Electronic sector scan using multiple beam-forming networks

This item was submitted to Loughborough University’s Institutional Repository by the/an author.

Additional Information:

- A Doctoral Thesis. Submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy at Loughborough University.

Metadata Record: https://dspace.lboro.ac.uk/2134/36180

Publisher: © Anthony John Copping

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.
ELECTRONIC SECTOR SCAN

USING MULTIPLE BEAM-FORMING NETWORKS

by

ANTHONY JOHN COPPING, B.Sc.(Hons.)


Supervisor: A.R. PRATT, Ph.D.

Department of Electronic and Electrical Engineering.

SUMMARY

A particular class of multiple beam-forming network, which performs the discrete Fourier transform operation on the outputs of a linear receiving array, is examined in detail. A general synthesis procedure, based on the 'fast Fourier transform' method of computing the discrete Fourier transform, is described, which also encompasses the microwave networks known as Butler matrices.

The problem of generating a within-pulse sector scan display from the outputs of a multiple beam-forming network used in a sonar receiving system is discussed, and one particular technique and its practical implementation is examined in detail.

The design of two, pulsed carrier, sonar receiving systems is described. One, working at a carrier frequency of 500 kHz, has a digital beam-forming processor. The other, working at a carrier frequency of 32 kHz, has an analogue beam-forming processor which makes extensive use of low cost microelectronic devices. The practical limitations of both systems are investigated with a view to providing design data for future work.

Experimental results, obtained from the digital system working in a water tank, are presented. In the absence of a transducer array working at 32 kHz, the analogue processor was tested by using artificially generated signals which simulated the outputs of a receiving array.
ACKNOWLEDGEMENTS

I would like to thank Professor J.W.R. Griffiths for suggesting the topic under investigation in this thesis, for helpful comments, and for the use of the facilities in his department. My thanks are also due to my supervisor, Dr. A.R. Pratt, for the useful discussions during my three years at Loughborough. Lastly, a thank you to my wife, whose help in the preparation of this document was invaluable.

Unless otherwise acknowledged, the content of this thesis is the original work of the author. The network synthesis procedure and the resulting analogue processor for use in a sonar receiver have not been described elsewhere. The processor, in conjunction with the particular method suggested for the generation of a sector scan display, has resulted in a new receiver which is worthy of further investigation for commercial fish-finding applications.
| CONTENTS |
|------------------------|--------|
| **CHAPTER 1. Introduction** | 1 |
| **CHAPTER 2. The Aperture and the Fourier Transform** | 3 |
| 2.1 The Field Model. | 3 |
| 2.2 Fourier Transform Relations for a Continuous Aperture. | 5 |
| 2.3 The Sampled Aperture. | 8 |
| 2.4 Computation of the Fourier Transform. | 11 |
| 2.5 The Discrete Fourier Transform. | 13 |
| 2.6 The DFT and Multiple Beam Formation. | 14 |
| **CHAPTER 3. Synthesis of DFT Networks** | 17 |
| 3.1 Introduction. | 17 |
| 3.2 Notation. | 17 |
| 3.3 Matrix Representation of Beam-Forming Networks. | 18 |
| 3.4 The Fast Fourier Transform. | 19 |
| 3.5 Network Synthesis. | 23 |
| 3.5.1 The DFT Network of Dimension 8. | 25 |
| 3.6 Advantages of the FFT. | 29 |
| 3.6.1 Digital Implementation of the FFT. | 29 |
| 3.6.2 Analogue Implementation of the FFT. | 34 |
| 3.7 Extension of Synthesis using Networks Related to the DFT. | 37 |
| 3.7.1 Related Networks. | 38 |
| 3.7.2 A Uniform Phase Progression across Device Inputs. | 41 |
| 3.7.3 An 8-element Microelectronic Beam-Forming Network. | 46 |
| 3.8 The Reciprocal Nature of the FFT Schematic. | 46 |
| **CHAPTER 4. The Reconstruction of the Far Field Source Distribution** | 48 |
| 4.1 Introduction. | 48 |
| 4.2 The Spectrum of the Far Field Source Distribution. | 48 |
| 4.3 Reconstruction as a Linear Complex Filtering Process. | 50 |
| 4.3.1 A Complex Filter. | 52 |
| 4.4 Reconstruction as a Linear Real Filtering Process. | 54 |
| 4.4.1 Shifting the Origin of the Aperture. | 54 |
4.5 Generation of the Intensity Modulation Waveform.
4.5.1 A Simplification.
4.6 Choice of Filter.
4.6.1 Amplitude and Phase Distortion within the Pass-Band.
   4.6.1.1 The Zero-Order Hold.
   4.6.1.2 An Illustration.
4.6.2 Transient Effects.
   4.6.2.1 An Illustration.
   4.6.2.2 Insertion of a Guard Band.

CHAPTER 5. Receiver Design

5.1 Introduction.
5.2 A Receiving System using a Digital Processor.
   5.2.1 Channel Amplifier.
   5.2.2 Demodulator and Low Pass Filter.
   5.2.3 Track/Store and Multiplex Unit.
   5.2.4 The A.C.C. Amplifier.
   5.2.5 The Digital Processor.
   5.2.6 The Interpolating Filter and Detector.
   5.2.7 Control Unit.
5.3 A Receiving System Using an Analogue Processor.
   5.3.1 The Analogue Processor.

CHAPTER 6. The Estimation of Receiver Performance

6.1 Introduction.
6.2 Measures of Performance.
   6.2.1 Mean Square Error.
   6.2.2 Beam- Pattern Characteristics.
   6.2.3 Comparison of Performance Criteria.
6.3 The Performance of an Ideal Processor.
   6.3.1 The Noise Model.
   6.3.2 The Performance before the Detector.
   6.3.3 The Performance after the Detector.
6.4 The Performance of the Digital Processor.  
  6.4.1 Analogue to Digital Conversion Error.  
  6.4.2 Processor Error.  
    6.4.2.1 Scaling Error in Addition and Subtraction.  
    6.4.2.2 Multiplication Error.  
  6.4.3 Digital Processor Performance Graphs.  
6.5 The Performance of the Analogue Processor.  
  6.5.1 The Description of Practical Units.  
  6.5.2 Plane Wave Response of Processor.  
  6.5.3 The Noise Response of the Processor.  
    6.5.3.1 Amplifier Gain Variations.  
  6.5.4 Performance Graph for an Analogue Processor.  
6.6 The Error Introduced by a Non-Ideal Interpolating Filter.  
  6.6.1 Response to a Plane Wave Signal in the Absence of Noise.  
  6.6.2 Response to a Plane Wave Signal Plus Noise.  
  6.6.3 Transient Effects.  
6.7 Summary of Chapter 6.  

CHAPTER 7. Test Results  
7.1 Results of Tests on the Digital Processor.  
  7.1.1 Commentary.  
7.2 Results of Tests on the Analogue Processor.  
  7.2.1 Commentary.  

CHAPTER 8. Conclusions  

REFERENCES  

APPENDIX A. The DFT Matrix Factorisation  

APPENDIX B. On the Reduction of Computing Effort by using the FFT  

APPENDIX C. The Derivation of Expressions for the Number of Arithmetic Operations in the FFT
APPENDIX D. Copy of paper presented to British Acoustical Society. February 1970

APPENDIX E. The Mean Square Value of an Ideal Reconstruction.

APPENDIX F. A Polynomial Representation of the Ideal Reconstruction Filter Output

APPENDIX G. The Calculation of the Digital Processor Error

APPENDIX H. Component Tolerance Calculations on an Analogue DFT Network

Page
170
178
182
184
189
LIST OF PRINCIPAL SYMBOLS

( ) $\leftrightarrow$ ( ) the Fourier transform operation.

( ) $\ast$ ( ) the convolution operation.

( ) conjugate or time average, as defined in text.

$\overline{()}$ ensemble average.

$\langle()\rangle$ ensemble average.

e(t) an error waveform.

e_r a sample of e(t).

$e^*(t)$ uniformly sampled e(t).

$f_c$ filter cut-off frequency.

$f_o$ carrier frequency.

$f_T(t) \equiv Q_1^*(s)$ equivalent time waveform.

$F_T(f) \leftrightarrow f_T(t)$ filter impulse response.

$H(f) \leftrightarrow h(t)$ filter transfer function.

j $\sqrt{-1}$, or an integer.

$\ell$ distance measured in wavelengths, or an integer.

L sampling interval in aperture.

n number of elements in array.

N matrix representation of discrete Fourier transform.

$N_i$ mean square value of input noise.

$p(t,\ell,\phi)$ pressure at a point in the aperture.

$q(\ell)$ complex model of field on the aperture line.

$q_1(\ell)$ the observed field (limited by the aperture).

$q_1^*(\ell)$ uniformly sampled $q_1(\ell)$, origin at end of aperture.

$Q_1(s) \leftrightarrow q_1(\ell)$ far field source distribution function.
\( r_1(\mathbf{\tau}) \)  
\( R_1(s) \leftrightarrow r_1(\mathbf{\tau}) \)  
\( s \)  
\( s(t) \equiv R_1(s) \)  
\( s^*(t) \)  
\( s_1(t) \)  
\( S_1(f) \leftrightarrow s^*_1(t) \)  
\( S_o \)  
\( S_i \)  
\( t \)  
\( T \)  
\( u^*(t) = s^*(t) + e^*(t) = \text{input sequence to interpolating filter} \)  
\( v(t) \)  
\( x \)  
\( (\ )_x \)  
\( y \)  
\( (\ )_y \)  
\( \beta \)  
\( \gamma \)  
\( \delta \)  
\( \Lambda \)  
\( \varepsilon_1 \)  
\( \varepsilon_2 \)  
\( \theta \)  
\( \kappa \)  
\( \lambda \)  
\( \nu_k \)  

sampled field, origin at centre of aperture.

\( \sin \theta \), or an integer.

signal component in output of filter.

uniformly sampled \( s(t) \).

one cycle of \( s(t) \) which is periodic.

mean square value of output signal.

mean square value of input signal.

time, or an integer.

period of \( f_\tau(t) \).

input sequence to interpolating filter.

output of interpolating filter.

an input variable, or distance.

the real part of a complex variable.

an output variable.

the imaginary part of a complex variable.

number of bits in a binary number.

variable associated with error in digital processor.

delta function.

variable associated with error in analogue processor.

error measure before detector.

error measure after detector.

angle of incidence.

quantization step size.

wavelength.

complex noise model in channel \( k \).
\( \sigma^2 \) variance.

\( \tau \) sampling interval in time domain.

\( \phi \) phase angle.

\( \psi \) phase angle.

\( \omega = 2\pi f \)
CHAPTER 1

Introduction

Since the trials of the first equipment, reported in 1959\(^1\), within-pulse electronic sector scanning sonar has become established as a useful tool for locating fish shoals and observing their movements. The underlying principle is straightforward. A broad sector of water, extending perhaps ±30° about a centre line, is insonified by a pulse of acoustic energy, and objects within this sector will cause some energy to be reflected back to a fixed array of receiving transducers. The array outputs are processed in such a way as to effectively produce a narrow receiving beam which is swept electronically across the insonified sector within the duration of the pulse. Repeating this process throughout the working range of the sonar system enables a picture of the complete sector to be built up before the next pulse is transmitted. The information is usually presented on a B-scope display of range against bearing. This system has obvious operational advantages over a mechanically scanned system of similar resolution capability which can only observe a small part of the sector in the interval between transmitted pulses.

In spite of the advantages, electronic sector scanning sonar has not been adopted by commercial fishing fleets because of the expense involved\(^2\). A system is required which is inexpensive, reliable, and easy to operate and maintain.

One promising solution to the problem was suggested by Rairn\(^3\) in which digital techniques were used to measure the time difference between the zero-crossings of signals appearing at the outputs of adjacent
receiving transducers, and thence determine the presence of a target and its location. This method is attractive because of the low cost of microelectronic digital circuits, and coupling this with medium and large scale integrated circuit techniques should result in an inexpensive and reliable system. However, the performance is not as good as that of the earlier electronic scanning receiver, and cannot be so, because information about the zero crossings alone is not sufficient to describe the field at the receiving aperture when more than one target echo is present.

Recent advances in semiconductor technology have produced low cost devices, such as high gain amplifiers and analogue multiplexers, which have made possible an alternative means of producing a sector scan display. This thesis describes a method which can be implemented by analogue techniques throughout, or digital techniques in part.

The signal processor can be viewed as producing a number of narrow receiving beams whose positions are fixed in space relative to the receiving array, and which together cover the entire sector of interest. The bearing display at any range is generated by interpolating between the beam outputs, and should produce results similar to that obtained by the earlier electronic scanning sonar described in reference 1.

The analogue processor takes the form of an interconnected array of only two basic building blocks, both of which can be mass produced using thin or thick film hybrid circuit technology. The digital processor is more expensive at the present time, but this situation may well be reversed in the wake of large scale integrated circuit developments.
Figure 2.1
Pressure variation along the aperture
2.1 The Field Model

The problem to be investigated in this chapter may be stated as follows. Given a spatial distribution of sources of energy in a homogeneous, isotropic medium, how may the angular location of the sources be determined from measurements taken over a finite part of the resulting field? The variation of the field amplitude at any point is assumed to be sinusoidal with a constant, known, frequency.

Attention will be restricted to a linear aperture, with a single source a sufficient distance away from it for the incident wavefront to be considered plane. This situation, at a particular instant in time, is shown in Figure 2.1(a). The lines perpendicular to the direction of propagation, and separated by the wavelength, \( \lambda \), join points where the instantaneous incremental pressure is a maximum. At any point on the aperture line, the variation of pressure with time is sinusoidal as the wave propagates through the medium. If the direction of propagation is not normal to the aperture, the pressure variations measured at two points on the aperture line are displaced in phase by an amount proportional to the separation of the points. Figure 2.1(a) shows the conditions prevailing at time \( t = 0 \), with the direction of propagation making an angle \( \theta \) to the normal. The pressure variation, which is superimposed on the static pressure, at any point on the aperture line is expressed in equation 2.1

\[
p = \hat{p} \cos(\omega t + 2\pi \frac{\sin \theta \cdot x}{\lambda} + \phi)
\]  

- (2.1)
\( \hat{p} \cos \phi \) is the pressure at the point \( x = 0 \) at time \( t = 0 \). Equation 2.1 defines a wave travelling along the aperture line in the negative \( x \) direction with a wavelength \( \frac{\lambda}{\sin \theta} \), and is shown in Figure 2.1(b).

If the direction of the source had been on the other side of the normal, (i.e. \( \theta \) negative), equation 2.1 would describe a wave travelling along the aperture line in the positive \( x \) direction with the same wavelength, \( \frac{\lambda}{|\sin \theta|} \). So, the required information on the direction of the single source is contained in the wavelength of the sinusoidal pressure distribution along the aperture, and also in the direction in which the wave travels along the aperture. The two pieces of information may be combined in a single variable, called the spatial frequency, which may take on positive or negative values. The spatial frequency, \( f_s \), is defined as the number of cycles per unit distance along the aperture, as a direct analogy to the frequency of a sinusoidal time waveform.

\[
f_s = \frac{1}{\lambda / \sin \theta} = \frac{\sin \theta}{\lambda} \tag{2.2}
\]

It is convenient to normalise this quantity by measuring distances along the aperture in terms of the wavelength, \( \lambda \), in the medium, and then equation 2.2 reduces to

\[
s = \sin \theta \tag{2.3}
\]

where the normalised spatial frequency, \( s \), is independent of frequency.

Equation 2.1 may now be written as

\[
p = \hat{p} \cos (\omega t + 2\pi s l + \phi) \tag{2.4}
\]
Figure 2.2

Field model at a particular instant in time
where \( z = \frac{x}{\lambda} \), i.e. distance measured in wavelengths.

A complex variable \( q \) can be defined now,

\[
q = \tilde{p} \exp(j\omega t) \exp\{j(2\pi s + \phi)\}
\]

where \( \tilde{p} = \text{Re}\{q\} \)

This step is necessary because if the dependence of \( p \) on time is removed by observing the aperture at one particular instant, as in Figure 2.1(b), it would be impossible to decide whether the source was in a direction of +0 or -0. A processor will be described which samples the aperture at one instant, and the drawback just mentioned is overcome by resorting to the complex model expressed in equation 2.5. An attempt at a pictorial representation of the complex field as a function of distance along the aperture is shown in Figure 2.2. The complex value of \( q \) at any point, such as \( A \), in the aperture is represented by a phasor. The tips of the phasors for all points in the aperture are then joined to produce a helix, which has an anticlockwise progression along the \( z \) axis for positive \( s \), and a clockwise progression for negative \( s \).

Using a suitable demodulation technique, it is possible to obtain the real and imaginary components of the complex sample, and it is evident that it should be possible to distinguish between the distributions shown in Figures 2.2(a) and 2.2(b), even though they have the same real component, shown in Figure 2.2(c).

2.2 Fourier Transform Relations for a Continuous Aperture

It was shown in section 2.1 that the spatial frequency of the field
at the aperture is determined by the direction of the source. It follows that the angular location of the source may be found by evaluating the Fourier transform $Q(s)$, of the complex field, $q(\ell)$.

Adopting the convention used in (4) to show that two functions are Fourier transform pairs

$$Q(s) \leftrightarrow q(\ell)$$  \hspace{1cm} (2.6)

where

$$Q(s) = \int_{-\infty}^{\infty} q(\ell) \exp(-j2\pi s\ell) d\ell$$  \hspace{1cm} (2.7)

and

$$q(\ell) = \int_{-\infty}^{\infty} Q(s) \exp(j2\pi s\ell) ds$$  \hspace{1cm} (2.8)

For example, if the length of the aperture is infinite, and the direction of the source is given by $s = s_1$, then from equations 2.5 and 2.7

$$Q(s) = \exp\{j(\omega t + \phi)\} \int_{-\infty}^{\infty} \delta(\ell - s_1) \exp(-j2\pi s_1 \ell) d\ell$$  \hspace{1cm} (2.9)

Now, $\exp(j2\pi s_1 \ell) \leftrightarrow \delta(s - s_1)$  \hspace{1cm} (2.10)

where $\delta(s)$ is a delta function of unit strength. Therefore, equation 2.9 becomes

$$Q(s) = \exp\{j(\omega t + \phi)\} \cdot \delta(s - s_1)$$  \hspace{1cm} (2.11)

The exponential term in equation 2.11 represents the time varying phase of the Fourier transform which would be expected as the wave travels along the aperture. Henceforth, this term will be dropped, since it does not affect the structure of $Q(s)$. 
Figure 2.3

$|Q(s)|$ v. $s$, direction of propagation corresponding to $s = s_1$, aperture of infinite length.

Figure 2.4

(a) Aperture function $t(L)$

(b) $T(s) \leftrightarrow t(L)$

Figure 2.5

$|Q_1(s)|$ v. $s$, direction of propagation corresponding to $s = s_1$, aperture of length $d$.
So \[ Q(s) = \hat{p} \cdot \delta(s - s_1) \] \hspace{1cm} -(2.12)

and a source in a direction given by \( \sin \theta = s_1 \) is represented by a delta function of strength \( \hat{p} \) at the point \( s_1 \) on the \( s \) axis. Figure 2.3 shows \( |Q(s)| \) plotted against \( s \).

In practice the aperture has a finite length, \( d \), and the observed field, \( q_1(\ell) \), can be expressed in the form

\[ q_1(\ell) = q(\ell) \cdot t(\ell) \] \hspace{1cm} -(2.13)

where \( t(\ell) = 1, \ 0 \leq \ell < d \)
otherwise \( t(\ell) = 0 \)

If \( T(s) \leftrightarrow t(\ell) \), and \( Q_1(s) \leftrightarrow q_1(\ell) \)

then \( Q_1(s) = Q(s) \ast T(s) \) \hspace{1cm} -(2.14)

The symbol '\( \ast \)' indicates the convolution of the functions on either side of it.

Here, \( T(s) = d \cdot \sin \frac{\pi sd}{\pi sd} \cdot \exp(-j\pi sd) \) \hspace{1cm} -(2.15)

Figure 2.4 shows \( t(\ell) \) and the modulus of its transform, and Figure 2.5 shows \( |Q_1(s)| \) plotted against \( s \). \( Q_1(s) \) is unique, and the transform corresponding to a source at \( s = s_1 \) is different from the transform for a source at \( s = s_2 \), no matter how small the difference between \( s_1 \) and \( s_2 \). In the absence of noise, in principle, it should be possible to detect the difference and determine the precise location of the source.

The field due to two or more sources is assumed to be the superposition
of the fields due to each source taken separately. The Fourier transform is a linear process, and so the transform of the total field is the superposition of the transforms of the fields due to each source. Once again, in the absence of noise, it should be possible to resolve two adjacent sources, radiating at the same frequency, no matter how close together they may be.

The transform, $T(s)$, is characteristic of the aperture. The width of the main lobe is inversely proportional to the number, $N$, of wavelengths, $\lambda$, contained in the aperture. For an aperture of fixed length, the width of the main lobe, and the level of the sidelobes, can be varied by 'shading' the aperture. A simple detection scheme involving inspection of the modulus of $Q_j(s)$ would not reliably resolve two sources separated by less than half the width of the main lobe between nulls, known as the Rayleigh distance, in approximate terms. The Rayleigh distance decreases as the length of the aperture is increased, and, borrowing a term from Welsby, the 'fineness of detail' to be observed by an aperture using this simple detection scheme sets a lower limit on the length of the aperture.

2.3 The Sampled Aperture

If the field at the aperture is represented by a set of point samples taken at a fixed interval, the Fourier transform of the sampled field can be made to approximate, as closely as desired, the Fourier transform of the continuous field at the aperture.

It can be shown that a sampled function can be represented by the original function multiplied by a uniform train of delta functions spaced apart by the sampling interval.
The sampling function $s_L(\ell)$

(b)

$S_L(s) \leftrightarrow s_L(\ell)$

Figure 2.5

$|Q^*_1(s)|$ v. s, number of samples in aperture = 8

Figure 2.7
Figure 2.6(a) shows an infinite train of delta functions, \( s_L(\ell) \), of unit strength, with a spacing \( L \). \( s_L(\ell) \) will be called the sampling function.

\[ s_L(\ell) = \sum_{k=-\infty}^{\infty} \delta(\ell - kL) \quad - (2.16) \]

\[ S_L(s) \leftrightarrow s_L(\ell) \quad - (2.17) \]

where

\[ S_L(s) = \frac{1}{L} \sum_{k=-\infty}^{\infty} \delta(s - kS) \quad - (2.18) \]

and \( S = \frac{1}{L} \).

i.e. the Fourier transform of the sampling function, shown in Figure 2.6(b), is an infinite train of delta functions, of strength \( \frac{1}{L} \) and spacing \( \frac{1}{L} \).

The effect of sampling the observed field, \( q_1(\ell) \), can be deduced now.

Let \( q_1^*(\ell) = q_1(\ell) \cdot s_L(\ell) \)

then

\[ q_1^*(\ell) \leftrightarrow Q_1^*(s) \quad - (2.20) \]

where

\[ Q_1^*(s) = Q_1(s) \ast S_L(s) \quad - (2.21) \]

\( |Q_1^*(s)| \) is plotted against \( s \), for \( L = 1 \) and \( d = 3 \), in Figure 2.7.

