *Principles of aperture and array system design.* John Wiley and Sons, 1976.

[151] R. H. Stolt. Migration by Fourier transform. *Geophysics*, 43(1):23–48, February 1978.

[152] D. W. Stowe, L. H. Wallman, J. W. Follin Jr., and P. J. Luke. Stability of the acoustic pathlength of the ocean deduced from the medium stability experiment. Technical Report TG 1230, Applied Physics Lab., the John Hopkins University, January 1974.

[153] K. Tomiyasu. Tutorial review of synthetic-aperture-radar (SAR) with applications to imaging of the ocean surface. *Proceedings of the IEEE*, 66(5):563–583, May 1978.

[154] G. E. Trahey and L. F. Nock. Synthetic receive aperture imaging with phase correction for motion and for tissue inhomogeneities–Part II: effects of and correction for motion. *IEEE Ultrasonics, Ferroelectrics, and Frequency Control*, 39(4):496–501, July 1992.

[155] T. K. Truong, I. S. Reed, R. G. Lipes, A. L. Rubin, and S. A. Butman. Digital SAR processing using a fast polynomial transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(2):419–425, April 1984.

[156] R. J. Urick. *Principles of underwater sound.* McGraw Hill Book Company, 2nd edition, 1975.

[157] H. Urkowitz, C. A. Hauer, and J. F. Koval. Generalized resolution in radar systems. *Proceedings of the IRE*, 50(10):2093–2105, October 1962.

[158] D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and C. V. Jakowatz. Phase gradient autofocus - a robust tool for high resolution SAR phase correction. *IEEE Transactions on Aerospace and Electronic Systems*, 30(3):827–835, July 1994.

[159] D. E. Wahl, C. V. Jakowatz, and P. A. Thompson. New approach to strip-map SAR autofocus. *Sixth IEEE Digital Signal Processing Workshop*, pages 53–56, October 1994.

[160] J.L. Walker. Range-doppler imaging of rotating objects. *IEEE Transactions on Aerospace and Electronic Systems*, 16(1):23–52, January 1980.

[161] G. M. Walsh. Acoustic mapping apparatus. *Journal of the Acoustical Society of America*, 47(5):1205, December 1969. Review of U.S. Patent 3,484,737.

[162] S. Wang, M. L. Grabb, and T. G. Birdsall. Design of periodic signals using fm sweeps and amplitude modulation for ocean acoustic travel-time measurements. *IEEE Journal of Oceanic Engineering*, 19(4):611–618, October 1994.

[163] D. B. Ward and R. A. Kennedy. Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns. *Journal of the Acoustical Society of America*, 97(2):1023–1034, February 1995.

[164] C.A. Wiley. Synthetic aperture radars. *IEEE Transactions on Aerospace and Electronic Systems*, 21(3):440–443, May 1985.

[165] R.E. Williams. Creating an acoustic synthetic aperture in the ocean. *Journal of the Acoustical Society of America*, 60(1):60–73, July 1976.

[166] C. Wu, K. Y. Liu, and M. Jin. Modeling and a correlation algorithm for spaceborne SAR signals. *IEEE Transactions on Aerospace and Electronic Systems*, 18(5):563–575, September 1982.

[167] M. E. Zakharia, J. Chatillon, and M. E. Bouhier. Synthetic aperture sonar: a wide-band approach. *Proceedings of the IEEE Ultrasonics Symposium*, pages 1133–1136, 1990.