\( Q_1^*(s) \) is periodic, being the superimposition of \( Q_1(s) \) convolved with every delta function in \( S_L(s) \). It can be seen that as the sampling interval, \( L \), is reduced, the period of \( Q_1^*(s) \) increases and the influence between the individual terms in the summation representing \( Q_1^*(s) \), (combine equations 2.18 and 2.21), is reduced. In the limit, as \( L \rightarrow 0 \), \( Q_1^*(s) \rightarrow Q_1(s) \) apart from the multiplying factor, \( \frac{1}{L} \), contained in
equation 2.18. Papoulis\(^{4}\) overcomes this by saying that the sampling function should have strength \( L \). Using the detection scheme, mentioned before, which inspects \( |Q_1(s)| \), ambiguity will arise if the Fourier transform of the sampled field repeats within the observed range of angles. Provided that the sampling interval is chosen so that this does not happen, \( Q_1(s) \) can be used instead of \( Q_1^*(s) \) to determine the direction of a source. For instance, if \(-90^\circ < 0 < 90^\circ\), then \(-1 < s < 1\), and the minimum value of the period, \( \frac{1}{L} \), of \( Q_1(s) \) is 2. This means that the maximum allowable sample spacing is 0.5 (wavelengths). If it is known that the source direction is restricted to \(-30^\circ < 0 < 30^\circ\), or \(-0.5 < s < 0.5\), then similar reasoning shows that the maximum sample spacing may be increased to 1 (wavelengths).

A practical field sampling probe will be a finite size and will exhibit directional properties. For example, the pressure amplitude at the sample points in the aperture may be sensed by a plate of piezoelectric material whose width could be anything up to the sampling interval, \( L \). The effect of this can be represented in the aperture by the convolution of each delta function, corresponding to the samples, with an element function, \( e(\lambda) \).

\[
e(\lambda) = 1, \quad |\lambda| \leq \frac{a}{2}
\]

- (2.22)

where \( a \) = width of the sampling element, and \( e(\lambda) = 0 \) elsewhere.

\[
E(s) \leftrightarrow e(\lambda)
\]

- (2.23)

Here, \( E(s) = a \cdot \frac{\sin \frac{\pi sa}{\pi a}}{\frac{\pi a}{\pi a}} \)

- (2.24)
Figure 2.8
The element function and its transform

Figure 2.9
$|E(s)Q_1^*(s)|$ v. $s$, number of elements = 8, width of element = $L$
Now, \( e(\ell) \ast q_1^*(\ell) \leftrightarrow E(s) \cdot Q_1^*(s) \) \hspace{1cm} (2.25)

Figure 2.8 shows the element function and its transform when \( a = L \), and Figure 2.9 shows the result of the Fourier transform operation on samples of the field taken at intervals of \( \lambda \) using a flat probe of width \( \lambda \). It is possible to correct the tapering effect on the transform \( Q_1^*(s) \), if desired, by multiplying the transform in equation 2.25 by \( \frac{1}{E(s)} \) in the range of interest. This will be done in the processors to be described in later chapters, so the development will continue by assuming that point samples of the field are available.

2.4 Computation of the Fourier Transform

Combining equations 2.19 and 2.16, the sampled field, \( q_1^*(\ell) \), within the aperture may be expressed as

\[
q_1^*(\ell) = q_1(\ell) \sum_{k=-\infty}^{\infty} \delta(\ell - kL)
\]

The number of samples, \( n \), with spacing, \( L \), in the aperture of length, \( d \), is \( \frac{d}{L} \).

Therefore, equation 2.26 becomes

\[
q_1^*(\ell) = q_1(\ell) \sum_{k=0}^{n-1} \delta(\ell - kL)
\]
or

\[
q_1^*(\ell) = \sum_{k=0}^{n-1} q_1(kL) \delta(\ell - kL)
\]

\[
\therefore Q_1^*(s) = \int_{-\infty}^{\infty} \exp(-j2\pi s \ell) \sum_{k=0}^{n-1} q_1(kL) \delta(\ell - kL) \, d\ell.
\]

\hspace{1cm} (2.27)

\hspace{1cm} (2.28)
\[ Q_i^*(s) = \sum_{k=0}^{n-1} q_i(kL) \exp(-j2\pi skL) \quad - (2.29) \]

i.e. \[ Q_i^*(s) = q_i(0) + q_i(L) \exp(-j2\pi sL) + q_i(2L)\exp(-j2\pi s2L) + \ldots \]
\[ + q_i[L(n - 1)] \exp(-j2\pi s(n - 1)L) \quad - (2.30) \]

It is convenient to generate \( Q_i^*(s) \) as a time waveform for display purposes, and this can be done by the transformation

\[ s = s_1 + bt \quad - (2.31) \]

where \( s_1 \) is the direction at the start of the scan when \( t = 0 \), and \( b \) is a constant scale factor which determines the scan rate.

The \((k + 1)\)th term in equation 2.30 now becomes

\[ q_i(kL) \exp(-j2\pi s_1kL) \exp(-j2\pi kbLt). \]

The exponential part, corresponding to a phase shift, is composed of a constant term determined by the initial direction, and a time varying term whose value is proportional to the distance of the sample point from the origin.

So, the process of generating the Fourier transform of the sampled field as a time waveform is achieved by inserting a time varying progressive phase shift at each sample point and summing the outputs. The schematic for this operation is shown in Figure 2.10, with \( \alpha(t) = -2\pi L(s_1 + bt) \).

This technique is well known, although the method of producing a time varying phase shift and summation varies \(^8, 9\).
transducer array

\[ \sum (n-1) \alpha(t) \]

Figure 2.10

Electronic scan schematic

\[ |Q_1^*(s)| \]

\[ 0, \frac{1}{nL}, \frac{2}{nL}, \ldots, \frac{(n-1)}{nL}, \frac{1}{L} \]

Figure 2.11

Samples of \( Q_1^*(s) \) generated by the DFT
2.5 The Discrete Fourier Transform

An alternative method of computing the Fourier transform of the sampled field is to compute, first, the discrete Fourier transform (DFT) of the samples. The DFT of a sequence of \( n \) samples is itself a sequence of \( n \) samples defined by,

\[
y_r = \sum_{k=0}^{n-1} x_k \exp(-j2\pi rk/n)
\]

\((r = 0, 1, 2, \ldots, n-1)\)

where \( y_r \) is the \( r \)th coefficient of the DFT, and \( x_k \) is the \( k \)th sample of the input sequence.

The inverse DFT is defined by

\[
x_k = \frac{1}{n} \sum_{r=0}^{n-1} y_r \exp(j2\pi rk/n)
\]

Applying the DFT to the field samples,

\[
q_1(kL) \equiv x_k
\]

Comparing equations 2.32 and 2.29,

\[
Q_1^n(rL) = y_r
\]

i.e. the DFT generates samples of the transform of the sampled field, the \( n \) samples being taken uniformly through the interval \( 0 \leq s < \frac{1}{L} \). This situation is shown in Figure 2.11.
$Q_1^*(s)$ is a periodic function, with period $\frac{1}{L}$, and so

$$Q_1^*(\frac{r}{nL}) = Q_1^*(\frac{r}{nL} - \frac{1}{L})$$

(2.36)

In the situation where the sector of interest is defined by $\frac{1}{2L} < s < \frac{1}{2L}$, then the coefficients of the DFT are interpreted as

$$y_r = Q_1^*(\frac{r}{nL}), \quad 0 \leq r \leq \frac{n}{2}$$

(2.37)

$$y_r = Q_1^*(\frac{r}{nL} - \frac{1}{L}), \quad \frac{n}{2} < r < n-1$$

$Q_1^*(s)$ is the transform of a function which is zero outside the range $0 < \xi < nL$. Accordingly, $Q_1^*(s)$ may be completely represented by samples taken uniformly throughout $s$ with a minimum spacing of $\frac{1}{nL}$. This is analogous to the sampling theorem in the frequency domain on the representation of the spectrum of a time limited waveform (4). Also, $Q_1^*(s)$ is a periodic function with a period $\frac{1}{L}$, so only $n$ samples uniformly spaced by $\frac{1}{nL}$ are required to completely represent the function. Therefore, in principle, the continuous function $Q_1^*(s)$ can be recovered from the DFT of the field samples by a suitable interpolation procedure.

The remaining chapters in this thesis explore the practical implementation of the last statement. Before this is done, the DFT will be shown to be equivalent to the formation of multiple beams.

2.6 The DFT and Multiple Beam Formation

The schematic diagram for a conventional summing array of $n$ elements is shown in Figure 2.10, with the phase shifts for each array element remaining constant with time.
Let the phase increment, $\alpha$, between adjacent elements be determined by

$$\alpha = -2\pi s_1 L$$

where $s_1$ defines a particular direction.

The output, $O(s)$, of the summing device is

$$O(s) = \sum_{k=0}^{n-1} q_1(kL) \exp(-j2\pi s_1 kL)$$

In response to a field of unit amplitude due to a source in a direction $s$, equation 2.39 becomes

$$O(s) = \sum_{k=0}^{n-1} \exp(j2\pi s kL) \exp(-j2\pi s_1 kL)$$

i.e.

$$O(s) = \sin \frac{n\pi L(s - s_1)}{\sin \pi L(s - s_1)} \exp[j\pi L(n-1)(s - s_1)]$$

$|O(s)|$ is the beam pattern associated with a receiving array having a progressive phase shift, defined by equation 2.38, across the aperture. The main lobe is pointing in the direction corresponding to $s = s_1$. Comparing equations 2.39 and 2.29, the output of this receiving array is the value of the transform, $Q_1^*(s)$, at the sample point $s = s_1$.

It has already been shown that the DFT computes samples of $Q_1^*(s)$, so the DFT operation can be interpreted as forming $n$ beams whose positions are fixed in space, and whose main lobes point in directions corresponding to the values of $s_1 = \frac{r}{nL}$, at the sample points.
Multiple-beam forming networks producing orthogonal beams have been in use in radar systems for a number of years (11). Their use does not appear to have been associated directly with the DFT, and the beam outputs are treated independently rather than as a set which defines $Q_1(s)$. 
CHAPTER 3

Synthesis of DFT Networks

3.1 Introduction

In Chapter 2, it was shown that an estimate of the angular location of sources, in the far field of an aperture could be obtained from the DFT of the samples of the field in the aperture. The DFT could be regarded, equivalently, as forming a number of overlapping beams, each of which is pointing in a fixed direction in space.

In this chapter, a technique is developed for the rapid design of a class of network which performs the DFT of the field samples. These networks could be fabricated using analogue or digital devices, and both types will be considered in more detail later on.

3.2 Notation

Throughout this chapter, and Appendix A, much use will be made of matrix notation. A common convention\(^{(12)}\) for printing matrix operations is to use bold-faced type, with column vectors represented by lower-case letters, and rectangular matrices represented by upper-case letters. The elements of the matrices are often printed in fine lower-case italics. Since it is not possible to call upon this repertoire of type faces, the following convention will be adopted:– all matrices and vectors will be represented by upper-case letters, and their individual elements will be printed in normal lower-case type.
Figure 3.1
Summing array

Figure 3.2
Alternative configuration for a 4-element summing array
3.3 Matrix Representation of Beam-Forming Networks

In view of the matrix notation adopted, it has been found necessary to use the identities 2.34 and 2.35 for the field samples and their DFT coefficients in the development of this chapter.

Consider the summing network, shown in Figure 3.1, which produces output \( y_1 \) in response to the field samples \( x_k \), \( (k = 0, 1, \ldots, n-1) \). Here, \( \alpha = -2\pi/n \), and putting \( a_k = \exp(jk\alpha) \), the network is conveniently represented by the matrix \( A \), of dimension \( 1 \times n \).

\[
A = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-1} \end{bmatrix}
\tag{3.1}
\]

The output, \( y_1 \), is then given by

\[
y_1 = A \cdot X
\tag{3.2}
\]

where \( X \) is a column vector, of dimension \( n \times 1 \), containing the samples of the field. For example, if \( n = 4 \)

\[
y_1 = \begin{bmatrix} 1 & a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}
\tag{3.3}
\]

Equation 3.3 may be factorised, to give

\[
y_1 = \begin{bmatrix} 1 & a_1 \\ . & 1 & a_2 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix}
\tag{3.4}
\]
i.e. \( y_1 = A_1 \cdot A_2 \cdot x \) \hspace{1cm} -(3.5)

where \( A_1 = \begin{bmatrix} 1 & a_1 \end{bmatrix} \)

and \( A_2 = \begin{bmatrix} 1 & a_2 \\ 1 & a_2 \end{bmatrix} \)

The spaces in \( A_2 \) are zero elements which have been left out for clarity.

Equation 3.4 shows that the process of forming \( y_1 \) can take place in two stages. The four array samples are first combined in a network represented by \( A_2 \), and the two outputs of \( A_2 \) are combined in the network represented by \( A_1 \). The schematic diagram for this sequence is shown in Figure 3.2.

The usefulness of the matrix representation of such networks lies in the fact that the inter-connection of the different components is implied by the structure of the matrix. Each row corresponds to one output of the network, and the disposition of the elements in each row indicates which outputs from the preceding stage are to be combined. Figure 3.2 was obtained directly from equation 3.4 by using this technique.

3.4 The Fast Fourier Transform

Repeating equation 2.32 as equation 3.7, the DFT of a sequence of \( n \) samples is defined by

\[
y_r = \sum_{k=0}^{n-1} x_k \exp(-j2\pi rk/n), \quad r = 0, 1, \ldots, n-1 \hspace{1cm} -(3.7)
\]

Putting \( n = \exp(-j2\pi/n) \)

\[
\text{equation 3.7 may be rewritten as}
\]
Each coefficient is the sum of a sequence formed by multiplying each input sample by a unit rotation vector whose value depends on the number of the coefficient, \( r \), and the number of the sample, \( k \).

Equation 3.9 can be expressed in matrix form

\[
Y = N \cdot X
\]

where \( X \) is a column vector containing the \( n \) input samples, \( N \) is a square matrix of dimension \( n \times n \) describing the DFT operation, and \( Y \) is a column vector containing the \( n \) coefficients formed by the DFT. Labelling the rows and columns from 0 to \( n-1 \), the element in row \( r \) and column \( k \) in \( N \) is \( N^{rk} \).

Examination of equation 3.9 shows that there are \( n^2 \) rotation vectors (including \( N^0 \)) involved in forming all \( n \) coefficients. However, only \( n \) of these are distinct when written modulo \( 2\pi \), and they divide the unit circle into \( n \) equal segments.

If \( n \) is not a prime number, and can be expressed as the product

\[
n = \prod_{i=1}^{t} n_i
\]

then the transform can be decomposed into \( t \) steps with \( \frac{n}{n_i} \) transforms of dimension \( n_i \) within each step (13). As it will be shown later, this leads to a reduction in the total number of operations which are necessary to carry out the \( n \)-point DFT in comparison with the straightforward method of deriving each coefficient from equation 3.9.
The reduction in the number of rotation operations is of particular interest since this means fewer phase shift devices in analogue processors, and fewer complex multiplications, which are relatively time consuming when compared with a simple addition operation, in digital processors. The latter reason has given rise to the term 'Fast Fourier Transform' (FFT) to describe the result of the decomposition procedure, that is, an algorithm for the more efficient computation of the DFT.

Much has been written about the FFT, its derivation, and uses - see the bibliography attached to reference 10. Some writers have described the FFT in terms of a factorisation of the matrix, \( N \), defining the DFT operation \((14, 15)\). The matrix approach is particularly useful because, as it has already been shown, a beam-forming network is readily represented in matrix form.

The procedure for writing the DFT matrix, \( N \), as the product of \( t \) matrices,

\[
N = F_t F_{t-1} \ldots F_2 F_1
\]

where \( t \) is the number of factors in the number \( n \), is now well documented \((16, 17, 18)\).

If \( n = \sum_{i=1}^{t} n_i \), the \( i \)th factor matrix, \( F_i \), contains the elements of \( \frac{n}{n_i} \) transforms of dimension \( n_i \). Each matrix also contains rotation vectors which may be removed as a further factor matrix, \( R_i \), containing only rotation vectors on the main diagonal. This leads to

\[
N = N_t R_{t-1} N_{t-1} \ldots R_2 N_2 R_1 N_1
\]
where \( N_i \) now contains only transforms of dimension \( n_i \). So equation 3.13 shows that a network for performing the DFT can be synthesised from units which perform the DFT of dimension \( n_i \). If \( n \) is a power of 2, the basic unit is a 2-point transform which is achieved by a simple sum and difference operation.

The factorisation procedure described in Appendix A was developed as a design technique for synthesising multiple beam-forming networks from devices which perform a DFT of dimension, \( n_i \), which is a factor of the dimension, \( n \), of the complete network. As stated before, this technique has the desirable feature of reducing the number of rotation operations, in particular, compared with the direct implementation of equation 3.7.

Equations 3.12 and 3.13 are adequate for digital computer implementation of the FFT(18). However, it has been found useful to carry the factorisation procedure a stage further when dealing with small-scale processors using either analogue or digital devices. The rows and columns of the matrix, \( N_i \), in equation 3.13 are rearranged so that the matrix can be partitioned and written in quasi-diagonal(12) form. The purpose of this is to group the elements of the DFT's, of dimension \( n_i \), in compact submatrices so that the schematic diagram for the whole network can be drawn directly. This process will be illustrated later for a network with \( n = 8 \). The design technique was described by the author at a meeting of the British Acoustical Society in February 1970(19), and Appendix D is a copy of the paper which was presented.

The form of the resulting network for \( n \) a power of 2 is well known to radar antenna designers. In 1961, Butler and Lowe(20) described a network for forming multiple beams using phase shifters and 90° hybrid couplers which have 2 input ports and 2 output ports. Allen(21)
describes a similar network using sum and difference hybrids, and Shelton and Kelleher\(^{(11)}\) introduce more complicated couplers having more than 2 input and output ports for use in networks where the number of inputs is not a power of 2. All such networks have become known generically as Butler matrices, but no generalised synthesis procedure linking the different designs appears to have been published. Moody\(^{(22)}\) describes a systematic approach to the particular form which uses the 90° hybrid coupler with \(n\) a power of 2. However, his procedure amounts to following a set of rules which are not derived and which contribute little to an understanding of the problem.

The FFT factorisation of the DFT matrix provides the essential link between all forms of Butler matrix, and it leads to a design procedure which is flexible with regard to the number of elements in the array, and also the particular type of coupler used. The similarity between the FFT schematic and Butler matrices has already been noted\(^{(23)}\), and, with hindsight, it is perhaps surprising that the Butler matrix designs did not lead to the FFT a good deal sooner than 1965\(^{(13)}\).

Hitherto, published work on Butler matrices has been concerned with their use at radar frequencies, but the networks may also be used to advantage at sonar frequencies. The low frequency analogue equivalents of the microwave hybrid couplers can be fabricated using microelectronic circuit techniques\(^{(19)}\), and perhaps transformers could be used at the higher frequencies.

3.5 Network Synthesis

The only requirement for this procedure to work is that \(n\) is not prime, and it should be expressible, initially, as the product of just
2 factors,

\[ n = p \cdot q \]  

It is shown in Appendix A that the DFT matrix, \( N \), can then be expressed as

\[ N = C \cdot Q_Q \cdot C^T \cdot R \cdot P_p \cdot C \]  

where all the factors are matrices of dimension \( n \times n \).

\( P_p \) has been partitioned into submatrices of dimension \( p \times p \). The matrix is quasi-diagonal with the \( q \) identical submatrices, \( P \), on the main diagonal containing a DFT of dimension \( p \), and all other elements are zero.

\( Q_Q \) has been partitioned into submatrices of dimension \( q \times q \). The matrix is quasi-diagonal with the \( p \) identical submatrices, \( Q \), on the main diagonal containing a DFT of dimension \( q \), and all other elements are zero.

\( C \) is a reordering matrix which has only one element in each row and column of value 1, specified by

\[ c_{m+sp,s+mq} = 1 \]  

for \( m = 0, 1, 2, \ldots, p-1 \)

\[ s = 0, 1, 2, \ldots, q-1 \]

and all other elements are zero. \( C^T \) is the transpose of \( C \). \( R \) has been partitioned into submatrices of dimension \( p \times p \). The matrix is
quasidiagonal, and the $s$th submatrix on the diagonal is $D_s$ which contains $d_{\lambda,m}$ defined by

$$d_{\lambda,m} = \delta_{\lambda,m} \cdot \eta_{s}$$  \hspace{1cm} (3.17)

where $\delta_{\lambda,m} = 1$, $\lambda = m$

$= 0$, $\lambda \neq m$

The experimental work described in this thesis uses a receiving array of 8 elements. The design of the DFT network for this array will be used as an example to illustrate the general technique.

3.5.1 The DFT Network of Dimension 8

$n = 8 = 4.2$, initially.

From equation 3.14, $p = 4$, and $q = 2$.

The first step is to determine the reordering matrix, $C$. From equation 3.16

$$c_{m+4s,s+2m} = 1$$  \hspace{1cm} (3.18)

for $m = 0, 1, 2, 3$

$s = 0, 1$

and all other elements are zero.

The elements defined by equation 3.18 may be found systematically by constructing a table which includes all the possible combinations of $s$ and $m$. 
Table 3.1

Locations of unit coefficients in reordering matrix, C

From Table 3.1, \( c_{0,0} = 1; \ c_{1,2} = 1; \ c_{2,4} = 1; \) etc. Hence

\[
C = \begin{bmatrix}
1 & \cdots & \cdots & \cdots & \cdots \\
\cdots & 1 & \cdots & \cdots & \cdots \\
\cdots & \cdots & 1 & \cdots & \cdots \\
\cdots & \cdots & \cdots & 1 & \cdots \\
\cdots & \cdots & \cdots & \cdots & 1
\end{bmatrix}
- (3.19)
\]

Now, the rotation vectors in matrix \( R \) are obtained from equation 3.17. There are 2 submatrices of dimension \( 4 \times 4 \) on the main diagonal of the partitioned matrix \( R \). The elements in the first submatrix, \( D_0 \), are given by

\[
d_{\lambda,m} = \delta_{\lambda m} \\
\lambda = 0, 1, 2, 3 \quad m = 0, 1, 2, 3
- (3.20)
\]
The elements in the second submatrix, $D_1$, are given by

$$d_{\lambda, m} = \delta_{\lambda m} \cdot n^\lambda \quad \lambda = 0, 1, 2, 3 \quad m = 0, 1, 2, 3$$  \hspace{1cm} (3.21)

where $n = \exp(-j\frac{\pi}{4})$.

Combining equations 3.20 and 3.21 gives

$$R = \begin{bmatrix}
1 & \cdots & \cdots & \cdots \\
\vdots & \ddots & \ddots & \ddots \\
\vdots & \vdots & \ddots & \ddots \\
\vdots & \vdots & \cdots & 1 \\
\cdots & \cdots & \cdots & n \\
\cdots & \cdots & \cdots & n^2 \\
\cdots & \cdots & \cdots & n^3
\end{bmatrix}$$  \hspace{1cm} (3.22)

In equation 3.15, matrices $P_P$ and $Q_Q$ are defined as follows:

$$P_P = \begin{bmatrix}
P & \cdots & \cdots & \cdots \\
\vdots & \ddots & \ddots & \ddots \\
\vdots & \vdots & \ddots & \ddots \\
\vdots & \vdots & \ddots & \ddots \\
\cdots & \cdots & \cdots & P
\end{bmatrix}$$  \hspace{1cm} (3.23)
Initial decomposition of a DFT of dimension 8

Decomposition of a DFT of dimension 4
where \( P \) defines a DFT of dimension 4.

\[
Q_Q = \begin{bmatrix}
  Q & & \\
  & \ddots & \\
  & & Q
\end{bmatrix}
\]

where \( Q \) defines a DFT of dimension 2. Combining equations 3.19, 3.22, 3.23 and 3.24, the product in equation 3.15 can be drawn schematically by using the technique described in section 3.3. The result is shown in Figure 3.3, which also shows the correspondence between the different parts of the network and the matrix factors.

The DFT of dimension 8 has been decomposed into 2 DFT's of dimension 4, plus 4 DFT's of dimension 2, plus 3 complex multiplications or phase shifts. Using the same technique, each DFT of dimension 4 can be decomposed into DFT's of dimension 2, and this is shown in Figure 3.4. If the network in Figure 3.4 is substituted for each DFT of dimension 4 in Figure 3.3, the final decomposition of the original DFT in terms of DFT's of dimension 2 is obtained. The schematic for this is shown in Figure 3.5, and the DFT of two samples is obtained by a simple sum and difference operation, since when \( n = 2 \)

\[
N = \begin{bmatrix}
  1 & 1 \\
  1 & -1
\end{bmatrix}
\]
Figure 3.5

Final decomposition of a DFT of dimension 8
This example has illustrated the general procedure to follow when n is expressible as the product of more than two prime factors, i.e. the repeated application of the technique used for just two factors.

3.6 Advantages of the FFT

It was stated in section 3.4 that the FFT offers a saving in computation over the direct method of generating the DFT of a sequence of samples. The nature of this saving will be investigated now in more detail, and since different criteria arise depending on whether the implementation is by digital or analogue means, the two situations will be dealt with separately.

3.6.1 Digital Implementation of the FFT

The detailed design of a special purpose digital processor is considered to be beyond the scope of this thesis. Consequently, the relative economic advantage of one particular hardware implementation compared with another cannot be assessed. So, a basic computer structure consisting of a data store, arithmetic unit, and control unit is assumed, and the savings to be considered arise solely from the reduction in the number of additions and complex multiplications involved in computing the DFT by the FFT method.

In the digital implementation of the FFT, a schematic diagram of the type shown in Figure 3.5 is interpreted as a form of flow diagram for the FFT algorithm.

Firstly, it is shown in Appendix B that if \( n = p.q \), then the FFT decomposition, always leads to a reduction in,
(i) the number of additions, taking two quantities at a time,
(ii) the number of complex multiplications,
for the practical case of both \( p \) and \( q \) being integers greater than 1.

In Appendix C, expressions are derived for the total number of
additions and complex multiplications involved in the FFT computation,
for the general case of \( n \) being the product of any number of factors.

If there are \( t \) factors in \( n \), so that,

\[
u = \prod_{i=1}^{t} n_i
\]  \hspace{1cm} -(3.25)

and the total number of additions (including subtractions) in the FFT is
\( m^- \), then

\[
m^- = n\left\{-t + \sum_{i=1}^{t} n_i\right\}
\]  \hspace{1cm} -(3.26)

If \( \ell \) of the \( t \) factors are equal to 2, so that

\[
u = 2^\ell \cdot \prod_{i=\ell+1}^{t} n_i
\]  \hspace{1cm} -(3.27)

and the total number of complex multiplications in the FFT is \( m^- \), then

\[
m^- = n\left\{3\ell - (t + 1) + \sum_{i=\ell+1}^{t} n_i\right\} + 1
\]  \hspace{1cm} -(3.28)

When \( n \) contains a power of 2 as a factor, the specific form in
equation 3.27 has to be used in the derivation of an expression for \( m^- \),
otherwise an error will result, as explained in Appendix C. However,
Gain of FFT processing - summations

Gain of FFT processing - multiplications
equation 3.26 for $m_a^*$ is still applicable in this case.

Equations 3.26 and 3.28 have been used to construct Table 3.2. For convenience, the straightforward method of computing the DFT, according to equation 3.7, has been called the slow Fourier transform (SFT). Table 3.2 lists the number of additions (including subtractions) and the number of complex multiplications involved in computing the SFT and the FFT, for any number of inputs between 2 and 32.

In the column listing the number of multiplications in the SFT, the reason why a particular number of inputs (e.g. 24) may have fewer multiplications than a SFT of lower dimension (e.g. 23) should be explained. If the number $n$ is prime, then the number of complex multiplications in the DFT matrix is $(n-1)^2$, since the elements in the first row and first column are all equal to 1. If $n$ contains 2 as a factor, the number of complex multiplications is less than $(n-1)^2$ because of the presence of some elements equal to $+1$ or $-1$ in the DFT matrix. This occurs in row $r$ and column $k$ when $\frac{2rk}{n}$ is an integer, say $i$; then $n^{rk} = \exp(-jim)$, which gives rise to the apparent anomaly just described.

The gain factors, $\nu_a$ and $\nu_m$, in the last two columns of Table 3.2 show the gains which are due to the reduction in computation required by the FFT. For instance, if the FFT has $\frac{1}{5}$ of the multiplications in the SFT, the gain factor, $\nu_m$, is 5. $\nu_a$ and $\nu_m$ are presented graphically, as a function of $n$ in Figures 3.6 and 3.7. Both $\nu_a$ and $\nu_m$ contribute to a reduction in the time taken to compute the DFT, but the most noticeable feature is the large gain obtainable from the multiplications when $n$ is a power of 2, (Figure 3.7). This is significant because complex multiplications consume much more processing time than complex additions or subtractions.
<table>
<thead>
<tr>
<th>n</th>
<th>SFT</th>
<th>FFT</th>
<th>Gain Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\pm \frac{m_a}{m})</td>
<td>(\pm \frac{m}{m})</td>
<td>(\pm \frac{m_a}{m})</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>7</td>
<td>42</td>
<td>36</td>
<td>42</td>
</tr>
<tr>
<td>8</td>
<td>56</td>
<td>32</td>
<td>24</td>
</tr>
<tr>
<td>9</td>
<td>72</td>
<td>64</td>
<td>36</td>
</tr>
<tr>
<td>10</td>
<td>90</td>
<td>64</td>
<td>50</td>
</tr>
<tr>
<td>11</td>
<td>110</td>
<td>100</td>
<td>110</td>
</tr>
<tr>
<td>12</td>
<td>132</td>
<td>84</td>
<td>48</td>
</tr>
<tr>
<td>13</td>
<td>156</td>
<td>144</td>
<td>156</td>
</tr>
<tr>
<td>14</td>
<td>182</td>
<td>144</td>
<td>98</td>
</tr>
<tr>
<td>15</td>
<td>210</td>
<td>196</td>
<td>90</td>
</tr>
<tr>
<td>16</td>
<td>240</td>
<td>176</td>
<td>64</td>
</tr>
<tr>
<td>17</td>
<td>272</td>
<td>256</td>
<td>272</td>
</tr>
<tr>
<td>18</td>
<td>306</td>
<td>240</td>
<td>90</td>
</tr>
<tr>
<td>19</td>
<td>342</td>
<td>324</td>
<td>342</td>
</tr>
<tr>
<td>20</td>
<td>380</td>
<td>292</td>
<td>120</td>
</tr>
<tr>
<td>21</td>
<td>420</td>
<td>400</td>
<td>168</td>
</tr>
<tr>
<td>22</td>
<td>462</td>
<td>400</td>
<td>242</td>
</tr>
<tr>
<td>23</td>
<td>506</td>
<td>484</td>
<td>506</td>
</tr>
<tr>
<td>24</td>
<td>552</td>
<td>416</td>
<td>120</td>
</tr>
<tr>
<td>25</td>
<td>600</td>
<td>576</td>
<td>200</td>
</tr>
<tr>
<td>26</td>
<td>650</td>
<td>576</td>
<td>338</td>
</tr>
<tr>
<td>27</td>
<td>702</td>
<td>676</td>
<td>162</td>
</tr>
<tr>
<td>28</td>
<td>756</td>
<td>628</td>
<td>224</td>
</tr>
<tr>
<td>29</td>
<td>812</td>
<td>784</td>
<td>812</td>
</tr>
<tr>
<td>30</td>
<td>870</td>
<td>720</td>
<td>210</td>
</tr>
<tr>
<td>31</td>
<td>930</td>
<td>900</td>
<td>930</td>
</tr>
<tr>
<td>32</td>
<td>992</td>
<td>832</td>
<td>160</td>
</tr>
</tbody>
</table>

Table 3.2
Operations involved in computing the SFT and FFT.
The reduction in the time taken to compute the DFT by using the FFT algorithm is estimated as follows. It is assumed that the processor has a single arithmetic unit, and that the time taken to perform a complex multiplication is \( b T_a \), where \( T_a \) is the time taken to perform a complex addition. \( b \) is a constant, determined by the processor, and it is assumed that the time taken to perform a complex subtraction is also \( T_a \). If \( T_S \) is the time taken to compute the SFT and \( T_F \) is the time taken to compute the FFT, then

\[
T_S = m_a T_a + m_m b T_a \quad \text{-- (3.29)}
\]

and

\[
T_F = m_a T_a + m_m b T_a \quad \text{-- (3.30)}
\]

The gain in processing time \( \nu_t \) is given by

\[
\nu_t = \frac{T_S}{T_F}
\]

or

\[
\nu_t = \frac{m_a + b m_m}{m_a + b m_m} \quad \text{-- (3.31)}
\]

Take, as an example, the CTL Modular One computer. Here, \( b = 5 \), and if \( n = 32 \)

\[
\nu_t = \frac{992 + 5 \times 832}{160 + 5 \times 49} = 13
\]

Further reductions in computing time \(^{(24)}\) can be achieved by noting that a multiplication by \( \exp(-j\pi/2) \) is equivalent to interchanging the real and imaginary parts of the complex quantity, together with a single change of sign. Also, multiplication by \( \exp(-j\pi/4) \) requires only two real multiplications rather than four in the general case of a complex multiplication.
It is possible to use more than one arithmetic unit in a special purpose FFT processor designed to give the ultimate in computing speed.

To summarise, the advantage of using the FFT algorithm in a digital processor is in the reduction in computing time which is obtained, particularly when the number of input samples is a power of 2. This opens up the possibility of real-time digital signal processing in situations where it could not have been entertained before.

3.6.2 Analogue Implementation of the FFT

Using analogue components, it is quite possible to implement the DFT directly, and so obtain a multiple beamforming network. This would consist of \( n \) structures, of the type shown in Figure 3.1, connected in parallel. The FFT decomposition leads to a schematic diagram, for example Figure 3.5, which gives an alternative way of connecting the basic elements for summing and phase shifting, so that \( n \) beams are formed.

Any advantage of the FFT structure over the SFT structure would be measured in terms of the relative cost for comparable performance. Table 3.3 shows the number of summing devices and phase shifts required for both configurations. The number of phase shifts corresponds to the number of complex multiplications used in the digital implementation. Fewer summing devices are required in the analogue networks because the restriction of summing only two quantities at a time is unnecessary, and it has been assumed that only one summing device is required for each row of a DFT matrix.

A comparison of the two analogue networks shows that the FFT structure requires fewer phase shifts but more summing devices than the SFT.
<table>
<thead>
<tr>
<th>n</th>
<th>[\text{SFT}] Sums</th>
<th>[\text{SFT}] Phase Shifts</th>
<th>[\text{FFT}] Sums</th>
<th>[\text{FFT}] Phase Shifts</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>16</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>16</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>36</td>
<td>7</td>
<td>36</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>32</td>
<td>24</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>64</td>
<td>18</td>
<td>28</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>64</td>
<td>20</td>
<td>36</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>100</td>
<td>11</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>84</td>
<td>36</td>
<td>25</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>144</td>
<td>13</td>
<td>144</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>144</td>
<td>28</td>
<td>78</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>196</td>
<td>30</td>
<td>76</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>176</td>
<td>64</td>
<td>17</td>
</tr>
<tr>
<td>17</td>
<td>17</td>
<td>256</td>
<td>17</td>
<td>256</td>
</tr>
<tr>
<td>18</td>
<td>18</td>
<td>240</td>
<td>54</td>
<td>64</td>
</tr>
<tr>
<td>19</td>
<td>19</td>
<td>324</td>
<td>19</td>
<td>324</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>292</td>
<td>60</td>
<td>81</td>
</tr>
<tr>
<td>21</td>
<td>21</td>
<td>400</td>
<td>42</td>
<td>148</td>
</tr>
<tr>
<td>22</td>
<td>22</td>
<td>400</td>
<td>44</td>
<td>210</td>
</tr>
<tr>
<td>23</td>
<td>23</td>
<td>484</td>
<td>23</td>
<td>484</td>
</tr>
<tr>
<td>24</td>
<td>24</td>
<td>416</td>
<td>96</td>
<td>61</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>576</td>
<td>50</td>
<td>176</td>
</tr>
<tr>
<td>26</td>
<td>26</td>
<td>576</td>
<td>52</td>
<td>300</td>
</tr>
<tr>
<td>27</td>
<td>27</td>
<td>676</td>
<td>81</td>
<td>136</td>
</tr>
<tr>
<td>28</td>
<td>28</td>
<td>628</td>
<td>84</td>
<td>169</td>
</tr>
<tr>
<td>29</td>
<td>29</td>
<td>784</td>
<td>29</td>
<td>784</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>720</td>
<td>90</td>
<td>166</td>
</tr>
<tr>
<td>31</td>
<td>31</td>
<td>900</td>
<td>31</td>
<td>900</td>
</tr>
<tr>
<td>32</td>
<td>32</td>
<td>832</td>
<td>160</td>
<td>49</td>
</tr>
</tbody>
</table>

**Table 3.3**

Operations involved in analogue networks
Phase shift circuits using a differential amplifier

(a) Summing amplifier

(b) Difference amplifier

Figure 3.8

Figure 3.9
structure. The cost of the two structures must depend on the techniques used to achieve summing and phase shifting. The type 709 microelectronic differential amplifier could form the basis of beam-forming networks at frequencies lower than 40 kHz. This amplifier is a low cost unit, currently less than 30p, which could be used in conjunction with thick film circuits to produce hybrid building blocks capable of performing the desired operations. The circuit configurations for phase retard and phase advance operations are shown in Figure 3.8 - the operation of these circuits will be analysed in detail at a later stage.

The economic case for the FFT structure with $n$ a power of 2 is the easiest one to demonstrate. The circuits for the basic sum and difference operations on two inputs are shown in Figure 3.9. It can be seen that all the circuits in Figures 3.6 and 3.9 require four components connected to the amplifier to define the required operation. These components also need to be tightly tolerated to the order of 1% or less. It is more expensive to fabricate a capacitor and three resistors on the same substrate than to fabricate four resistors, because more manufacturing operations are involved. This means that a phase shift circuit costs more than a sum or difference circuit with two inputs.

Table 3.4 lists the number of phase shifts for the SFT structure and the total number of operations (phase shift, sum, and difference) for the FFT structure with $n$ a power of 2.
Table 3.4

Comparison of Analogue Structures

<table>
<thead>
<tr>
<th>n</th>
<th>SFT phase shifts</th>
<th>FFT operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>29</td>
</tr>
<tr>
<td>16</td>
<td>176</td>
<td>81</td>
</tr>
<tr>
<td>32</td>
<td>832</td>
<td>209</td>
</tr>
</tbody>
</table>

On the basis of this comparison alone, without counting the cost of the multiple input summing devices required in the SFT structure, the FFT structure is cheaper than the SFT structure when the number of inputs is 8 or more.

3.7 Extension of Synthesis using Networks Related to the DFT

The FFT decomposition enables analogue beam-forming networks, which compute the DFT of the aperture samples, to be synthesised from devices or circuits which are themselves DFT networks with relatively few inputs compared with the total number of samples, n. An extension of this technique allows other devices or circuits, which are simply related to DFT networks, to be used.

It will be shown, in section 3.7.3, that a slight economic advantage can be gained when using the hybrid circuit technology which has already been mentioned. However, a more general treatment is covered in sections 3.7.1 and 3.7.2 because this unifies the design of the multiple beam-forming microwave networks described by Butler and Lowe(20), Shelton and Kelleher(11), and Allen(21).
3.7.1 Related Networks

The general device which will be considered is equivalent to a DFT network with a phase shift in series with each input and output. If the behaviour of this device is represented by the matrix $G_p$, and the DFT network, of dimension $p$, is represented by the matrix $P$, then

$$G_p = S_O \cdot P \cdot S_I \quad - (3.32)$$

All matrices in equation 3.32 are of dimension $p$, and $S_O$ and $S_I$ are diagonal matrices containing the phase shift terms. The element in row $\ell$ and column $m$ of matrix $S_O$ is

$$s_O(\ell,m) = \delta_{\ell m} \exp(j\psi_0\ell) \quad - (3.33)$$

for $\ell = 0, 1, 2, \ldots, (p-1)$

$m = 0, 1, 2, \ldots, (p-1)$

and $j = \sqrt{-1}$

The corresponding element in matrix $S_I$ is

$$s_I(\ell,m) = \delta_{\ell m} \exp(j\psi_I\ell) \quad - (3.34)$$

for $\ell = 0, 1, 2, \ldots, (p-1)$

$m = 0, 1, 2, \ldots, (p-1)$

with, $\psi_{I0} = 0 \quad - (3.35)$

Equation 3.35 is not restrictive in any way, since the equations describing the performance of any network of the type under consideration can be modified so that equation 3.35 is satisfied.
Figure 3.10
Decomposition of a DFT of dimension 6

Figure 3.11
Network equivalents

Figure 3.12
DFT of dimension 6 using network equivalents
It follows from equation 3.32 that

\[ p = S_0^{-1} \cdot G_p \cdot S_1^{-1} \]  

- (3.36)

where \((\cdot)^{-1}\) indicates the inverse of a matrix. Since \(S_0\) is diagonal, an element in \(S_0^{-1}\) is simply the complex conjugate of the corresponding element in \(S_0\). Similarly, the same is true for \(S_1\). Equation 3.36 can be expressed, alternatively, as

\[ p = \overline{S_0} \cdot \overline{G_p} \cdot \overline{S_1} \]  

- (3.37)

where the bar indicates the complex conjugate.

Equation 3.37 shows that, if the FFT decomposition has been carried out in terms of a DFT of dimension \(p\), defined by \(P\), then each network corresponding to \(P\) in the FFT schematic can be replaced by the device equivalent to \(G_p\), with a phase shift in series with each input and output. This substitution will be illustrated for a beam-forming network having 6 inputs. The FFT decomposition for \(n = 6 = 3 \cdot 2 = p \cdot q\) was carried out in Appendix A as an example. The FFT schematic for this case is shown in Figure 3.10. Figure 3.11 shows the general network equivalent of the DFT's of dimensions 3 and 2, and Figure 3.12 shows the resulting network which is equivalent to that shown in Figure 3.10.

The effect of the substitution is to introduce \(q \cdot (p-1)\) phase shifts before the first set of devices \(G_p\), a further \((p + q - 1)\) phase shifts between the sets of devices \(G_p\) and \(G_Q\), and \(n\) phase shifts after the last set of devices \(G_Q\). This increase in the number of phase shifts may not be as much of a disadvantage as it might appear. In the first place an important simplification can be carried out. This results from the
observation that any phase shift appearing in the FFT schematic can be transferred from one side of a device to the other side. At the same time, all of the other phase shifts attached to the inputs and outputs of the same device are altered in value, but the technique does enable the number of phase shifts in the FFT schematic to be reduced.

Consider a device, with \( p \) inputs and \( p \) outputs, which is represented by the matrix \( G_p \) and which has a phase shift in series with each input. This situation is represented by the product of two matrices,

\[
G_p \cdot \begin{bmatrix}
  e^{j\psi_0} & & & \\
  & e^{j\psi_1} & & \\
  & & \ddots & \\
  & & & e^{j\psi_{p-1}}
\end{bmatrix}
= G_p \cdot \begin{bmatrix}
  e^{j\psi_0} & & & \\
  & e^{j\psi_0} & & \\
  & & \ddots & \\
  & & & e^{j\psi_0}
\end{bmatrix} \cdot \begin{bmatrix}
  1 & & & \\
  e^{j(\psi_1-\psi_0)} & & & \\
  & \ddots & & \\
  & & e^{j(\psi_{p-1}-\psi_0)}
\end{bmatrix}
\]

\[
\begin{bmatrix}
  e^{j\psi_0} \\
  & e^{j\psi_0} \\
  & & \ddots \\
  & & & e^{j\psi_0}
\end{bmatrix} \cdot G_p \cdot \begin{bmatrix}
  1 & & & \\
  e^{j(\psi_1-\psi_0)} & & & \\
  & \ddots & & \\
  & & e^{j(\psi_{p-1}-\psi_0)}
\end{bmatrix}
\]
Figure 3.13

Alternative form of DFT using network equivalents
The phase shift represented by $e^{j\psi_0}$ has been transferred to the output side of the device, where it can be combined with any other phase shifts already in series with the outputs. If the device had a phase shift in series with each input and output, the net effect of this procedure is to reduce the number of these phase shifts by one.

Applying this technique to the schematic in Figure 3.12, the device blocks labelled $G_2$ have a phase shift associated with each input and output. Transferring a phase shift from the input of each device to the output gives a schematic, shown in Figure 3.13, which has three less phase shifts than that shown in Figure 3.12. In general, there will be $p(q-1)$ phase shifts between the sets of devices $G_p$ and $G_q$.

Of the phase shifts remaining at the inputs and outputs of FFT schematic, if the interpolation technique to be described in Chapter 4 is used, then this requires a set of compensating phase shifts at the inputs and outputs of the network in any case.

If the beam outputs are not to be interpolated and only amplitude information retained, as is commonly the case with microwave networks, then the phase shifts at the outputs of the FFT schematic can be discarded. If the device, represented by matrix $G_p$, is equivalent to the DFT matrix, $P$, with a linear phase progression across the inputs then a further simplification is possible, and this is examined in section 3.7.2.

### 3.7.2 A Uniform Phase Progression across Device Inputs

Repeating equation 3.32,

$$G_p = S_0 \cdot P \cdot S_I$$

- (3.32)
and equations 3.34 and 3.35,

\[ s_I(\ell, m) = \delta_{\ell,m} \exp(j\psi_{I\ell}) \]  \hspace{1cm} (3.34)

where:
\[ \ell = 0, 1, 2, \ldots, (p-1) \]
\[ m = 0, 1, 2, \ldots, (p-1) \]

with, \[ \psi_{I0} = 0 \]  \hspace{1cm} (3.35)

The situation to be examined in this section is defined by

\[ \psi_{I\ell} = 2\cdot\psi_{I1} \]  \hspace{1cm} (3.38)

Consider, first, a uniformly spaced linear array which has a linear phase progression applied to the outputs of the transducers before being connected to a DFT network. The phase increment between adjacent array elements is \( \alpha \), so that the kth input to the DFT network is \( x_k \cdot \exp(jk\alpha) \). The rth coefficient of the DFT of this set of inputs is, from equation 2.32,

\[ y_r = \sum_{k=0}^{n-1} x_k \cdot \exp(jk\alpha) \cdot \exp(-j2\pi rk/n) \]

i.e.

\[ y_r = \sum_{k=0}^{n-1} x_k \cdot \exp\left\{-j2\pi k\left(\frac{r}{n} - \frac{\alpha}{2\pi}\right)\right\} \]  \hspace{1cm} (3.39)

Comparing equations 3.39 and 2.29, using identity 2.34,

\[ y_r \equiv Q_1^*(\frac{r}{nL} - \frac{\alpha}{2\pi}) \]  \hspace{1cm} (3.40)

Comparing equations 3.40 and 2.35, the effect of the linear phase progression across the array has been to shift the positions of the sampling
points of $Q_1(s)$ by an amount $s = \frac{-\alpha}{2\pi L}$.

As a result of the FFT decomposition, the first stage in the combination of the signals from the transducer array is performed by a set of $\frac{n}{p}$ DFT's of dimension $p$. This set of DFT's is represented by the matrix $P_p$ in equation 3.15. The reordering matrix, $C$, which precedes $P_p$, causes the set of $p$ inputs for each DFT matrix, $P$, to be taken from points in the aperture which are separated by $\frac{n}{p}$ element spacings. This can be verified by referring back to Figure 3.10, where $p = 3$ and $\frac{n}{p} = 2$.

Now, when each matrix, $P$, is replaced by

$$
P = \tilde{S}_0 \cdot C_p \cdot \tilde{S}_I$$

it is convenient to distinguish between the $\frac{n}{p}$ matrices, $C_p$, by the addition of a further subscript $i$, where $i = 0, 1, 2, \ldots \left(\frac{n}{p} - 1\right)$. Then the $i$th input to $C_{Fi}$ is connected to array element $k = i + \frac{2.3}{p}$.

It will be demonstrated, now, that the phase shifts which appear in series with each array element, as a result of the substitution expressed in equation 3.37, can be made into a progressive sequence provided that equation 3.38 is satisfied. Then, if this sequence of phase shifts is removed from the schematic diagram, the effect is equivalent to inserting the conjugate of the sequence in front of the DFT network. It has just been shown that this displaces the main lobes of all the beams by a constant amount.

The phase shifts, which are in series with each element, form a progressive sequence if the phase shift in series with element $k$ is
Equation 3.41 contains a component which depends on the device number, \( i \), and a component which depends on the input number, \( \lambda \). As a result of the substitution expressed in equation 3.37, there already exists a phase shift of \( -2\Psi_{II} \) radians in series with input \( \lambda \). Therefore, if \( \frac{n}{p}\psi = -\Psi_{II} \), from equation 3.41, or

\[
\psi = -\frac{p}{n} \Psi_{II}
\]

then a further phase shift of

\[
\psi = -i \cdot \frac{p}{n} \Psi_{II}
\]

must be added in series with each input of \( G_{pi} \) for equation 3.41 to be satisfied. This may be achieved by application of the technique described in section 3.7.1. This amounts to subtracting \( -i \cdot \frac{p}{n} \Psi_{II} \) from the phase shift in series with each output of \( G_{pi} \), and transferring it to the input side.

There exists now, within the boundaries of the FFT schematic, a phase shift of \( -k \cdot \frac{p}{n} \Psi_{II} \) in series with element \( k \) in the array of transducers. If these phase shifts are discarded, this is equivalent to inserting the sequence of conjugate phase shifts across the array, in front of the DFT network. The phase increment between adjacent elements for the conjugate sequence is \( \frac{p}{n} \Psi_{II} \), so the rth output of the analogue beam-forming network is, from identity 3.40

\[
Y_r = Q_1^*(\frac{r}{nL} - \frac{p\Psi_{II}}{2\pi nL})
\]

- (3.44)
Figure 3.14

Development of network for 4 elements
Equation 3.44 shows that the modified DFT network generates samples of the Fourier transform of the sampled field, as before, but all sample points are shifted by an amount \( s = -\frac{p\psi_{II}}{2 nL} \).

As an illustration, consider the microwave hybrid coupler(11), with two input ports and two output ports, which is specified by the matrix:

\[
G_2 = \begin{bmatrix}
1 & \exp(-j\pi/2) \\
\exp(-j\pi/2) & 1
\end{bmatrix}
\]

\[
= \begin{bmatrix}
1 & 1 \\
\exp(-j\pi/2) & 1 -1
\end{bmatrix}
\begin{bmatrix}
1 & 1 \\
1 & \exp(-j\pi/2)
\end{bmatrix}
\]

\[
(3.45)
\]

Here, \( \psi_{II} = -\frac{\pi}{2} \). From equation 3.46,

\[
\begin{bmatrix}
1 & 1 \\
1 & -1
\end{bmatrix} = \begin{bmatrix}
1 & \exp(j\pi/2)
\end{bmatrix} \cdot G_2 \cdot \begin{bmatrix}
1 & \exp(j\pi/2)
\end{bmatrix}
\]

\[
(3.47)
\]

Figure 3.4 shows the FFT schematic for a DFT of dimension 4. Each DFT of dimension 2 can be replaced according to equation 3.47, and the result is shown in Figure 3.14(a). From equation 3.43, with \( p = 2 \) and \( n = 4 \), a phase shift of \( \frac{\pi}{4} \) radians is to be transferred from the output side of the device labelled (2) to the input side. The result is shown in Figure 3.14(b). A further simplification is made by transferring a phase shift of \( \frac{\pi}{4} \) radians from the inputs of device (4) to the outputs. Removing the phase shifts at the inputs and outputs of the network gives the 4-element multiple-beam forming network shown in Figure 3.14(c). This network appears in reference (11).
Figure 3.15

Surrounding amplifier

\[ V_3 = -(V_1 + V_2) \]

Figure 3.16

8 element beam forming network
3.7.3 An 8-element Microelectronic Beam-Forming Network

If the summing configuration shown in Figure 3.15 can be used instead of that shown in Figure 3.9(a), then a small advantage can be gained. This is because only three closely matched resistors are required to define the summing operation in Figure 3.15, and the value of $R_1$ is not critical in this application. A functional block $G_2$ can now be defined,

$$G_2 = \begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix}$$

which will be used to replace the 2-input DFT in the FFT schematic in Figure 3.5, according to the relation,

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} -1 & \end{bmatrix} \cdot G_2$$

The result of the substitution and simplification procedure is shown in Figure 3.16. A network, corresponding to the schematic in Figure 3.16 and designed to operate at 32 kHz, was built and tested. A detailed analysis of its behaviour is presented in later chapters.

3.8 The Reciprocal Nature of the FFT Schematic

The DFT matrix, $N$, is symmetric, so

$$N = N^T$$

- (3.48)

Using the rule for the transpose of a product, equation 3.15 becomes,

$$N = C^T \cdot P^T \cdot R^T \cdot C \cdot Q^T \cdot C^T$$

- (3.49)
Now

\[ P_P^T = P_P \]
\[ R^T = R \]  \hspace{1cm} - (3.50) \]

and

\[ Q_Q^T = Q_Q \]

Therefore, equation 3.49 becomes

\[ N = C^T \cdot P_P \cdot R \cdot C \cdot Q_Q \cdot C^T \]  \hspace{1cm} - (3.51) \]

The schematic for equation 3.51 is simply related to that for equation 3.15. The forms of the schematics are identical, but the input arrays are interchanged. This means that the DFT can be computed by applying the input samples to either end of the FFT schematic.
4.1 Introduction

Within the angular resolution limit of the system\(^{(5, 27)}\) it is possible to make an estimate of the locations of sources in the far field of an aperture by inspecting the far field source distribution function \(Q_1(s)\). A convenient method to implement is to generate the required section of \(|Q_1(s)|\) as a function of time, and use this to intensity modulate a sweep on a cathode ray tube. The sweep axis is calibrated in angular measure, and a source shows up as a bright spot on the screen. When used in systems which also provide for discrimination in range, the familiar B-scope display is generated. An example of this can be found in reference 1.

In Chapter 2, the \(n\) outputs of a DFT operation, on the \(n\) samples of the field within an aperture, were shown to be samples of \(Q_1^*(s)\). Using a digital processor, the field in the aperture is sampled both in space and in time. For an analogue processor, the field is sampled in space, and the set of DFT outputs are sampled in time. Both processors, assumed to be ideal, produce identical sets of samples of \(Q_1^*(s)\) at the same point in time. In this chapter, the problems associated with the reconstruction of \(Q_1^*(s)\) from this set of samples are investigated.

4.2 The Spectrum of the Far Field Source Distribution

The far field source distribution can be reconstructed as a time waveform, suitable for use in displays, by regarding the samples of
$Q_1(s)$ as a set of samples of the time waveform $f_T(t)$. The two functions are identical in form, so that

$$f_T(t) \equiv Q_1^*(s) \quad - (4.1)$$

The independent variable transformation, $t = s$, is normalised, so that $s = 1$ corresponds to $t = 1$ second. Any scaling of the time axis can be considered at a later stage.

$f_T(t)$ is a periodic complex function with period $T$, corresponding to $1/L$; and the interval between samples of $f_T(t)$ is $T = T/n$, corresponding to $1/\lambda L$. The amplitude spectrum, $F_T(f)$, of $f_T(t)$ is defined by

$$F_T(f) = \int_{-\infty}^{\infty} f_T(t) \exp(-j2\pi ft) \, dt \quad - (4.2)$$

The sampled field in the aperture is the inverse Fourier transform of $Q_1(s)$, from equation 2.20,

$$q_1^*(\lambda) = \int_{-\infty}^{\infty} Q_1^*(s) \exp(j2\pi s\lambda) \, ds \quad - (4.3)$$

Comparing equations 4.2 and 4.3, and using the identity 4.1, it can be seen that

$$F_T(f) \equiv q_1^*(-\lambda) \quad - (4.4)$$

Now, $q_1^*(\lambda)$, is a sequence of $n$ uniformly spaced delta functions, separated by $\lambda$, and of strength, $q_1(\lambda L)$, corresponding to the field sample value at array element $k$. 
Figure 4.1
Relation between $F_T(f)$ and $q_1(n)$

Figure 4.2
$F_T^*(f)$ and complex filter characteristic, $H(f)$, to recover $f_T(t)$
\[ q_1^*(k) = \sum_{k=0}^{n-1} q_1(kL) \cdot \delta(k - kL) \] - (4.5)

Therefore, from equation 4.4,

\[ F_T(f) = \sum_{k=0}^{n-1} F_T(-\frac{k}{T}) \cdot \delta(f + \frac{k}{T}) \] - (4.6)

where \( F_T(-\frac{k}{T}) = q_1(kL) \) - (4.7)

The spectrum, \( F_T(f) \), consisting of \( n \) lines, is shown in Figure 4.1 in relation to \( q_1^*(k) \).

4.3 Reconstruction as a Linear Complex Filtering Process

The function \( f_T(t) \) is periodic, so the sample sequence, \( f_T^*(t) \), of the entire function, \( -\infty < t < \infty \), may be generated by repeating the set of samples, taken uniformly through the period \( 0 \leq t < T \), which are formed by the DFT operation.

If,

\[ s_T(t) = \sum_{r=-\infty}^{\infty} \delta(t - rt) \] - (4.8)

then

\[ f_T^*(t) = f_T(t) \cdot s_T(t) \] - (4.9)

where

\[ f_T^*(rt) = f_T((r - n)t) \] - (4.10)

The spectrum, \( F_T^*(f) \), of \( f_T^*(t) \) may be obtained simply, since

\[ P_T^*(f) = P_T(f) \ast S_T(f) \] - (4.11)

where \( S_T(f) \leftrightarrow s_T^*(t) \)
i.e. \[ F^*_T(f) = F_T(f) \ast \frac{1}{T} \sum_{r=-\infty}^{\infty} \delta(f - \frac{r}{T}) \] 

In the range, \( 0 \leq k \leq n-1 \)
\[ F^*_T(-\frac{k}{T}) = \frac{1}{T} F_T(-\frac{k}{T}) \] 

and elsewhere
\[ F^*_T(\frac{k}{T}) = F^*_T(\frac{k-n}{T}) \] 

From equation 4.13, the required continuous function, \( f_T(t) \), can be reconstructed by passing the sample sequence, \( F^*_T(t) \), through a filter whose transfer function \( H(f) \) is defined by,
\[ H(f) = \frac{1}{T}, \quad \frac{(n-1)}{T} - \Delta f < f < \Delta f \]
where \( \Delta f < \frac{1}{T} \) 

and \( H(f) = 0 \) elsewhere.

The frequency response, \( H(f) \) v. \( f \), for this filter in conjunction with \( F^*_T(f) \) is shown in Figure 4.2.

The impulse response, \( h(t) \), of the filter is defined by
\[ h(t) \leftrightarrow H(f) \]

and \( h(t) \) is a complex function of time since \( H(f) \neq \overline{H}(-f) \). For this reason, the filter will be referred to as a complex filter to distinguish it from a real filter, which has a real impulse response.
4.3.1 A Complex Filter

Crystal and Ehrman (28) use the technique of forming a complex band-pass filter, by frequency shifting a low-pass filter, to design a digital filter which has complex coefficients. The possibility of using a digital filter to construct $f_T(t)$ is not ruled out in principle. However, the initial investigation was directed at producing an analogue filter, and since this gave an acceptable solution to the problem, digital filters were not considered further.

When the frequency band of interest lies wholly in either the positive half or the negative half of the frequency plane, the problem is resolved into synthesising a filter which has zero transmission in one half of the plane, and a passband suitably positioned in the other half plane.

Using results established in analytic signal theory (29), the frequency response, $H(f)$, of a filter with an impulse response,

$$h(t) = h_1(t) - j\hat{h}_1(t)$$  \hspace{1cm} (4.17)

is given by

$$H(f) = 2H_1(f), \quad f < 0$$ \hspace{1cm} (4.18)

$$H(f) = 0, \quad f > 0$$

where $\hat{h}_1(t)$ is the Hilbert transform of the impulse response of the real filter defined by

$$h_1(t) \leftrightarrow H_1(f)$$  \hspace{1cm} (4.19)
(a) Real filter frequency response

(b) Complex filter frequency response

Figure 4.3

Complex filter schematic

Figure 4.4
\( \hat{h}_1(t) \) can be obtained by passing \( h_1(t) \) through a circuit which imposes a phase shift of \( -\pi/2 \) radians on all frequency components. The frequency responses of the real filter and its associated complex filter are shown in Figure 4.3.

Applying the complex sequence,

\[
x_T^*(t) = f_x^*(t) + jf_y^*(t)
\]

- (4.10)

to the complex filter defined by equation 4.17, the output

\[
g(t) = g_x(t) + jg_y(t)
\]

- (4.21)

is given by

\[
g(t) = f_T^*(t) * h(t)
\]

i.e.

\[
g_x(t) + jg_y(t) = \{f_x^*(t) + jf_y^*(t)\} * \{h_1(t) - j\hat{h}_1(t)\}
\]

\[
= \{f_x^*(t) * h_1(t) + f_y^*(t) * \hat{h}_1(t)\}
\]

\[
+ j\{f_y^*(t) * h_1(t) - f_x^*(t) * \hat{h}_1(t)\}
\]

- (4.22)

The schematic diagram for the filtering operation defined by equation 4.22 is shown in Figure 4.4. The practical difficulties in constructing such a filter involving four closely matched real filters and two wideband phase shifters, are considerable. For this reason, the approach described in section 4.4 was adopted.
Figure 4.5
Relation between $q_1(\ell)$ and $r_1(\ell)$

Figure 4.6
$S(f)$ v. $f$, for $n = 8$

Figure 4.7
$S(f)$ v. $f$, for $n$ odd
4.4 Reconstruction as a Linear Real Filtering Process

The one-sided nature of the spectrum of $Q_1^*(s)$, and hence $f_1(t)$, is a direct result of taking one end of the aperture as the origin for the DFT network. If the origin is moved to the centre of the aperture, the Fourier transform of the field samples is a new function, $R_1^*(s)$, which is simply related to $Q_1^*(s)$. The two functions have the same modulus and differ only in phase. The important difference in their spectra is that the spectrum of $R_1^*(s)$ extends equally into both the positive and negative halves of the frequency plane, and so the corresponding function of time should be recoverable by a real filtering process.

4.4.1 Shifting the Origin of the Aperture

Figure 4.5 shows the original sequence of field samples, $q_1^*(\xi)$, with the origin taken at one end of the aperture, and the shifted sequence, $r_1^*(\xi)$, with the origin taken at the centre of the aperture. For $n$ samples,

$$r_1^*(\xi) = q_1^*\{\xi + (n-1) \frac{L}{2}\}$$  \hspace{1cm} (4.23)

Now, $r_1^*(\xi) \leftrightarrow R_1^*(s)$ \hspace{1cm} (4.24)

and $q_1^*\{\xi + (n-1) \frac{L}{2}\} \leftrightarrow Q_1^*(s) \cdot \exp\{j2\pi s(n-1) \frac{L}{2}\}$ \hspace{1cm} (4.25)

Combining equations 4.23, 4.24, and 4.25, gives

$$R_1^*(s) = Q_1^*(s) \cdot \exp\{j\pi s(n-1) L\}$$  \hspace{1cm} (4.26)
At the sample points  \( s = \frac{r}{nL} \), (\( r = 0, 1, \ldots, n-1 \))

\[
R_{1}^*(\frac{r}{nL}) = Q_{1}^*(\frac{r}{nL}) \cdot \exp\{j\pi r \frac{(n-1)}{n}\} \tag{4.27}
\]

or \( R_{1}^*(\frac{r}{nL}) = Q_{1}^*(\frac{r}{nL}) \cdot C_r \tag{4.27(a)} \)

where \( C_r = \exp\{j\pi r \frac{(n-1)}{n}\} \)

Equation 4.27(a) states that samples of the new far field source distribution function, \( R_{1}^*(s) \), are obtained by inserting the phase shift \( C_r \) in series with output \( r \) of the DFT network.

Defining a new function of time,

\[
s(t) \equiv R_{1}^*(s) \tag{4.28}
\]

then, from equation 4.26,

\[
s(t) = f_T(t) \exp\{j\pi t \frac{(n-1)}{T}\} \tag{4.29}
\]

If \( s(t) \leftrightarrow S(f) \tag{4.30} \)

then, from equation 4.29,

\[
S(f) = F_T\{ f - \frac{(n-1)}{2T}\} \tag{4.31}
\]

The spectrum of \( s(t) \), for \( n = 8 \), is shown in Figure 4.6, and the important feature is that the centre of the band occupied is at zero frequency.

Figure 4.7 shows the spectrum of \( s(t) \) when \( n \) is odd. In this case, \( s(t) \) is a periodic function, with period \( T \), and so the sample
Figure 4.8

Baseband of \( S^*(f) \)

\( S^*(f) \) v. \( f \)

Figure 4.9

Real filter schematic

Figure 4.10

Detector Schematic
sequence, \( s^*(t) \), for the entire function, \( -\infty < t < \infty \), can be constructed by repeating the set of \( n \) samples in the period \( 0 \leq t < T \) which are generated by the DFT. This situation corresponds to that described in section 4.3. The sample sequence with spacing \( \tau \) is given by

\[
s^*(t) = s(t) \cdot \sum_{r=-\infty}^{\infty} \delta(t - r\tau)
\]  

(4.32)

When \( n \) is even, the spectrum of \( s(t) \) is of the form shown in Figure 4.6. This indicates that \( s(t) \) is periodic, with period \( 2T \), and possesses half wave symmetry because of the absence of even harmonics of the fundamental frequency, \( 1/2T \). Here,

\[
s(t + T) = -s(t)
\]  

(4.33)

In this case, \( s^*(t) \) is created in the same way as when \( n \) is odd, except that alternate blocks of \( n \) samples are inverted.

The spectrum, \( S^*(f) \), of the sequence \( s^*(t) \) is given by

\[
S^*(f) = S(f) \ast \frac{1}{T} \sum_{r=-\infty}^{\infty} \delta(f - \frac{r}{T})
\]  

(4.34)

\( S^*(f) \) is periodic, with period \( 1/T \), as shown in Figure 4.8 for \( n = 8 \), and the continuous function \( s(t) \) can be reconstructed by passing the sequence \( s^*(t) \) through a filter with the specification,

\[
H(f) = \mathbb{T}, \quad -(n-1) \frac{2}{2T} - \Delta f < f < (n-1) \frac{2}{2T} + \Delta f
\]  

(4.35)

\[
= 0, \quad |f| > (n-1) \frac{2}{2T} + \Delta f
\]

where \( \Delta f < \frac{1}{T} \).
The filter specified by equation 4.35 has a real impulse response, $h(t)$, where

$$h(t) \leftrightarrow H(f)$$

The output, $v(t)$, of the filter is given by

$$v(t) = s^*(t) \ast h(t)$$

i.e.

$$v_x(t) + j v_y(t) = [s_x^*(t) + j s_y^*(t)] \ast h(t)$$

$$= e_x^*(t) \ast h(t) + j s_y^*(t) \ast h(t) \quad (4.36)$$

The schematic diagram for this filtering operation is shown in Figure 4.9, and just two real filters are required, with no cross coupling between channels, and no wideband phase shifters. This real filter is easier to implement than the complex filter shown in Figure 4.4, at the expense of introducing (n-1) narrowband phase shifts at the outputs of the DFT network.

The specification in equation 4.35 cannot be met in practice but realisable filters can be designed to have sufficiently high attenuation outside the band of interest ($f > \frac{n-1}{2T}$) to give an acceptable reconstruction of $s(t)$.

### 4.5 Generation of the Intensity Modulation Waveform

The modulus of the output of the reconstruction filter is used to intensity modulate the display.

$$|v(t)| = \sqrt{v_x^2(t) + v_y^2(t)} \quad (4.37)$$
The schematic diagram for the 'detector' which implements equation 4.37 is shown in Figure 4.10.

4.5.1 A Simplification

The realisation of circuits which perform the squaring and square root operations necessary in equation 4.37 has been made easier with the advent of high quality analogue multiplier integrated circuits. Even so, it is possible to remove the need for these circuits and still generate a display which should be acceptable in many applications of the sonar.

Taking the centre of the aperture as the origin, consider the field, $r_1(\ell)$, in the aperture due to a single source in a direction corresponding to $s = s_1$

$$r_1(\ell) = q_1\{\ell + \frac{(n-1)}{2}, L\}$$

(4.38)

From equation 2.5,

$$r_1(\ell) = q_1\{\frac{(n-1)}{2}, L\} \exp(j2\pi s_1\ell)$$

$$= r_1(0) \exp(j2\pi s_1\ell)$$

(4.39)

Removing the phase term from $r_1(0)$ in equation 4.39 and rearranging, leads to

$$r_1(\ell) = \exp(j\phi_0) \cdot |r_1(0)| \cdot \exp(j2\pi s_1\ell)$$

$$= \exp(j\phi_0) \cdot \{r_x(\ell) + j \cdot r_y(\ell)\}$$

(4.40)

where $r_x(\ell)$ is an even function, and $r_y(\ell)$ is an odd function because of the conjugate symmetry of $\exp(j2\pi s_1\ell)$ about the origin.
Now, 
\[ R_1^*(s) = \int_{-\infty}^{\infty} r_1^*(l) \exp(-j2\pi s l) \, dl \]  
\[ = \exp(j\phi) \int_{-\infty}^{\infty} \{ r_x^*(l) + j r_y^*(l) \} \exp(-j2\pi s l) \, dl \]  
\[ : R_1^*(s) = \exp(j\phi) \{ \int_{-\infty}^{\infty} \{ r_x^*(l) \cos 2\pi s l + r_y^*(l) \sin 2\pi s l \} \, dl \]  
\[ + j \int_{-\infty}^{\infty} \{ r_y^*(l) \cos 2\pi s l - r_x^*(l) \sin 2\pi s l \} \, dl \]  
\[ = (4.41) \]  
\[ (4.42) \]  
Since \( r_x^*(l) \) is an even function and \( r_y^*(l) \) is an odd function, the second integral in equation 4.42 is zero. The first integral in equation 4.42 is wholly real, and 
\[ \left| R_1^*(s) \right| = \left| \int_{-\infty}^{\infty} \{ r_x^*(l) \cos 2\pi s l + r_y^*(l) \sin 2\pi s l \} \, dl \right| \]  
\[ = (4.43) \]  
If, \( R_1^*(s) = R_x^*(s) + j R_y^*(s) \)  
\[ (4.44) \]  
then from equations 4.44, 4.43 and 4.42 
\[ \left| R_x^*(s) \right| = |\cos \phi_0| \cdot \left| R_1^*(s) \right| \]  
\[ = (4.45) \]  
and \( \left| R_y^*(s) \right| = |\sin \phi_0| \cdot \left| R_1^*(s) \right| \)  
\[ = (4.46) \]  
Equations 4.45 and 4.46 indicate that the modulus of either the real or the imaginary part of \( R_1^*(s) \) differs from the modulus of the complex function by only a scale factor. If the larger of \( |R_x^*(s)| \) and \( |R_y^*(s)| \) is selected then the scale factor will always be greater than \( 1/\sqrt{2} \).
Figure 4.11

Alternative detector schematic

K = scale factor

> 1/\sqrt{2}
Since \( s(t) \equiv R_1^*(s) \), \( R_x^*(s) \) and \( R_y^*(s) \) correspond to the outputs of the reconstruction filter.

An alternative schematic diagram for the generation of the intensity modulation waveform can now be drawn, and this is shown in Figure 4.11. A circuit which takes the modulus of a real waveform simply inverts the signal when it goes negative. The reconstruction scheme described in this chapter generates more than one cycle of \(|R_1^*(s)|\), and so it is possible to decide which of the real and imaginary components is the larger before it is used for display. This function is performed by a peak level memory in each channel and a comparator circuit.

Whilst this alternative arrangement works very well for a single source, the behaviour in a multiple source situation cannot be predicted except in average terms. This is because the field in the aperture no longer exhibits conjugate symmetry about the centre of the aperture, and consequently the moduli of the real and imaginary parts of \( R_1^*(s) \) are not related to the modulus of the complex function by a scale factor. The effect on the display will be to make the sources appear to fluctuate in intensity as the relative positions of the sources and the aperture change. This will reduce the detection capability of the sonar, but the effect may be tolerable in many applications.

4.6 Choice of Filter

So far, the reconstruction technique has been based on filtering the infinite sequence of samples, \( s^*(t) \), and so the required waveform is the steady state output of the filter. A within-pulse sector scan sonar requires a suitable intensity modulating waveform to be generated in the duration of one pulse. Consequently, a limited number of cycles of \( s^*(t) \),
perhaps one, can be filtered in the time available, and the filter output will not have achieved its steady state.

The continuous waveform obtained by passing the sample sequence through a realisable filter will differ from \( s(t) \) due to the superposition of effects caused by:

(i) amplitude and phase distortion within the passband of the filter,
(ii) finite attenuation outside the passband,
(iii) the filter output being still in the transient state,
(iv) non-identical filters in the real and imaginary channels.

Effects due to (i), (ii), and (iii), will be examined qualitatively here, by using a digital computer simulation of a reconstruction filter. A quantitative assessment of the error contribution of a particular filter will be deferred to Chapter 6, where the deficiencies of the analogue and digital processors are also discussed.

4.6.1 Amplitude and Phase Distortion within the Pass-Band

The filter characteristic required for distortionless reconstruction of \( s(t) \) has a uniform amplitude v. frequency response and a linear phase v. frequency response within the pass-band. Any departure from these characteristics will produce distortion in the output waveform, and this effect will be objectionable using a filter having a high enough cut-off rate to be useful in this application. However, it is possible to compensate non-ideal filter characteristics by altering the spectrum of the input sequence. This is achieved by inserting the appropriate gain and phase shift in series with each transducer element, before the DFT processor.
The transfer function, \( H(f) \), of the filter may be written in the form

\[
H(f) = A(f) \cdot \exp\{j\phi(f)\}
\]  

(4.47)

where \( A(f) \) represents the amplitude characteristic and \( \phi(f) \) represents the phase characteristic. The spectrum, \( V(f) \), of the waveform at the output of the filter, in response to the input sequence \( s^*(t) \) is given by

\[
V(f) = H(f) \cdot S^*(f)
\]  

(4.48)

In the band \( \frac{- (n-1)}{2T} \leq f \leq \frac{n-1}{2T} \),

\[
S^*\left(\frac{n-1}{2T} - \frac{k}{T}\right) = \frac{q_1(kL)}{T}
\]  

(4.49)

\[ k = 0, 1, \ldots, n-1 \]

from equations 4.7, 4.13, and 4.31. Inserting amplitude and phase compensation, \( B_k \), in series with element \( k \) modifies \( S^*(f) \), so that now equation 4.49 becomes

\[
S^*\left(\frac{n-1}{2T} - \frac{k}{T}\right) = \frac{B_k \cdot q_1(kL)}{T}
\]  

(4.50)

From equation 4.48,

\[
V\left(\frac{n-1}{2T} - \frac{k}{T}\right) = q_1(kL) \cdot \frac{B_k}{T} \cdot H\left(\frac{n-1}{2T} - \frac{k}{T}\right)
\]  

(4.51)

Now, from equations 4.31 and 4.7

\[
S\left(\frac{n-1}{2T} - \frac{k}{T}\right) = q_1(kL)
\]  

(4.52)
so that if $B_k$ is chosen such that

$$
B_k = \frac{1}{H\left(\frac{n-1}{2T} - \frac{k}{T}\right)}
$$

(4.53)

equation 4.51 becomes

$$
V\left(\frac{n-1}{2T} - \frac{k}{T}\right) = S\left(\frac{n-1}{2T} - \frac{k}{T}\right)
$$

(4.54)

for $-(n-1) \leq f \leq \frac{n-1}{2T}$

If the attenuation outside the pass-band is such that all other frequency components may be neglected, equation 4.54 is equivalent to

$$
v(t) = s(t)
$$

(4.55)

Equation 4.55 states that the required waveform, $s(t)$, can be obtained in the steady state by using non-ideal filters in conjunction with the compensation defined by equation 4.53.

Combining equations 4.53 and 4.47 gives

$$
B_k = \frac{1}{A\left(\frac{n-1}{2T} - \frac{k}{T}\right)} \cdot \exp\{-j\phi\left(\frac{n-1}{2T} - \frac{k}{T}\right)\}
$$

(4.56)

Amplitude compensation may not be necessary since the effect of the amplitude characteristic of the filter is equivalent to shading the aperture symmetrically about its centre. This gives a different main-lobe width and side-lobe levels in the point source response, compared with a uniformly shaded aperture, and this may be acceptable, or desirable. Any additional shading can be incorporated in $B_k$. 
1. Interpolating filter

\[ C_r = \frac{7\pi r}{8} \]

**Figure 4.12**

Schematic for reconstruction of far field source distribution

**Figure 4.13**

Steady state point source response of system with an ideal low pass filter
Taking a linear phase term out of $\phi(f)$,

$$\phi(f) = -b_{\phi} \cdot f + \phi_1(f)$$

where $b_{\phi}$ is a constant phase gradient, and redefining $B_k$ as

$$B_k = \frac{i}{\Lambda(n-1/k)} \cdot \exp\{-j\phi_1(n-1/2T - k/T)\}$$

gives

$$V(n-1/2T - k/T) = S(n-1/2T - k/T) \exp\{-j b_{\phi}(n-1/2T - k/T)\}$$

for $-(n-1)/2T \leq f \leq n-1/2T$.

Equation 4.59 leads to

$$v(t) = s(t - \frac{b_{\phi}}{2\pi})$$

The output of the filter is now the required waveform delayed by a time $t_d = \frac{b_{\phi}}{2\pi}$. The reason for taking the linear phase term out of $B_k$ as defined in equation 4.56 to give the definition in equation 4.58, is that appropriate choice of $b_{\phi}$ will give sufficiently low values of phase compensation to be applied to the elements near the centre of the aperture that they may be omitted.

The schematic diagram for an 8-element array processor, using the reconstruction technique described in this chapter, is shown in Figure 4.12. Figure 4.13 shows the normalised modulus of the hypothetical point source response of the system for a uniformly shaded aperture, and is the modulus of the steady state output of the ideal low-pass filter specified in equation 4.35. The point source is in a direction corresponding to
Input and output waveforms for zero order hold

Form of amplitude spectrum at output of zero-order hold
s = 0, so the principal maxima of \( v(t) \) occur at \( t = 0 \) and at integral multiples of \( T = 8t \). Even though the period of \( v(t) \) is \( 2T \), the period of \( |v(t)| \) is \( T \) because of halfwave symmetry.

4.6.1.1 The Zero-Order Hold. In the process of reconstructing a continuous waveform from a sequence of samples using a practical filter, it is usual to first pass the samples through a hold circuit (29). The zero-order hold is the simplest type, and the response of the circuit to an impulse of unit strength is a rectangular pulse of unit amplitude. If the duration of the impulse response equals the interval between samples, the impulse sequence at the input is converted to a voltage waveform of the type shown in Figure 4.14.

The impulse response,

\[
h_H(t) = \begin{cases} 1, & 0 \leq t < T \\ 0, & t < 0, \ t \geq T \end{cases}
\]

corresponds to a transfer function

\[
H_H(f) = T \cdot \frac{\sin \pi ft}{\pi ft} \cdot \exp(-j\pi ft)
\]

The spectrum of the waveform resulting from passing the sequence \( s^*(t) \) through a zero-order hold is \( S^*_H(f) \). \( H_H(f) \), and the form of this is shown in Figure 4.15. Obviously, a further stage of filtering is necessary to isolate the band, \( |f| < \frac{n-1}{2T} \).

4.6.1.2 An Illustration. The effect, on the point source response of the system, of compensation for filter characteristics will be demonstrated for a practically realisable filter which is used in conjunction
Figure 4.16
Amplitude characteristic for 9-pole Chebyshev filter, with 0.5 dB ripple

Figure 4.17
Steady state point source response using uncompensated 9-pole Chebyshev filter
with a zero-hold. The filter is a 9-pole Chebyshev low pass filter with 0.5 dB ripple in the pass-band. The amplitude characteristic is shown in Figure 4.16, and the cut-off frequency, \( f_c \),

\[ \{ H(f_c) = -0.5 \text{ dB relative to } H(0) \} \text{ is equal to } \frac{n-1}{2T} \left(= \frac{7}{2T} \right). \]

With no compensation at the input to the DFT network 
\( B_k = 1 \left[ 0^\circ ; \ k = 0, 1, 2, \ldots, 7 \right] \), the point source response corresponding to \( s = 0 \) is shown in Figure 4.17. Once again, this waveform is the modulus of the steady state response of the filter, and it should be compared with the ideal in Figure 4.13.

Now, compensation for the hold circuit and filter is inserted in series with each array element to give overall uniform amplitude shading and a linear phase progression across the array. The phase gradient is chosen so that the phase compensation at the two innermost elements is zero. These two elements, \( k = 3 \) and 4, correspond to the spectral lines at \( f = \pm \frac{1}{2T} \) in \( S^*(f) \), as in equation 4.49. The combined phase lag of the hold circuit and filter at \( f = \frac{1}{2T} \) is 76.8°.

\[ b_\phi = 76.8 \times \frac{\pi}{180} \times \frac{1}{1/2T} \text{ rad./Hz} \]

\[ \therefore \text{ from equation 4.60,} \]

\[ t_d = \frac{76.8 \times \pi \times 2T}{180} \cdot \frac{1}{2\pi} \]

\[ = 3.41 \times \frac{T}{8} \]

\[ = 3.41T \]

The compensation, \( B_k \), required in series with element \( k \) is shown in Table 4.1.
Compensated 9 pole Chebyshev filter, 0.5 dB ripple

Ideal low pass filter

Figure 4.18

Steady state point source responses
### Table 4.1

<table>
<thead>
<tr>
<th>k</th>
<th>$b_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.48</td>
</tr>
<tr>
<td>1</td>
<td>1.22</td>
</tr>
<tr>
<td>2</td>
<td>1.10</td>
</tr>
<tr>
<td>3</td>
<td>1.06</td>
</tr>
<tr>
<td>4</td>
<td>1.06</td>
</tr>
<tr>
<td>5</td>
<td>1.10</td>
</tr>
<tr>
<td>6</td>
<td>1.22</td>
</tr>
<tr>
<td>7</td>
<td>1.48</td>
</tr>
</tbody>
</table>

**Compensation for zero-order hold and 9-pole Chebyshev filter with 0.5 dB ripple**

Figure 4.18(a) shows the normalised steady state system response to a point source at $s = 0$. For comparison, Figure 4.18(b) shows the response to the same source, but using an ideal low-pass filter having the same phase gradient. The difference between the ordinates of the two curves is less than 0.01 at any point along the time axis, and the two reconstructions are indistinguishable for display purposes.

#### 4.6.2 Transient Effects

Assuming that one display sweep is generated per pulse length, $t_p$, the filter must be chosen so that

$$t_p > t_t + T$$

- (4.63)
Figure 4.19

Amplitude characteristic for 5-Pole Chebyshev filter

with 0.5 dB ripple
where, \( t_t \) is the transient delay time taken for the filter output to attain an acceptably close approximation to its steady state;

\[ T = nT \] is the period of the display intensity modulating waveform.

There will also be a settling time during which the filter output is decaying after the input sequence has been removed; however, this period can run concurrently with \( t_t \) for the next input sequence and so is not included in equation 4.63.

4.6.2.1 An Illustration. The transient delay time effect will be illustrated for two different filters; one being the 9-pole Chebyshev filter used in section 4.6.1, and the other is a 5-pole Chebyshev low-pass filter having the same pass-band specification. The amplitude characteristic of the filter is shown in Figure 4.19.

For the 5-pole filter, the compensation, \( E_k \), required in series with element \( k \) is shown in Table 4.2; and \( b_\phi = 1.57T \) radians/Hz, corresponding to \( t_d = 2.0T \).

Figure 4.20(a) shows the normalised steady state response of this filter for a point source at \( s = 0 \); Figure 4.20(b) shows the response of an ideal filter having the same phase gradient; and Figure 4.20(c) shows the modulus of the difference, \( e(t) \), between these two curves. Whilst the response of the 5-pole filter is not as good as that of the 9-pole filter, the maximum deviation from the ideal is still less than 0.03, and the most noticeable effect is in the asymmetric side-lobe levels shown in Figure 4.20(a).
Point source response using 5 pole Chebyshev filter, 0.5 dB ripple

Figure 4.20

- **(a)** steady state point source response ($s = 0$)
- **(b)** ideal steady state point source response ($s = 0$)
- **(c)**

$$ |e(t)| = |v(t)| - |y(t)| $$
To compare the transient responses of the two filters used in conjunction with a zero order hold, just one cycle of $s^*(t)$, of length $2T = 16\tau$, is applied to the filter inputs. Figure 4.21(a) shows the response of the 9-pole filter, and the difference between this curve and the ideal filter steady state response is shown in Figure 4.21(b). The corresponding response and error curve for the 5-pole filter is shown in Figures 4.22(a) and (b).

Inspection of Figures 4.21 and 4.22 shows that the transient response of the 5-pole filter is considerably shorter in duration than that of the 9-pole filter, as would be expected. Furthermore, from Figures 4.22(b) and 4.20(c), little improvement can be expected by allowing for a transient delay time greater than $t_t = 8\tau$ for the 5-pole filter. By extrapolation of Figure 4.22(b), a comparable error for the 9-pole filter will be achieved only after a delay of $t_t > 32\tau$. Assuming for the moment

<table>
<thead>
<tr>
<th>k</th>
<th>$B_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.48</td>
</tr>
<tr>
<td>1</td>
<td>1.22</td>
</tr>
<tr>
<td>2</td>
<td>1.10</td>
</tr>
<tr>
<td>3</td>
<td>1.03</td>
</tr>
<tr>
<td>4</td>
<td>1.03</td>
</tr>
<tr>
<td>5</td>
<td>1.10</td>
</tr>
<tr>
<td>6</td>
<td>1.22</td>
</tr>
<tr>
<td>7</td>
<td>1.48</td>
</tr>
</tbody>
</table>

Table 4.2

Compensation for zero-order hold and 5-pole Chebyshev filter with 0.5 dB ripple
Figure 4.21

(a)

Transient response

(b)

$|e(t)| \text{ v. } t$

Transient point source response and error for 9 pole Chebyshev filter
Transient response

(a)

Transient point source response and error for 5 pole Chebyshev filter

(b)

Figure 4.22
that this error, which is the steady state error for the 5-pole filter, is acceptable, it can be used as a basis for the filter specification.

Using the limiting condition in equation 4.63, for the 5-pole filter,

\[ t_p = 8\tau + 8\tau = 16\tau \] \hspace{1cm} (4.64)

and for the 9-pole filter

\[ t_p = 32\tau + 8\tau = 40\tau \] \hspace{1cm} (4.65)

For a given pulse length, the sample interval \( \tau \), and consequently the filter cut-off frequency, \( f_c \), can be determined from equations 4.64 and 4.65. In the examples used,

\[ f_c = \frac{n-1}{2T} = \frac{n-1}{2n\tau} \]

with \( n = 8 \), for the 5-pole filter

\[ f_{c5} = \frac{7}{16} \times \frac{16}{t_p} = \frac{7}{t_p} \] \hspace{1cm} (4.66)

and for the 9-pole filter

\[ f_{c9} = \frac{7}{16} \times \frac{40}{t_p} = \frac{17.5}{t_p} \] \hspace{1cm} (4.67)

For equivalent performance, as measured above, there is more than a 2:1 ratio between the cut-off frequencies of the two filters. In the receiver using an analogue processor, a high gain amplifier is used to raise the level of the samples before filtering (as described in Chapter 5), and the effect of thermal noise originating in this amplifier is minimised
by making the noise equivalent bandwidth (29) of the reconstruction filter as narrow as possible.

The noise equivalent bandwidth of the filter depends on the rate of attenuation outside the pass-band as well as the nominal cut-off frequency. Comparing the two filters on this basis, if the cut-off frequencies for both are normalised to 1 \( (f_c = 1) \), the 9-pole filter has a narrower noise equivalent bandwidth than the 5-pole filter. However, the difference between the noise bandwidths is expected to be quite small, of the order of less than 10\%, and so scaling the frequency axes so that \( f_c = \frac{7}{t_p} \) and \( f_c = \frac{17.5}{t_p} \), the 5-pole filter still has a significant noise advantage over the 9-pole filter. Another positive advantage is that the 5-pole filter design is easier to implement than the 9-pole filter design.

The preceding argument could have been expressed more quantitatively if a figure for what was called an 'acceptable' error could have been given. The 'acceptability' of a specified error level can only be assessed by comparative testing of the display against that corresponding to a different error level in a known sonar environment. In the absence of such test results, an absolute figure for an acceptable error level cannot be defined with any confidence. In Chapter 6, an average measure of error is suggested for the comparison of the performance of the digital and analogue processors, and the filter specification is viewed in terms of the additional degradation of the intensity modulation waveform which is caused by the filter.

For a specified error contribution which is attributable to the filter, it should be possible to search for an optimum filter which has the minimum noise equivalent bandwidth associated with the given level of
error. For the moment, it is thought that the 5-pole Chebyshev filter will give an adequate reconstruction of the display intensity modulation waveform, and this filter is used in the prototype receiver.

4.6.2.2 Insertion of a Guard Band. The rate of attenuation of the reconstruction filter outside the pass-band can be reduced, and thus the transient response improved, if one or more of the spectral lines of $S(f)$ adjacent to the required band, $-(n-1)/(2T) \leq f \leq n-1/(2T)$, are made zero. This guard band can be inserted by using a DFT network which has more inputs, $n_N$, available than the number of elements, $n$, in the array. The first $n$ inputs to the network, $(k = 0, 1, ..., n-1)$, are connected to the array elements, and the remaining $n_N - n$ inputs, $(k = n, n+1, ..., n_N-1)$, are zero. All $n_N$ outputs of the network are used to form the sequence $s(t)$ to be filtered, but equation 4.27 which defines the phase shift in series with each output, needs to be modified.

It was shown in Chapter 3 that the FFT factorisation was most advantageous when the number of inputs to the DFT network was a power of 2. For this reason, although the technique is quite general, $n_N$ might well be a power of 2 and $n$ some other number less than $n_N$.

Taking the centre of the $n$ element array as the origin, the far field source distribution $R_1^*(s)$ is defined by equation 4.26 as

$$R_1^*(s) = Q_1^*(s) \exp(j\pi s(n - 1)L)$$  \hspace{1cm} -(4.26)$$

where

$$Q_1^*(s) = \sum_{k=0}^{n-1} q_1(kL) \exp(-j2\pi skL)$$  \hspace{1cm} -(2.29)$$
Now, from equation 2.32, the rth output of the DFT network of
dimension $n_N$ is

$$\begin{align*}
y_r &= \sum_{k=0}^{n_N-1} x_k \exp(-j2\pi rk/n_N) \\
    &= \sum_{k=0}^{n_N-1} x_k \exp(-j2\pi r_k/n_N)
\end{align*}$$

where $x_k \equiv q_1(kL)$.

Comparing equations 4.68 and 2.29

$$y_r = Q_1\left(\frac{r}{n_NL}\right)$$

since $x_k = 0$ for $k = n, n+1, \ldots, n_N-1$.

i.e. the $n_N$ outputs of the DFT network are samples of $Q_1(s)$ taken at
the points $s = \frac{r}{n_NL}$.

At the sample points, from equation 4.26

$$R_1^*\left(\frac{r}{n_NL}\right) = Q_1^*\left(\frac{r}{n_NL}\right) \exp\left[\frac{j\pi r(n-1)}{n_N}\right]$$

Equation 4.70 defines the phase shift to be inserted in series with
the output $r (r = 0, 1, \ldots, n_N-1)$ of the DFT network.

If the samples of the outputs of the network are translated now into
the time sequence $s^*(t)$ with sample spacing $\tau$, the spectrum $S^*(f)$
is given by

$$S^*(f) = S(f) \ast \frac{1}{\tau} \sum_{r=-\infty}^{\infty} \delta(f - \frac{r}{\tau})$$

where $\tau = \frac{T}{n_N}$. 
\[ \frac{1}{T} q_1(7L) = 0 \]

\[ \frac{1}{T} q_1(0) = \frac{1}{T} \]

\[ \frac{1}{T} = \frac{n_N}{J} \]

Figure 4.23

\( S^*(t) \) v. \( f \)

for 7-element array feeding a DFT processor of dimension 8

Figure 4.24

Schematic for 7 element array feeding DFT processor of dimension 8
$S^*(f)$ is periodic with period $\frac{1}{T} = \frac{n_N}{T}$.

Now, from equations 4.31 and 4.7

$$S\left(\frac{n-1}{2T} - \frac{k}{T}\right) = q_1(kL)$$

$$k = 0, 1, ..., n_N - 1$$

Since $q_1(kL) = 0$ for $n \leq k \leq n_N - 1$, $S(f)$ consists of only $n$ spectral lines ($k = 0, 1, ..., n-1$) with spacing $\frac{1}{T}$. This means that the baseband width of $S^*(f)$ is narrower than the period, $\frac{n_N}{T}$, and a gap has been introduced between adjacent bands in $S^*(f)$. This is illustrated in Figure 4.23 for a 7 element array feeding a 8 element DFT network. A single zero has been introduced between adjacent bands in $S^*(f)$ which enables the reconstruction filter rate of attenuation to be reduced. The schematic for this situation is shown in Figure 4.24.
CHAPTER 5

Receiver Design

5.1 Introduction

This chapter describes the design of two types of receiver for a pulse-modulated transmission which incorporate the ideas discussed in previous chapters. One receiver is based on the digital DFT processor, and works at a carrier frequency of 500 kHz; this system was subsequently tested using targets immersed in a water tank. The other receiver has an analogue DFT processor working at a carrier frequency of 32 kHz, and in the absence of a working transducer array the processor was tested with an artificially generated plane wave input.

A considerable amount of circuit design work was required during the construction of the receivers, but since established design principles were adhered to, a detailed description of many of the circuits is not included. A few circuits are described where it is felt that an improvement in performance is necessary, or where the circuit detail may be of help to further work.

5.2 A receiving system using a digital processor

A schematic diagram of the receiver is shown in Figure 5.1. Only the processing for a single channel is shown, but all channels are identical and are combined in front of the gain controlled amplifier. The function of each block is described now in more detail.
Figure 5.1

Receiver schematic using digital processor
5.2.1 Channel Amplifier

Each receiving transducer is connected to a low noise preamplifier, situated in the transducer head assembly, which drives the long connecting cable to the receiver equipment. The channel amplifier is inserted before the demodulator to provide gain and phase adjustment in each channel, so that the characteristics of the reconstruction filter can be equalised, and any additional aperture shading can be applied.

5.2.2 Demodulator and Low Pass Filter

Both the amplitudes and relative phases of the signals in each channel are required for the computation of the DFT. Equivalently, the real and imaginary components of the signal phasor must be determined. This is achieved by the two-path demodulator and filter shown in Figure 5.1. This scheme is used in the coherent detector described by Tucker and Griffiths (31).

After low pass filtering the output of the balanced demodulators to remove the sum frequency, the signals $x_{kx}(t)$ and $x_{ky}(t)$ in the real and imaginary component branches of channel $k$ are given by,

$$x_{kx}(t) \propto X_k \cos\{(\omega_r - \omega_o)t + \phi_k\}$$  \hspace{1cm} (5.1)

and

$$x_{ky}(t) \propto X_k \sin\{(\omega_r - \omega_o)t + \phi_k\}$$  \hspace{1cm} (5.2)

in response to the signal, having a phasor representation $x_k = X_k \exp\{j(\omega_r t + \phi_k)\}$, at the input to the demodulators. A scale factor depending on the gain of the circuit and the magnitude of the local oscillator has been omitted from equations 5.1 and 5.2.
In the absence of Doppler shift in the received signal \((\omega_r = \omega_o)\), \(x_{kx}(t)\) and \(x_{ky}(t)\) are proportional to the resolved components of \(x_k\) at \(t = 0\). If there is a Doppler shift in the carrier frequency of the received signal \((\omega_r \neq \omega_o)\), then an additional time varying component is present in \(x_{kx}(t)\) and \(x_{ky}(t)\). This is of no consequence since the relative phases of the signals in all transducer channels are preserved.

Equations 5.1 and 5.2 were obtained for a continuous sinusoidal input. Using a pulse modulated carrier the bandwidth of the filter following the demodulator should be of the order of the reciprocal of the pulse length\(^{(32)}\), allowing for Doppler shift as well, so as to maximize the signal to noise ratio at this point. Although, if the principal source of interference is reverberation\(^{(32)}\), the filter will have little effect in reducing this since reverberation looks like narrowband noise whose spectrum occupies approximately the same frequency band as the signal. The effect of noise on the performance of the receiver is considered in Chapter 6.

In the prototype receiver, an integrated circuit demodulator, MC1596G, is used in the configuration recommended in the data sheet. The low pass filter following each demodulator is a 2 section RC ladder network having a 3 dB cut-off frequency of 11 kHz. This simple filter was used because all channels could be matched to better than 1° without difficulty, and the attenuation at the sum frequency (1 MHz) is greater than 60 dB. The bandwidth was chosen to be large enough for subsequent variations in pulse length to be accommodated, although the results discussed in this thesis were obtained with a pulse duration of 400\(\mu\)s.

Tests on the demodulator, using a 500 kHz local oscillator input
Plate 5.1
Input to quadrature demodulators

Plate 5.2
Output of real component channel

Plate 5.3
Output of imaginary component channel

All Plates: vertical axis - 50 mV/cm
horizontal axis - 0.5 ms/cm
of amplitude 44mV (r.m.s.), and a variable amplitude c.w. signal input of frequency 501 kHz, gave the following results. At the difference frequency of 1 kHz, the circuit signal gain was typically 1.3. A gain control is included for fine adjustment. For output signal levels less than 200 mV, the maximum departure of the output v. input voltage characteristic from a straight line, was 2 mV (i.e. 1%).

Plate 5.1 shows a photograph of the signal appearing at the output of one of the transducer channel amplifiers. The large echo was reflected from a flat plate. Plates 5.2 and 5.3 show the corresponding outputs from the demodulator filters in the real and imaginary channels.

5.2.3 Track/Store and Multiplex Unit

The function of the track/store and multiplex unit in each channel is to sample simultaneously the signals in all channels, and then multiplex the samples sequentially to an amplifier preceding the analogue-to-digital (A/D) convertor.

The component parts of the unit are shown in more detail in Figure 5.2. The input CR network removes the d.c. level from the output of the demodulator, and the unity gain buffer amplifier provides a low impedance source for the track/store switch. This is a conventional F.E.T. series switch. In the tracking mode, the channel resistance is low (200 Ω) and the voltage across the sample capacitor tracks the signal with a circuit time constant of 0.2μs. At the sample instant, the switch is opened and the sample is stored on the capacitor. The gate switching waveform is coupled to the sample capacitor by the interelectrode capacitances of the F.E.T. and imparts a negative step (on the order of 10mV) to the held voltage. Subsequently the capacitor voltage is subject to change due to
Figure 5.2
Track/store and multiplex unit
leakage currents, principally through the switch and the input stage to
the following amplifier when a high quality capacitor is used. The M.O.S.
transistor input to the high impedance read-amplifier has a gate pro-
tection diode, and the leakage through this tends to cancel the leakage
through the gate source diode of the switch. The net result is a very
low drift rate of the capacitor voltage (as low as 5\mu V/100\mu s).

The drift rate of the capacitor voltage varies from one amplifier to
the next, but its effect can be compensated by the offset control on the
read-amplifier. This is because the net leakage current has a constant
current nature (being due to reverse biassed diodes), and the time delay
between the sampling instant and the instant that the voltage is digitised
by the A/D convertor, is constant for any one channel. This offset control
also compensates for the switch waveform breakthrough already described,
and any other constant d.c. offset present in each channel. The leakage
currents are very temperature sensitive, but, once again, they tend to
track each other, although this may be a problem if the operating tem-
perature range is wide.

The most serious deficiency of the circuit is in the read-amplifier.
The transistors, type ML 102, were chosen because they are available as a
matched pair on a common substrate, and so their characteristics should
track well with temperature. However, the output of the amplifier was
found to drift slowly over a range of 1 mV in less than one hour. The
performance of integrated circuit amplifiers with F.E.T. inputs is ex-
pected to be better than this, but these are relatively expensive at the
present.

The multiplexer switch is another series connected F.E.T. The time
during which this switch is closed (low resistance) is determined by the performance of the A/D convertor. In this case, at least 16\(\mu\)s was required for a digitisation and store cycle, so 20\(\mu\)s was chosen for the multiplex time. The total time taken to digitise all 16 samples (2 from each transducer channel) is 320\(\mu\)s, and this governed the choice of 400\(\mu\)s as the transmitted pulse length.

5.2.4 The A.C.C. Amplifier

The received signal strength, including reverberation, will vary over a wide range in an underwater sonar system for a number of reasons: range spreading loss, attenuation in the medium, variability of target strength, variability in back-scattering properties of the medium and its equivalent target strength as the volume insonified increases with range. The function of the a.g.c. amplifier is to utilise the fixed input range of the A/D convertor effectively; both to prevent saturation by too high a level of input signal, and to ensure that the signal level is not so low that quantization effects and subsequent processing error cause excessive degradation of the signal.

It has been suggested by Vural,\(^{(33)}\), in a study of quantization effects on the performance of a conventional summing array, that the amplifier gain is controlled by the noise level in the received signal. In the simplest case, the control voltage is derived from the noise power in a single channel. Kay,\(^{(1)}\) discusses the use of reverberation level to control amplifier gain in a sector scan receiver in order to match the signal to the limited dynamic range of the intensity modulated display. The receiver described here uses a similar display, and controlling the signal level to suit the A/D convertor in the first instance should avoid
the need for further control to suit the display.

A single amplifier is used, and all samples are multiplexed through it, to avoid the problem of closely matching separate variable gain amplifiers in each channel. The gain of the amplifier varies during the range sweep, but it should remain constant during the time that the set of samples from one range cell is being multiplexed through the amplifier - this implies the use of a sample and hold circuit on the gain control voltage. The bandwidth of the amplifier should extend from zero frequency up to as high a frequency as necessary to allow each amplified sample to 'settle' before being converted to digital form.

In the present test equipment, a fixed gain amplifier was used because of the limited test environment available; firstly, a variable gain was not necessary, and secondly, its function could not have been tested adequately. However, an initial survey was made of the problems that might be encountered in the design of a wideband, direct-coupled, variable gain amplifier. The most obvious difficulty arises in preventing the gain control mechanism causing a change in the d.c. offset level at the output of the amplifier. This, ideally, should be zero over the entire range of gain variation.

5.2.5 The Digital Processor

The idea of using an on-line general purpose digital computer to perform the DFT of the field samples, using the FFT algorithm, has been discussed by Griffiths and Hudson\(^{(34)}\). It was rejected on the grounds that the computation time, for an unspecified computer, was too long for short-pulse high resolution systems. More detailed calculations, based on the figures contained in Table 3.2, and using the CTL Modular One
computer lead to a similar conclusion. This particular computer was chosen because of its fast multiply time (3.1μs). Even so, the minimum computation time for the FFT for an 8-element array is 440μs, and for a 32-element array the computation takes 3.64 ms. Bearing in mind the high cost of this equipment, the time spent on array processing should be less than 10%, say, of the total workload of the computer (as a very approximate order of magnitude), and this obviously limits its use to long pulse durations.

For short pulse lengths, the only solution is to use a special purpose processor which is dedicated to computing the FFT of the array samples. An increase in speed by a factor of 10 should not be difficult to achieve, using a fast semiconductor store. Further reductions in computing time can be gained by using more than one arithmetic unit\(^{26}\).

The objective is to reduce the computation time to less than the duration of the transmitted pulse, otherwise the processor would present a bottleneck to the flow of data. On this basis, the computation time would not be the limiting factor in assessing the feasibility of using a digital processor in the fish-finding type of sonar being considered.

The detailed hardware design of a digital processor was outside the scope of this investigation. However, it is possible to study the performance of such a processor, having a limited word length, by using a general purpose machine with a suitable stored program. The Modular-One computer was used on-line, and the problem of the computation time being longer than the 400μs pulse duration was overcome by reading all the array samples, from the 8 range cells contained within the tank, directly into the store. In the time remaining before the next transmitted pulse, the FFT of the sample set from each range cell was computed.
Figure 5.3

In-place computation
in turn and a B-scope display generated. The result, using this quasi real-time approach, is identical to that which would be obtained from a strictly real-time processor.

One further point may be made about the use of the FFT algorithm. If the order of the input samples is shuffled in a predetermined manner, then all subsequent computations can take place in such a way as to minimize the amount of intermediate storage required. That is, if any pair of samples take part in the basic sum and difference operation, then the outputs of the operation may be stored in the locations originally occupied by the inputs. This 'in place' property of the FFT is well known (10), and it may be demonstrated by rearranging the schematic in Figure 3.5 to give Figure 5.3. The inputs and outputs of any sum and difference block go to the same locations in store.

After the FFT operation, each complex coefficient so formed is multiplied by a complex number, within the digital processor. This compensates for the finite transducer size, and also shifts the spectrum of the set, regarded as a time sequence, so that it is centred at zero frequency.

5.2.6 The Interpolating Filter and Detector

The single channel detector scheme, described in section 4.5.1, was used to generate the display intensity modulation waveform in the prototype receiver. One cycle, $s_1(t)$, of the sequence $s^*(t)$ is applied to the interpolating filter, and a decision is made at the end of the first half-cycle on which one of the real and imaginary channel outputs is to be used to intensity modulate the display - the component having the larger peak amplitude being chosen. The display period then occupies the
Figure 5.4(a)
Filter and Detector

Real D/A Conversion

\[ S/H_1 \]

Imag. D/A Conversion

\[ S/H_2 \]

Real channel

\[ \text{peak det.} \]

\[ \text{imaginary channel} \]

\[ \text{peak det. comparator} \]

\[ \text{o/p} \]

Figure 5.4(b)
Timing Waveforms for D/A conversion and channel delay

\[ s_1(t) \]

Buffer amp. gain

Display channel decision

Display enable

Filter enable

Display period

Figure 5.5
Timing waveforms for filter and detector
second half-cycle of $s^*_1(t)$.

The real and imaginary components of the complex sample sequence, $s^*_1(t)$, must be applied simultaneously to the two channels of the interpolating filter. If a single digital-to-analogue (D/A) convertor is used, output samples can only be generated sequentially, and an additional delay must be inserted in one channel to regain synchronism between the real and imaginary components.

The schematic diagram for the filter and detector is shown in Figure 5.4(a). The switch(F.E.T.) and capacitor combinations, $S/H_1$, $S/H_2$, and $S/H_3$, give a zero-order hold characteristic. The waveforms driving the switches are $SW_1$, $SW_2$, and $SW_3$ respectively, shown in Figure 5.4(b), with the 'high' level indicating a switch closed condition. At the start of an interpolation cycle, ($t = 0$), the inputs to the filters in the real and imaginary channels are zero. This condition is ensured by shorting the capacitors in $S/H_2$ and $S/H_3$ at the end of the preceding sequence. The D/A conversion of the real component of an output sample from the processor is initiated and completed in the time interval $t_1$ (Figure 5.4(b)). The output of the convertor is sampled and the voltage held by unit $S/H_1$ during interval $t_2$. The D/A convertor then produces the imaginary component of the complex valued output of the processor during interval $t_3$. At the start of interval $t_4$, the imaginary component at the output of the convertor is sampled and appears at the input to the filter in the imaginary channel; the real component is transferred from $S/H_1$ to $S/H_2$ and also appears at the input to the filter in the real channel. $S/H_1$ can now be used to store the real component of the next output from the processor and the conversion cycle repeats.
The time, \( T \), for which one output sample is held at the input to the filter is 40\( \mu \)s in the prototype receiver. The channel filters are the 5-pole Chebyshev type with 0.5 dB ripple, whose performance was examined in section 4.6.2.1, and the cutoff frequency is \( f_c = \frac{n-1}{2\pi T} \), i.e. \( f_c = 10.93 \) kHz. The filter design and component values were taken from Tables (30). The full wave rectifier in Figure 5.4(a) is an operational amplifier circuit (35) giving good rectification of input signals as low as a few millivolts.

Now \( s^*(t) \) is one cycle of the periodic waveform \( s^*(t) \), and so the particular output of the processor which is used as the start of the sequence can be chosen as a matter of convenience. Referring to Figure 4.22(a), the compensated 5-pole Chebyshev filter gives a delay of approximately \( 2\pi T \), so if output 6 is chosen to start the sequence, the interpolated response to a plane wave incident normal to the array will reach a maximum amplitude approximately half-way through the second half of the input sequence \( s_1(t) \). Consequently, the 'bright-up' for a target normal to the array will be in the centre of the screen.

Remembering that \( s^*(t) \) consists of two cycles of the output samples of the processor with one cycle inverted, the inversion must always commence with the first output of the processor, \( s_0 \), even though this may not be the first sample in the sequence to be interpolated. The inversion is accomplished by making the buffer amplifier, preceding the filter, with a switchable gain of \( \pm 1 \). The principal timing waveforms in the interpolation cycle are shown in Figure 5.5.

The time duration of the interpolation cycle is approximately \( 2\pi T \), in this case 640\( \mu \)s. Using the single channel detector, it is preferable to allow the filter output time to decay sufficiently so as not to inter-
fere with the display waveform decision process associated with the following interpolation and display sequence. Consequently, a total of 800µs was allowed for the generation of the display waveform for a sector scan. This was acceptable in the prototype receiver because of the quasi real-time processing scheme which had to be adopted. However, in a strictly real-time application, the interpolation cycle time could be reduced by as much as a factor of 10 by using a high speed D/A convertor and a faster sampling switch in the zero-order hold. This means that the interpolation cycle time for a 16-element processor could be as low as 200µs.

5.2.7 Control Unit

The control unit causes all of the operations involved in generating a sector scan display to take place in the correct order. Each scan sequence is then repeated in every range cell within the working range of the sonar set. The detailed design of this unit will not be discussed, but since the feasibility of using this type of receiver with short pulse lengths depends upon whether a sector scan display can be generated at least once within the pulse length, a general point is worth making.

The operations taking place within the receiver can be divided into three distinct phases,

(i) sampling, A/D conversion, and storage of data for input to the DFT processor,

(ii) the DFT operation on the stored data, and

(iii) the generation of a sector scan display from the DFT processor outputs.
filter and detector

track/store and multiplex

control unit

Tx modulator

channel amplifiers

demodulators

500 kHz oscillator
The minimum pulse length that can be used with the receiver is governed not by the sum of the time intervals required for each of (i), (ii), and (iii) above, but only by the longest of the individual time intervals if buffer storage is made available. In this situation, all three phases can be operating concurrently on data from three different range cells. Using this 'conveyor belt' technique, there will be an initial time lag corresponding to two range cells whilst the system fills up with data, but this is immaterial.

The time limitation will be caused by either phase (i) or phase (iii). In phase (i), all $2n$ samples must pass through the a.g.c. amplifier and A/D convertor. In phase (iii) a total of $4n$ samples (2 cycles) must pass through the D/A convertor, or at best $2n$ samples through two D/A convertors - one for the real component channel and one for the imaginary component channel. On this basis, a 200μs pulse length used in conjunction with a 16-element array is technically possible. In phase (ii), the DFT operation can be performed well within the shortest pulse length likely to be required, just by increasing the number of arithmetic units - although at the present time this will obviously incur an economic penalty.

Plate 5.4 shows a rear view of the complete receiver.

5.3 A Receiving System Using an Analogue Processor

A schematic diagram of the receiver, showing only one channel, is shown in Figure 5.6. Apart from the analogue FFT processor, all other units in the diagram operate in an identical manner to their counterparts in the receiver described in section 5.2, and so they will not be described again. Using a carrier frequency of 32 kHz, the transmitted pulse
Figure 5.6

Receiver schematic using analogue processor
Figure 5.7
The sum and difference unit

(a)
Phase retard unit

(b)
Phase advance unit

Figure 5.8
length is unlikely to be less than 500μs, because of transducer Q limitation. Also the pulse must be sufficiently long for the aperture to be fully covered. The interpolating scheme suggested should easily cope with the display waveform for a 16-element array in this time.

The main point of difference between the two systems is that the analogue processor operates at the carrier frequency, and so the order of demodulation and multiple-beam formation is reversed.

5.3.1 The Analogue Processor

The basic sum and difference circuit is shown in Figure 5.7 and the phase retard and advance circuits are shown in Figure 5.8. The values of R and C in the phase shift circuits are determined from equations 6.71(b) and 6.72(b). These circuits were interconnected according to the schematic diagram in Figure 3.16, but a slight modification was incorporated at the outputs of the processor. Since only the relative phases of the outputs are important in the interpolation process, it is permissible to apply a phase shift of π radians to all outputs, effectively moving the phase shift circuits from outputs 0, 3, 5, and 6 to outputs 1, 2, 4, and 7. This is done so that when the (n-1) additional phase shifts required for the real interpolation process (from equation 4.27) are added in series with the outputs, then no phase shift is required in series with output 0 - a small economy. The phase shift, arg(C_r), associated with each output of the processor for an 8-element array is listed in Table 5.1.
Plate 5.5

Analogue FFT processor for 8-element array
Table 5.1
Phase Compensation at output of processor

<table>
<thead>
<tr>
<th>Output r</th>
<th>arg(C_r) (radians)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>-π/8</td>
</tr>
<tr>
<td>2</td>
<td>3π/4</td>
</tr>
<tr>
<td>3</td>
<td>5π/8</td>
</tr>
<tr>
<td>4</td>
<td>π/2</td>
</tr>
<tr>
<td>5</td>
<td>3π/8</td>
</tr>
<tr>
<td>6</td>
<td>-3π/4</td>
</tr>
<tr>
<td>7</td>
<td>-7π/8</td>
</tr>
</tbody>
</table>

Using discrete components, three circuit boards were required for the 8-element processor, and these are shown in Plate 5.5. An assembly using thin or thick film hybrid modules would be expected to require less board area than this.

Each circuit within the processor adds noise to the signal, and it is of interest to calculate the effect of the internal noise sources at the outputs of the processor. Taking a single summing circuit for example, with both inputs at zero, the noise voltage appearing at the output is due to noise sources within the amplifier (35), and noise sources associated with the external resistors. Figure 5.9(a) shows the calculated typical value of the output noise voltage spectral density, \( G_o(f) \), in the range 2 kHz < \( f \) < 40 kHz, and also values measured for a single unit. In this range, \( G_o(f) = n_1 \), a constant.

In the analogue processor, with \( n = 2^m \), there are \( m \) ranks of Σ/Δ units. In the \( m \)th rank, the noise voltage spectral density, \( n_m \),
Output noise voltage spectral density, $G_o(f)$ v. $f$, for a summing unit

Input referred noise voltage spectral density, $G_r(f)$ v. $f$, for 8-element analogue processor

Figure 5.9
at output 0 of the processor consists of a contribution, \( \eta_1 \), from the \( \Sigma \) unit in the mth rank, and a contribution, \( \eta_{m-1} \), from each of two \( \Sigma \) units in the \((m-1)\)th rank.

i.e. \( \eta_m = \eta_1 + 2\eta_{m-1} \) \hspace{1cm} - (5.3)

Working backwards to the inputs of the network, equation 5.3 becomes

\[ \eta_m = \eta_1 + 2\eta_1 + 2\eta_{m-2} \] \hspace{1cm} - (5.4)

Repeating this procedure leads to a finite sum

\[ \eta_m = \eta_1(1 + 2 + 2^2 + \ldots + 2^{m-1}) \]

i.e. \( \eta_m = \eta_1(2^m - 1) \)

i.e. \( \eta_m = \eta_1(n - 1) \) \hspace{1cm} - (5.5)

The expressions for noise voltage spectral density in the other outputs of the processor will be different from equation 5.5, and will involve the noise voltage spectral densities associated with difference units and phase shift units. However, the difference is sufficiently small for equation 5.5 to be taken as an approximate guide to the noise voltage spectral densities to be expected at all outputs.

This value of \( \eta_m \) would have been obtained by connecting independent noise voltage sources, each having a spectral density \( \eta_{rn} = \eta_m/n \) in each input channel of an ideal noiseless processor. From equation 5.5,

\[ \eta_{rn} = \eta_1(1 - \frac{1}{n}) \] \hspace{1cm} - (5.6)
Equation 5.6 provides a useful rule of thumb for noise calculations involving large networks containing internal source, since for a single unit, \( n = 2 \) and \( \eta_{r2} = \eta_1/2 \); and for a network with \( n \geq 8 \), \( \eta_{rn} \approx \eta_1 = 2\eta_{r2} \). So, the input referred noise voltage spectral density for a large network is approximately twice that for a single \( \Sigma/\Delta \) unit.

Figure 5.9(b) shows the calculated and measured values of the input referred noise voltage spectral density, \( G_r(f) \), in the range

\[ 2 \text{kHz} < f < 40 \text{kHz} \]

for the 8-element analogue processor which was built.

From Figure 5.9(b), the input referred noise voltage is approximately 1\( \mu \text{V r.m.s.} \) in a 1 kHz band.
CHAPTER 6

The Estimation of Receiver Performance

6.1 Introduction

A receiver, of the type described in Chapter 5, will suffer from deficiencies due to the practical limitations of the circuits which are used. However, because of the noisy nature of the signal being processed, these deficiencies in the receiver are not necessarily greatly detrimental to its performance. In this chapter, an estimate is made of the performance of practical digital and analogue processors in comparison with an ideal processor. This enables certain design parameters to be specified; for example, the word length and the a.g.c. amplifier characteristic for the digital processor, and the tolerance of the components used in the analogue processor.

Since the performance of the receiver is judged ultimately by the display presented to the operator, it could be argued that an analysis should include the results of extensive trials of the different processors by a number of operators, since their individual subjective opinions are bound to be different. In the absence of such experimental data, the performance of a processor is assessed from the reconstructed far field source distribution waveform which is used to intensity modulate the display.

In the development of the system so far, the reconstructed far field source distribution, generated as a time waveform $s(t)$, has been defined as the response of an ideal receiver (ideal beam-forming network and ideal reconstruction filter) to samples of the field at the
aperture; and this field is the resultant of the fields due to point sources sufficiently far away from the aperture for the wavefronts to be considered plane. In a practical system, the output, v(t), of the reconstruction filter will be different from s(t) for the following reasons: the signal inputs to the receiver will be contaminated by noise, and the non-ideal receiver will contribute additional noise as well as systematic errors.

The sources of noise arising from within the medium are various. Isotropic noise may be assumed to be uncorrelated between channels, but reverberation signals in each channel are correlated to an extent which depends on the design of the transmitting and receiving arrays. Noise also arises in the receiving transducers and channel amplifiers; this is uncorrelated between channels, and is not usually dominant in well designed sonar receivers.

In the digital processor, it is possible to saturate the A/D converter on a high input signal level with the a.g.c. amplifier gain controlled by the noise level in one channel. This is equivalent to clipping the signal and introduces a saturation error in addition to the quantization error which accompanies the A/D conversion process. The use of a finite length word in the processor means that the stored multiplier constants are necessarily inaccurate, and also the results of all operations have to be scaled to remain within the word length, and this introduces a rounding error.

In the analogue processor, the circuit elements which are used do not perform the ideal operations of summing, differencing, and phase shifting. Errors are introduced by the non-ideal operational amplifiers
and the passive components which are specified within a certain tolerance band.

The behaviour of the analogue and digital processors is examined, in the first instance, under the assumption that a perfect low pass filter is available for the reconstruction of \( s(t) \). This reveals the error contribution due to the processors alone. The degradation of the waveform due to the use of a filter having finite attenuation outside the frequency band of interest is then assessed, together with the effect of using only the first one or two cycles of the periodic reconstruction.

In this initial study, which is aimed at identifying the importance of different system parameters, it will be assumed that the noise inputs to the receiver are uncorrelated between channels. Even though this eases the problems of analysis, a digital computer simulation has had to be used to obtain many of the results. Sufficient insight is gained for a 'sea-going' receiver to be designed, and the need for additional calculations taking into account interchannel noise correlation can be more readily assessed during trials of the equipment.

6.2 Measures of Performance

Two different ways of describing the behaviour of a sector scan receiver will be examined in this section. The first approach relies on an average measure of the difference between \( v(t) \) and \( s(t) \). In the second approach, the variation of one or more parameters associated with beamforming is studied; for example, the direction of the main lobe maximum, the 3 dB beamwidth, and the level of the sidelobes.

Since the performance of the DFT processor is independent of the
reconstruction filter characteristic, it is convenient to examine the
behaviour of the DFT processor alone by assuming that an ideal low pass
filter is available for the reconstruction of $s(t)$. The effect of a
practical filter on the receiver performance will be investigated in
section 6.6.

6.2.1 Mean Square Error

One measure of the difference between the two functions, $v(t)$, and
$s(t)$, is given by

$$
\varepsilon_1 = \frac{1}{2T} \int_0^{2T} |v(t) - s(t)|^2 \, dt
$$  \hspace{1cm} (6.1)

It is assumed that there are an even number of array elements, and con-
sequently the period of $v(t)$ and $s(t)$ is $2T$.

If the sample value at output $r$ of the DFT network is $u_r$, then
generally

$$
u_r = s_r + e_r
$$  \hspace{1cm} (6.2)

where $s_r$ is a sample of $s(t)$, and $e_r$ is an error component whose
source is unspecified as yet. The infinite periodic time sequence $u_r^*(t)$,
constructed as described in Chapter 4, will have $2n$ samples/cycle.

$$
u_r^*(t) = \sum_{r=-\infty}^{\infty} u_r \delta(t - rT)
$$  \hspace{1cm} (6.3)

where $u_r = u_{r-n} = u_{r-2n}$.
From equation 6.2
\[ u^*(t) = \sum_{r=-\infty}^{\infty} (s_r \delta(t - rT) + e_r \delta(t - rT)) \] \hspace{1cm} (6.4)

i.e. \( u^*(t) = s^*(t) + e^*(t) \) \hspace{1cm} (6.5a)

and \( U^*(f) = S^*(f) + E^*(f) \) \hspace{1cm} (6.5b)

The output of an ideal low pass filter, whose frequency response is defined by equation 4.35, in response to \( u^*(t) \) is given by
\[ v(t) = u(t) \] \hspace{1cm} (6.6)

\[ = s(t) + e(t) \] \hspace{1cm} (6.7)

\[ \therefore \text{from equation 6.1} \]
\[ \epsilon_1 = \frac{1}{2T} \int_{-T}^{T} |e(t)|^2 \, dt \] \hspace{1cm} (6.8)

It is shown in Appendix E that \( \epsilon_1 \) is related to the samples \( e_r \) (\( r = 0, 1, \ldots, n-1 \)) at the output of the DFT network by the equation
\[ \epsilon_1 = \frac{1}{n} \sum_{r=0}^{n-1} |e_r|^2 \] \hspace{1cm} (6.9)

A measure of the degradation suffered by the wanted signal is obtained by defining a signal/error power ratio.

The mean square value, \( S_o \), of the signal in the reconstructed waveform is given by
\[ S_o = \frac{1}{2T} \int_{-T}^{T} |s(t)|^2 \, dt \] \hspace{1cm} (6.10)
by changing the variables in equations 6.8 and 6.9. Combining equations 6.9 and 6.11

\[ S_0 = \frac{1}{n} \sum_{r=0}^{n-1} |s_r|^2 \]  \hspace{1cm} (6.11)

\[ \frac{S_0}{\varepsilon_1} = \frac{\sum_{r=0}^{n-1} |s_r|^2}{\sum_{r=0}^{n-1} |e_r|^2} \]  \hspace{1cm} (6.12)

\( \varepsilon_1 \) provides a measure of the difference between \( v(t) \) and \( s(t) \) directly after the reconstruction filter. The modulus of the reconstructed waveform is used to intensity modulate the display, and another error measure, \( \varepsilon_2 \), can be defined as

\[ \varepsilon_2 = \frac{1}{T} \int_0^T \left( |v(t)| - |s(t)| \right)^2 dt \]  \hspace{1cm} (6.13)

The interval of integration is \( T \) since \( |v(t)| \) and \( |s(t)| \) are periodic with period \( T \) because of the half wave symmetry of \( v(t) \) and \( s(t) \). The signal/error power ratio is given by \( S_0/\varepsilon_2 \), where \( S_0 \) is defined in equation 6.11.

6.2.2 Beam-Profile Characteristics

The directional behaviour of a transmitting or receiving antenna is usually described in terms of a measure of the main-lobe width and the level of the sidelobes in the radiation pattern. The design problem is then to produce an antenna in which these parameters do not exceed the levels which have been determined by a particular operational requirement. In the case of an electronically steered array, the angular steering
accuracy of the main lobe is another parameter which needs to be considered.

All of these performance criteria can be applied to the waveform which is used to intensity modulate the display for the sonar processor. A convenient analytical approach is to express the output, $v(t)$, of the ideal reconstruction filter as a Fourier series. This is done in Appendix F with the result that $v(t)$ may be written as a polynomial in a complex variable, $z_t$, where $z_t = \exp(j\frac{2\pi t}{T})$. The intensity modulation waveform is thus,

$$|v(t)| = t \left| c_0 + c_1 z_t^{-1} + c_2 z_t^{-2} + \ldots + c_{n-1} z_t^{-(n-1)} \right| - (6.14)$$

where $c_r = U(k)(\frac{n-1}{2T} - \frac{k}{T})$, a Fourier series coefficient.

Algebraic techniques can be applied to the polynomial within the modulus bars in equation 6.14 in order to find the maxima of $|v(t)|$. Evaluation of $|v(t)|$ at the positions of the maxima gives the main-lobe and side-lobe levels, and the position of the main-lobe is also defined. The 3 dB points on the main-lobe, where $|v(t)|^2 = 0.5|v(t)|^2_{\text{max}}$, can be found subsequently and hence the 3 dB bandwidth.

So, the response of the processor to a plane wave incident on the aperture can be found in terms of the usual beam-pattern parameters. The effect on these parameters of changing the word length or component tolerance in the two processors can be investigated numerically.

An interesting parallel can be drawn with Schelkunoff's work\(^{(37)}\) on the representation of the space factor of a linear summing array by a polynomial. Taking a linear array of $n$ equispaced point receivers
whose outputs are combined in a single summing junction, the output of the array in response to a plane wave approaching from any direction, \( s \), can be expressed as

\[
\Phi(s) = \sum_{k=0}^{n-1} B_k \exp(-j2\pi ks_1 L) z_s^k - (6.15)
\]

where \( z_s = \exp(j2\pi s_1 L) \); \( B_k \) is the amplitude weighting of element \( k \), and \( \exp(-j2\pi ks_1 L) \) is a phase shift in series with element \( k \) which steers the main-lobe in a direction \( s_1 \). In Schelkunoff’s paper, the steering phase shift is included in \( z_s \), but it is convenient to separate the terms here.

Now, from Appendix F, the response of the DFT processor to the same wavefront is

\[
v(t) = \exp\left\{j\pi t \frac{(n-1)}{T}\right\} \cdot T \sum_{k=0}^{n-1} U^\ast\left(\frac{n-1}{2T} - \frac{k}{T}\right) z_t^{-k} - (6.16)
\]

In the case of an ideal processor,

\[
U^\ast\left(\frac{n-1}{2T} - \frac{k}{T}\right) = S^\ast\left(\frac{n-1}{2T} - \frac{k}{T}\right) = u_1(kL) - (6.17)
\]

from equation 4.49.

Including the amplitude weighting factors, equation 6.16 becomes

\[
v(t) = \exp\left\{j\pi t \frac{(n-1)}{T}\right\} \cdot \sum_{k=0}^{n-1} B_k q(kL) z_t^{-k} - (6.18)
\]

In this case, \( v(t) = s(t) \), and using the equivalence of the time domain with the \( s \) domain, equation 6.18 becomes
\[ R^*_1(s) = \exp\{j\pi s(n-1)L\} \sum_{k=0}^{n-1} B_k q_1(kL) z_s^{-k} \quad - (6.19) \]

or, using equation 4.26

\[ Q^*_1(s) = \sum_{k=0}^{n-1} B_k q_1(kL) z_s^{-k} \quad - (6.20) \]

Equation 6.15 shows that the coefficients, of the polynomial representing the output of a linear summing array, are the complex weights associated with each element for controlling the beamshape and direction. Equation 6.20 shows that the far field source distribution, which is the output of a DFT processor, may be represented by a similar polynomial, except that the coefficients are now the field samples in the aperture, and the variable \( z_s \) is raised to a negative power.

6.2.3 Comparison of Performance Criteria

Two essentially different methods of measuring the performance of a DFT processor have been described in sections 6.2.1 and 6.2.2. On the one hand, the mean square error provides a single piece of information which cannot describe how the error varies in the interval over which the mean is taken. On the other hand, more detailed measures of the error - such as the variations in sidelobe levels, beamwidth, and direction indication - are necessarily more difficult to formulate, and interpretation is difficult when many reconstructions are to be compared. For instance, one reconstruction may offer a narrower main-lobe width, in response to a plane wave, but a higher side-lobe level than another reconstruction. Which is the better? This may depend on whether the direction indication is more or less accurate in one of them. This
qualitative argument is intended only to illustrate the difficulty in interpreting several pieces of information which need to be considered together. At least it can be asserted that one reconstruction is better overall if its associated mean square error is smaller than that of another reconstruction. For this reason, the mean square error is used subsequently to compare the performance of practical analogue and digital processors with an ideal processor. As it happens, the mean square error can be predicted quite easily in some situations, otherwise recourse is made to a numerical simulation of the processor.

6.3 The Performance of an Ideal Processor

Since the 'goodness' of a practical processor is to be judged by comparison with an ideal processor, this section is concerned with the behaviour of an ideal processor in response to a plane wave sinusoidal signal plus noise.

6.3.1 The Noise Model

It will be assumed in subsequent analysis that the receiving array is operating in a field giving a uniform power distribution across the aperture. The noise appearing at each input to the DFT processor, after filtering by the transducer and channel amplifier, can be represented by a narrow-band Gaussian random process with variance $\sigma_n^2$, and it will be assumed to be uncorrelated between channels. This model is a simplification of the noise characteristics to be met in many practical situations, and future work could include an examination of the additional effect of interchannel correlation; this is characteristic of reverberation, and the extent of the correlation depends on the design of the transmitting and receiving arrays. Correlation is also caused by
mutual acoustic coupling between array elements (38).

Each input channel has an effective bandwidth of 2B Hz centred on a frequency \( f_0 = \frac{\omega_0}{2\pi} \). The noise waveform, \( \rho(t) \), in each channel can be expressed as (29)

\[ \rho(t) = \rho_x(t) \cos \omega_0 t + \rho_y(t) \sin \omega_0 t \]  

where \( \rho_x(t) \) and \( \rho_y(t) \) are independent Gaussian random processes band-limited to B Hz.

Alternatively, from equation 6.21, \( \rho(t) \) may be written as

\[ \rho(t) = P(t) \cos(\omega_0 t - \zeta(t)) \]  

where \( P(t) = \sqrt{\rho_x^2(t) + \rho_y^2(t)} \)

\[ \zeta(t) = \tan^{-1} \frac{\rho_y(t)}{\rho_x(t)} \]

i.e. \( \rho_x(t) = P(t) \cos \zeta(t) \)  

and \( \rho_y(t) = P(t) \sin \zeta(t) \)

Also \( \rho^*(t) = \rho_x^*(t) = \rho_y^*(t) = \sigma_n^2 \)  

The bar indicates a time average.

Now, the complex samples at the output of the DFT network are the result of the DFT operation on the complex samples at the input to the network. A complex model, \( \psi_k(t) \), of the noise in input channel \( k \) is constructed in the same way as the complex model of the sinusoidal field
in Chapter 2, viz:

\[ v_k(t) = P_k(t) \exp\{j(\omega_0 t - \zeta_k(t))\} \]  

where \( \rho_k(t) = \text{Re}\{v_k(t)\} \)

\[ v_k(t) = \exp(j\omega_0 t) \cdot P_k(t) \exp(-j\zeta_k(t)) \]

The display waveform is generated from samples taken at a particular instant in time, and dropping the term \( \exp j\omega_0 t \), as in Chapter 2, the noise sample can be expressed as

\[ v_k = P_k \exp(-j\zeta_k) \]  

\[ v_k = v_{kx} + jv_{ky} \]

where \( v_{kx} = P_k \cos\zeta_k \)  

and \( v_{ky} = -P_k \sin\zeta_k \)

Comparing expressions 6.24 and 6.29, the real and imaginary parts of the complex sample correspond to samples taken from the uncorrelated processes \( \rho_x(t) \) and \( \rho_y(t) \).

It should be noted that the outputs of the real and imaginary channels of the demodulator, described in Chapter 5, in response to a noise input defined by equation 6.22, are proportional to \( P(t) \cos\zeta(t) \) and \( -P(t) \sin\zeta(t) \) respectively. Samples of the two channels taken at the same instant correspond then to \( v_{kx} \) and \( v_{ky} \).
6.3.2 The Performance before the Detector

The mean square value, $c_1$, of the error waveform before the detector, and the mean square value, $S_o$, of the signal waveform at the same point, can be calculated simply in terms of the input samples for the ideal processor.

The output, $v(t)$, of the reconstruction filter can be expressed as the Fourier series,

$$v(t) = \tau \sum_{k=0}^{n-1} U^* \left( \frac{n-1}{2T} - \frac{k}{T} \right) \exp \left\{ j2\pi t \left( \frac{n-1}{2T} - \frac{k}{T} \right) \right\}$$  \hspace{1cm} (6.30)

Equation 6.30 is derived from equation G.6 in Appendix G.

$$\frac{1}{2T} \int_0^{2T} |v(t)|^2 \, dt = \tau \sum_{k=0}^{n-1} U^* \left( \frac{n-1}{2T} - \frac{k}{T} \right) |^2$$  \hspace{1cm} (6.31)

Now, from equation 6.5b and equation 4.49, for the signal alone,

$$U^* \left( \frac{n-1}{2T} - \frac{k}{T} \right) = S^* \left( \frac{n-1}{2T} - \frac{k}{T} \right)$$

$$= q_1(kL) \cdot \tau$$  \hspace{1cm} (6.32)

and $v(t) = s(t)$.

Defining $q_1(kL) = q_k$ for convenience, equation 6.31 becomes

$$\frac{1}{2T} \int_0^{2T} |s(t)|^2 \, dt = S_o$$

$$= \sum_{k=0}^{n-1} |q_k|^2$$  \hspace{1cm} (6.33)
Similarly, for noise alone,

\[ u^*(\frac{n-1}{2T} - \frac{k}{T}) = E^*(\frac{n-1}{2T} - \frac{k}{T}) = \frac{v_k}{t} \]  \hspace{1cm} (6.34)

and \( v(t) = e(t) \).

\[ \frac{1}{2T} \int_0^{2T} |e(t)|^2 \, dt = \varepsilon_1 \]

\[ = \sum_{k=0}^{n-1} |v_k|^2 \]  \hspace{1cm} (6.35)

Considering the generation of the display waveform corresponding to a particular range cell, \( \varepsilon_1 \) will fluctuate from one range sweep to the next, and a quantity \( \bar{\varepsilon}_1 \) can be defined as the mean value of \( \varepsilon_1 \) over many reconstructions.

\[ \varepsilon_1 = \sum_{k=0}^{n-1} |v_k|^2 \]

\[ = \sum_{k=0}^{n-1} v_k^2 \]  \hspace{1cm} (6.36)

From equation 6.28,

\[ |v_k|^2 = v_{kx}^2 + v_{ky}^2 \]

\[ |v_k|^2 = \bar{v}_{kx}^2 + \bar{v}_{ky}^2 \]  \hspace{1cm} (6.37)

Provided that the characteristics of the noise do not change over the many reconstructions, then,

\[ \bar{v}_{kx}^2 = \rho_{x}^2(t) = \sigma_n^2 \]

and \[ \bar{v}_{ky}^2 = \rho_{y}^2(t) = \sigma_n^2 \]  \hspace{1cm} (6.38)
and, from equation 6.36,
\[ \varepsilon_1 = 2n \sigma_n^2 \] - (6.40)

If the signal in each input channel is a sinusoid of amplitude \( A \), then \( |q_k|^2 = A^2 \), and from equation 6.33,
\[ S_o = nA^2 \] - (6.41)
\[ \therefore \frac{S_o}{\varepsilon_1} = \frac{A^2}{2\sigma_n^2} \] - (6.42)

Considering any input channel, the mean square value of the signal is \( S_i = \frac{A^2}{2} \); and the mean square value of the noise is \( N_i = \sigma_n^2 \).
\[ \therefore \frac{S_i}{N_i} = \frac{A^2}{2\sigma_n^2} \] - (6.43)

Comparing equations 6.43 and 6.42, the mean signal to error power ratio at the output of the processor, before the detector, is the same as the signal to noise ratio at any input – assuming it to be uniform across the array.

6.3.3 The Performance after the Detector

Combining equations 6.32 and 6.34, the output of the reconstruction filter in response to a signal contaminated with noise is, from equation 6.30,
\[ v(t) = \sum_{k=0}^{n-1} (q_k + v_k) \exp\{j2\pi t (\frac{n-1}{2T} - \frac{k}{T})\} \] - (6.44)
Figure 6.1
Signal and noise components in $v(t)$
After the detector, the mean square error $C_2$ (defined in equation 6.13) is given by

$$C_2 = \frac{1}{T} \int \left[ \sum_{k=0}^{n-1} q_k \exp \{ j 2\pi t (\frac{n-1}{2T} - \frac{k}{T}) \} \right]$$

$$- \left[ \sum_{k=0}^{n-1} q_k \exp \{ j 2\pi t (\frac{n-1}{2T} - \frac{k}{T}) \} \right]^2 \ dt \quad (6.45)$$

Over many reconstructions, the mean value $\bar{C}_2$ can be defined, and hence a mean signal to error power ratio, $S_o/\bar{C}_2$. A relationship between $S_i/N_i$ and $S_o/\bar{C}_2$ has not been deduced, but the behaviour of $S_o/\bar{C}_2$ at high and low values of input signal to noise ratio can be estimated.

Consider, for the purpose of illustration only, the situation which would give rise to a single exponential term within the modulus bars in equation 6.45. This could be achieved by making inputs 1 to (n-1) zero, leaving signal and noise on input 0 only. The complex signal, $v(t)$, at some time during the reconstruction is shown with its component parts in Figure 6.1(a). In any one reconstruction, the angle between the noise phasor and signal phasor will remain constant, and the pair will rotate according to equation 6.44, as time progresses. Over many reconstructions, the angle between the noise phasor and signal phasor will be uniformly distributed in the range $0 \rightarrow 2\pi$.

Figure 6.1(b) shows the situation corresponding to low $S_i/N_i$ ratio, and it can be seen that $|v(t)| - |s(t)| = |v_o|$

i.e. $C_2 = |v_o|^2$

$\therefore \bar{C}_2 = |v_o|^2$
and hence $\varepsilon_2 \approx \varepsilon_1$, from equation 6.35.

Figure 6.1(c) shows the situation corresponding to high $S/N_0$, and here

$$|v(t)| - |s(t)| \approx v_{ox}$$

$\therefore \varepsilon_2 \approx v_{ox}^2$

$\therefore \tilde{\varepsilon}_2 \approx \tilde{v}_{ox}^2$

i.e. $\tilde{\varepsilon}_2 = \tilde{\varepsilon}_1/2$ from equations 6.37 and 6.38.

So, at high input signal to noise ratio, the detector gives a 3 dB improvement in the signal to error power ratio. The situation corresponding to equation 6.45 with all $n$ inputs present is much more difficult to analyse, but the same results at low and high input signal to noise ratios are thought to apply. To check this, a numerical simulation of an ideal processor with $8$ inputs was set up. 1000 trials were performed at different input signal to noise ratios, using 8 streams of complex noise samples taken from a random number generator program. Aitken (39) shows that an estimate of the variance of such a random number sequence, using 1000 samples, will have a fractional error with standard deviation of $\sqrt{2/1000}$, or 0.2 dB. The angle of incidence of the signal was normal to the array.

Initially, with the variance of the noise in each channel set at 1.0, a value for $\tilde{\varepsilon}_1$ was computed as 16.38. From equation 6.40, the theoretical value is $\tilde{\varepsilon}_1 = 16.0$, and so the error is 0.1 dB. Figure 6.2 shows the result of the simulation as a graph of the output signal to
Figure 6.2

Output signal/error ratio v. input signal/noise ratio for an ideal processor
error ratio, \( S_o/E_o \), (with \( E_o = \hat{e}_1 \) and \( \hat{e}_2 \)) plotted against input signal to noise ratio, \( S_i/N_i \). As predicted, the detector gives an improvement of up to 3 dB at high \( S_i/N_i \).

6.4 The Performance of the Digital Processor

In this section the factors which influence the performance of a digital processor are discussed.

6.4.1 Analogue to Digital Conversion Error

The A/D convertor is assumed to have a constant quantization step size, \( \kappa \), and the input-output characteristic is shown in Figure 6.3. If the number of bits available is \( \beta \), with the most significant bit acting as a sign bit, then the output range of numbers is \( -2^{\beta-1} \leq y \leq 2^{\beta-1} - 1 \). It will also be assumed that the convertor output is scaled so that the gain is unity with the number \( -2^{(\beta-1)} \) corresponding to -1 volt at the input. Then, the output range is interpreted as \( -1.0 \leq y \leq 1.2^{(\beta-1)} \), so that the decimal point is immediately to the right of the sign bit, and the quantization step size is \( \kappa = 1/2^{\beta-1} \).

Widrow\(^{40}\) has shown that the error, \( \gamma_q \), introduced by the quantization process is equivalent to the addition of noise, independent of the input, and having a uniform probability density function (p.d.f.) in the limited range \( -\kappa/2 \leq \gamma_q \leq \kappa/2 \), provided that the input is sufficiently finely quantized. This condition is satisfied even with the step size \( \kappa = \sigma_n \) for an input with a Gaussian p.d.f., and so the A/D conversion process will be treated here as adding independent random noise to the samples, with the p.d.f., \( p(\gamma_q) \), shown in Figure 6.4. The mean is zero, and the variance is given by
Figure 6.3

Input/output characteristic for A/D converter

Figure 6.4

P.d.f. of A/D conversion error
\[ \sigma_n^2 = \frac{\kappa^2}{12} \] - (6.46)

This is possible provided that the input signal stays within the input range of the convertor. If the input signal moves out of the convertor range then the output is limited to either the most positive or the most negative number, and considerable distortion can result - the convertor is said to be saturated in this condition. The function of the a.g.c. amplifier which precedes the A/D convertor is to amplify the signal level so that the convertor range is efficiently used - this reduces the effect of the quantization noise to a minimum. The gain of the amplifier is controlled so that the standard deviation, \( \sigma_n \), of the noise in any channel is amplified to a value, \( \sigma_{\text{ref}} \). The problem then becomes one of relating \( \sigma_{\text{ref}} \) to the input range of the convertor so that most signals of interest do not cause saturation.

Gray and Zeoli(41) discuss the choice of an optimum input level setting for a signal having a Gaussian p.d.f., and Vural(33) analyses the effect of word length and input level on the performance of a summing array in different sonar environments. The effect of convertor saturation on the performance of the digital DFT processor is not examined in detail here, but the results of a numerical simulation of the processor, which are discussed in section 6.4.3, enable \( \sigma_{\text{ref}} \) to be tentatively chosen, taking into account the expected range of input signal to noise ratios likely to be encountered.

6.4.2 Processor Error

The arithmetic operations performed by the DFT processor are that of addition, subtraction, and multiplication. For a processor having a
fixed word length, the result of each operation must be scaled to remain within the word length.

The result of the addition or subtraction of two numbers represented by $\beta$ bits will always be contained within $(\beta+1)$ bits, and so a division by two after each operation will ensure that the result is contained within $\beta$ bits. The multiplication of two complex numbers generates four intermediate products of length $(2\beta-1)$ bits. The intermediate products are scaled to $\beta$ bits before their subsequent addition and subtraction to form the final product. The final product will not require further scaling, provided that the modulus of the multiplicand was less than one, since the modulus of all complex multipliers is one.

The scheme described has the advantage of simplicity, even though it may not be necessary to scale after each rank of sum and difference operations in the FFT schematic diagram (Figure 3.5). Consequently, the error introduced as a result of the scaling process is larger than it might have been.

6.4.2.1 Scaling Error in Addition and Subtraction. The division of a binary number by a factor of 2 is simply accomplished by a single right shift, thus losing the least significant bit. If this bit is '0', there is no error. If this bit is '1', there will be an error whose properties depend on the convention used to represent numbers within the processor. Negative numbers may be stored as either a positive number accompanied by a sign bit, or as the 2's complement of the positive number. Either number representation could be used in a viable processor. The 2's complement representation leads to a simple arithmetic unit design and this will be adopted in the following analysis.
A right shift will round all numbers downwards (positive numbers towards zero and negative numbers away from zero). For example, consider a word length of 4 bits, where the least significant bit has a value $K = 1/2^2 = 0.125$.

$$\text{binary } 0.101 \equiv 1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3} \quad \text{decimal } 0.625$$

A right shift gives

$$0.0101 \equiv 0.25$$

This number represents the true result plus an error, $\gamma_s$, which is given by

$$\gamma_s = -\frac{K}{2}$$

For a negative number

$$1.011 \equiv -0.625$$

A right shift gives

$$1.101 \equiv -0.375$$

Again, $\gamma_s = -\frac{K}{2}$.

So, the result of a right shift is an error which is either zero or $-\frac{K}{2}$ occurring with equal probability, and over many operations of addition or subtraction followed by a right shift, the p.d.f. of the error is as shown in Figure 6.5.
Figure 6.5

P.d.f. of shift error

Figure 6.6

Sum and difference schematic
The mean value is given by

$$\bar{\bar{\gamma}}_s = -\frac{\kappa}{4}$$

and the mean square value is

$$\bar{\bar{\gamma}}_s^2 = \frac{\kappa^2}{8}$$

The propagation of errors through a $\Sigma/\Delta$ block can now be examined. The complex number at either input consists of the true number plus an error. The error will arise from a combination of noise in each transducer channel, quantization error, and processor error from any preceding operations. At one input the true number $(a_x + j a_y)$ is present, combined with an error $(\gamma_{ax} + j \gamma_{ay})$. At the other input, the true number $(b_x + j b_y)$ is present, combined with an error $(\gamma_{bx} + j \gamma_{by})$.

The schematic diagram for the sum and difference operation on the real parts of the numbers is shown in Figure 6.6. The result is the true number $c_x = (a_x + b_x)/2$, and an error, $\gamma_{cx}$, where

$$\gamma_{cx} = \frac{\gamma_{ax} + \gamma_{bx}}{2} + \gamma_s$$

The upper sign is read for the sum operation and the lower sign for the difference operation. The mean and mean square values of the error terms at the outputs of the $\Sigma/\Delta$ block are then, assuming that the individual terms are independent of each other,

$$\bar{\bar{\gamma}}_{cx} = \frac{1}{2} \bar{\bar{\gamma}}_{ax} \pm \frac{1}{2} \bar{\bar{\gamma}}_{bx} + \bar{\bar{\gamma}}_s$$

- (6.49)
Figure 6.7
P.d.f. of multiplication error
and

\[ \gamma_{cx}^2 = \frac{1}{4}\gamma_{ax}^2 + \frac{1}{4}\gamma_{ax}\gamma_{bx} + \frac{1}{2}\gamma_{s}\gamma_{s} + \gamma_{s}(\gamma_{ax} \pm \gamma_{bx}) \]  

Similar expressions would be obtained for the error in the imaginary part of the result.

6.4.2.2 Multiplication Error. The multiplication of two complex numbers generates four intermediate real products. It will be assumed that the digital multiplier acts on positive numbers only, and so a negative number in 2's complement form is converted to a positive number occupying \((\beta-1)\) bits, and the sign of the intermediate product is determined separately. The product of two \((\beta-1)\) bit numbers occupies \((2\beta-2)\) bits. Each intermediate product is then rounded to \(\beta\) bits, including the sign bit, to reduce intermediate storage requirements. The p.d.f. of the error, \(\gamma_m\), thus introduced depends on whether the number is positive or negative – both situations are shown in Figure 6.7(a) and (b), and assuming that positive and negative numbers are equally likely, the overall p.d.f. is shown in Figure 6.7(c).

Consider the complex product \(c = a.b\), where 'a' is the multiplicand and 'b' the stored multiplier. In fact,

\[ (c_x + \gamma_{cx}) + j(c_y + \gamma_{cy}) = \{(a_x + \gamma_{ax}) + j(a_y + \gamma_{ay})\} \times \{(b_x + \gamma_{bx}) + j(b_y + \gamma_{by})\} \]  

For the real part of the product only,

\[ c_x + \gamma_{cx} = \{(a_x + \gamma_{ax})(b_x + \gamma_{bx}) + \gamma_{m1}\} \]

\[ \{(a_y + \gamma_{ay})(b_y + \gamma_{by}) + \gamma_{m2}\} \]  

\[ (6.51) \]
where \( \gamma_m \) and \( \gamma_m' \) are the rounding errors associated with each intermediate product.

Equation 6.52 gives

\[
c_x = a_x b - a_y b
\]

and

\[
\gamma_{cx} = b \gamma_{ax} - b \gamma_{ay} + \gamma_m + \gamma_m' + \gamma_{bx}(a_x + \gamma_{ax}) - \gamma_{by}(a_y + \gamma_{ay})
\]

- (6.53)

In the last two terms on the R.H.S. of equation 6.53, \( \gamma_{bx} \) and \( \gamma_{by} \) are the errors incurred in representing the stored multiplier by a finite length word. As such, both \( \gamma_{bx} \) and \( \gamma_{by} \) are themselves constant, with a value in the range \(-\frac{K}{2}\) to \(\frac{K}{2}\). The effect of these two terms on the total error, \( \gamma_{cx} \), depends on the size of the multiplicand which lies in the range \(-1\) to \(+1\), and so these terms can be of the same order of significance as the rounding error. It is possible to make them insignificant by comparison with the rounding error by representing a stored multiplier by more bits than a multiplicand, say three more. It will be assumed, in what follows, that this is done. The intermediate products are still rounded to \( \beta \) bits and the value of \( \gamma_m' \) remains at \( \frac{\kappa^2}{12} \), approximately.

If the intermediate products were stored at double length, and the real and imaginary part of the product formed at double length, then only one rounding error would be involved subsequently. Such improvements are bought at the cost of more storage capacity, but since the detailed hardware design of the processor was not undertaken, the relative economics of different schemes cannot be assessed.
Equation 6.53 becomes

$$\gamma_{cx} = b_x \gamma_{ax} - b_y \gamma_{ay} + \gamma_{m1} - \gamma_{m2}$$  \hspace{1cm} - (6.54)

Consequently,

$$\tilde{\gamma}_{cx} = b_x \tilde{\gamma}_{ax} - b_y \tilde{\gamma}_{ay}$$  \hspace{1cm} - (6.55)

and

$$\tilde{\gamma}_{cx}^2 = b_x^2 \tilde{\gamma}_{ax}^2 + b_y^2 \tilde{\gamma}_{ay}^2 + 2 \tilde{\gamma}_{m1}^2 - 2b_x b_y \tilde{\gamma}_{ax} \tilde{\gamma}_{ay}$$ \hspace{1cm} - (6.56)

if the individual error terms are independent of each other.

Here $\tilde{\gamma}_{m1}^2 = \tilde{\gamma}_{m2}^2 = \gamma_m^2$, and $\tilde{\gamma}_{m1} = \tilde{\gamma}_{m2} = 0$

From equation 6.51, similar expressions for the true number and the associated error in the imaginary part of the product can be derived.

Hence,

$$c_y = a_{xy} + a_{yx} b_x$$

$$\gamma_{cy} = b_y \gamma_{ax} + b_x \gamma_{ay} + \gamma_{m1} + \gamma_{m2}$$  \hspace{1cm} - (6.57)

$$\tilde{\gamma}_{cy} = b_y \tilde{\gamma}_{ax} + b_x \tilde{\gamma}_{ay}$$  \hspace{1cm} - (6.58)

$$\tilde{\gamma}_{cy}^2 = b_y^2 \tilde{\gamma}_{ax}^2 + b_x^2 \tilde{\gamma}_{ay}^2 + 2 \tilde{\gamma}_{m1}^2 + 2b_x b_y \tilde{\gamma}_{ax} \tilde{\gamma}_{ay}$$ \hspace{1cm} - (6.59)

6.4.3 Digital Processor Performance Graphs

For a processor which does not incorporate scaling after each $\Sigma/\Delta$ operation, $S_o = nA^2$, from equation 6.41.

In the present case, scaling by a factor of 2 after each $\Sigma/\Delta$
operation amounts to an overall scale factor of $1/n$ from input to output, and so

$$ S_0 = \frac{1}{n^2} \cdot nA^2 = \frac{A^2}{n} \quad - (6.60) $$

Combining this with equation 6.9 gives,

$$ \frac{S_0}{\bar{E}_f^2} = \frac{A^2}{\sum_{r=0}^{n-1} |e_r|^2} \quad - (6.61) $$

Each term in the denominator of equation 6.61 can be deduced as a function of the noise variance in each input channel and the word length of the processor, by using the equations developed so far in section 6.4. This procedure is illustrated in Appendix G. For an 8 element processor with phase compensation at the output for a real reconstruction filter, from Table G.4 in Appendix G,

$$ \sum_{r=0}^{7} |e_r|^2 = 2\sigma^2_{\text{ref}} + \frac{8.32}{2^{2\beta-2}} \quad - (6.62) $$

Equation 6.62 enables performance graphs to be drawn for different values of $\sigma_{\text{ref}}$ and $\beta$, provided that the A/D convertor is not saturated.

A numerical simulation was performed to confirm the predicted performance figures, and to evaluate the effect of A/D convertor saturation. The same set of input noise samples were used as for the ideal processor, and the phase of the signal phasors, for a plane wavefront with normal incidence to the array, was allowed to vary uniformly throughout the range $0 + 2\pi$. 1000 trials were run at different $S_i/N_i$ ratios, with word lengths of 5, 6, 7, and 8 bits, and two different values of $\sigma_{\text{ref}}$, of 0.1 and 0.2.
The performance before the detector is shown in Figures 6.8 and 6.9, and after the detector in Figures 6.10 and 6.11. In Figure 6.8, in the unsaturated region of the A/D convertor the simulation results agree with the theoretical line (from equation 6.62) to better than 0.1 dB for $\beta = 5$ and 6, and 0.2 dB for $\beta = 7$ and 8.

Saturation effects would be expected when the amplitude of the sinusoidal signal is approximately 1 since the input range of the convertor is $-1 \rightarrow 1 - 2^{-(\beta-1)}$, i.e. $S_i/N_i = 0.5/\sigma^2_{\text{ref}}$. In this case, with $\sigma_{\text{ref}} = 0.1$, saturation will occur with $S_i/N_i = 17$ dB, and this is born out by the simulation results. No more results in this region, other than those indicated, could be obtained with the program which was used, and it was felt that the need for more information was not sufficient to warrant any program changes.

An improvement in the ratio $S_i/E_i$ would be expected with the signal amplitude just less than 1 since the noise is clipped and its mean square value reduced. The results obtained for $\sigma_{\text{ref}} = 0.2$ show this effect quite clearly in Figure 6.9; here, for saturation on the signal component, $S_i/N_i = 11$ dB. For both values of $\sigma_{\text{ref}}$, serious saturation effects do not occur for values of $S_i/N_i$ at least 1 dB greater than that calculated above. This is because the phase of the signal phasor is uniformly distributed in the range $0 \rightarrow 2\pi$, and so it is possible for the signal amplitude to be in the range $1 \rightarrow \sqrt{2}$ and not cause saturation in either the real or quadrature components of the sample. This leads to a convenient rule of thumb which can be adopted for the value of input signal to noise power ratio, $(S_i/N_i)_{\text{sat}}$, which can be reached before saturation effects become serious,
Figure 6.8
Digital processor performance before detector, for word lengths of 5, 6, 7, and 8 bits.
\[ \sigma_{\text{ref}} = 0.1 \]
Figure 6.9

Digital processor performance before detector, for word lengths of 5, 6, 7, and 8 bits.

$\sigma_{\text{ref}} = 0.2$
Figure 6.10

Digital processor performance after detector, for word lengths of 5, 6, 7, and 8 bits.

$\sigma_{\text{ref}} = 0.1$
Figure 6.11

Digital processor performance after detector, for word lengths of 5, 6, 7, and 8 bits.

\[ \sigma_{ref} = 0.2 \]
\[
\begin{align*}
\left( \frac{S_i}{N_i} \right)_{\text{sat}} & = 10 \log_{10} \left( \frac{1}{2 \sigma_{\text{ref}}^2} \right) + 1 \text{ dB} \\
\text{i.e.} \quad \left( \frac{S_i}{N_i} \right)_{\text{sat}} & = 20 \log_{10} \left( \frac{1}{\sigma_{\text{ref}}^2} \right) - 2 \text{ dB} \quad - (6.63)
\end{align*}
\]

\( \sigma_{\text{ref}} \) is chosen to suit the highest \( \frac{S_i}{N_i} \) ratio likely to be met, although there is no reason why this should not be under the control of the operator to some extent. The word length of the processor depends on the value of \( \sigma_{\text{ref}} \); higher values of \( \sigma_{\text{ref}} \) reduce the significance of the processor error contribution in equation 6.62. With \( \sigma_{\text{ref}} = 0.1 \), the performance of a digital processor, for an array of eight elements, is practically indistinguishable from the ideal if the word length is 8 bits.

From Figures 6.10 and 6.11, the improvement in the output signal to error power ratio obtained after the detector in the case of an ideal processor, is maintained for a digital processor having a finite word length.

6.5 The Performance of the Analogue Processor

The building blocks used in an analogue processor will suffer from practical deficiencies which affect the performance of the sonar receiver. In the present case, this will be due to the tolerance on the passive component values and the non-ideal characteristics of the operational amplifier in the frequency band of interest.

It will be assumed that each output of the basic units can still be expressed as a linear combination of the inputs, so that the practical
equivalent of the $\Sigma/\Delta$ unit can be represented by the matrix,

$$G = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{(6.64)}$$

and the phase shift unit is represented by the multiplier, $r$, where $a, b, c, d,$ and $r$ are defined in terms of the components in the equivalent circuits of the units.

If the matrix $G$, in equation 6.64, is substituted for the submatrices specifying the $\Sigma/\Delta$ operations in the DFT matrix factorisation, and likewise $r$ for the phase shift unit, the matrix product can be multiplied out to form an $n \times n$ matrix, $M$. $M$ now describes the practical processor, in place of the ideal DFT matrix, $N$. In fact,

$$M = N + E \quad \text{(6.66)}$$

where $E$ is an error matrix of dimension $n \times n$. Each element of $E$ will be a function of several variables contributed by the different units comprising the network.

From a practical viewpoint, it is necessary to know the likely effect on the performance of a network of an assumed production spread in all of the components. This is difficult to estimate in the case of an incident plane wave, and a numerical simulation of 1000 networks is used to provide the information. The circuit models used in the simulation are described in the next section.

6.5.1 The Description of Practical Units

At the proposed carrier frequency of 32 kHz, the simple model of the
An equivalent circuit of the LM709C amplifier

The Sum Unit

The Difference Unit
LM709C operational amplifier shown in Figure 6.12 was found to be quite adequate for predicting the behaviour of the \( \Sigma/\Delta \) and phase shift units.

The equivalent circuit for the sum unit is shown in Figure 6.13, and the output voltage is given by,

\[
v_{os} = \frac{-1}{1+\mu_s} \left( \frac{R_3}{R_1} v_1 + \frac{R_3}{R_2} v_2 \right)
\]

where

\[
\mu_s = \left\{ \frac{1 + R_3 \left( \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_1} \right)}{A_v - \frac{R_0}{R_1}} \right\}^+ \left[ 1 + \frac{R_2}{R_1} \right]
\]

and \( R_1^* = R_1 + R_4 \); \( A_v^* = A_v \cdot \frac{R_1^*}{R_1} \).

The nominal value of resistors \( R_1 \), \( R_2 \), and \( R_3 \) is 5.6 k\( \Omega \), and \( R_4 = 1.8 \) k\( \Omega \). The spread of \( R_1 \) and \( R_0 \) about their typical values of 250 k\( \Omega \) and 200 \( \Omega \) will have a negligible influence on \( \mu_s \), compared with the variation expected in \( A_v \), so \( R_1 \) and \( R_0 \) were taken as constant values throughout the simulation.

The output impedance, \( Z_{os} \), of the sum unit is given by

\[
Z_{os} = \frac{3R_0}{A_v}
\]

Since \( A_v = 100 \) \([-110^\circ] \) typically, the effect of \( Z_{os} \) on the performance of following stages is negligible.

The equivalent circuit for the difference unit is shown in Figure 6.14, and the output voltage is given by

\[
v_{od} = \frac{1}{1+\mu_d} \left\{ \left( \frac{1 + R_7/R_8}{1 + R_6/R_8} \right) \cdot v_2 - \frac{R_7}{R_8} v_1 \right\}
\]
Assumed in Figure 6.15

The Phase Shift Unit

Figure 6.16

Assumed p.d.f. for the fractional error, $\Delta$, in a passive component value (±2%)

Figure 6.17

Distribution of voltage gain measurements, $A_v$, for 40 samples of the LM709C amplifier at 32kHz
where \( \mu_d = \frac{1 + R_7 \left( \frac{1}{R_5} + \frac{1}{R_7'} \right)}{A'_v - \frac{R_9}{R_7}} \left\{ 1 + \frac{R_9}{R_7} \right\} \) \hspace{1cm} \text{-(6.70)}

and \( R_7' = R_1 + \frac{R_6 \cdot R_8}{R_6 + R_0} \), \( A'_v = A_v \cdot \frac{R_1}{R_7} \). The output impedance is given by

\[
Z_{od} = \frac{2R_8}{A'_v}
\]

The equations describing the behaviour of a phase shift unit, whose equivalent circuit is shown in Figure 6.15, can be obtained directly from equations 6.69 and 6.70 by putting \( v_2 = v_1 \), \( R_6 = Z_{12} \), and \( R_8 = Z_{10} \), \( R_7 = R_{11} \), and \( R_5 = R_9 \). Then, the output voltage is given by

\[
v_{op} = \frac{1}{1 + \mu_d} \left\{ \frac{1 - \frac{R_{11}}{R_9} \cdot \frac{Z_{16}}{Z_{12}}}{1 + \frac{Z_{16}}{Z_{12}}} \right\} v_1
\]

with \( \mu_d \) defined in equation 6.70, and input voltage = \( v_1 \).

For a phase retard circuit, \( Z_{10} = R \), and \( Z_{12} = \frac{1}{j\omega C} \), say. Then,

\[
v_{op} = \frac{1}{1 + \mu_d} \left\{ \frac{1 - \frac{R_{11}}{R_9} \cdot j\omega CR}{1 + \frac{1}{j\omega CR}} \right\} v
\]

\hspace{1cm} \text{-(6.71a)}

with \( R_{11} = R_9 \) nominally,

\[
|v_{op}| = |v_1|
\]

and \( \arg(v_{op}/v_1) = -2 \tan^{-1}(\omega CR) \) \hspace{1cm} \text{-(6.71b)}

For a phase advance circuit, \( Z_{10} = \frac{1}{j\omega C} \), and \( Z_{12} = R \).

\[
\therefore \quad v_{op} = \frac{-1}{1 + \mu_d} \left\{ \frac{R_{11}}{R_9} - j\omega CR \right\} v_1
\]

\hspace{1cm} \text{-(6.72a)}
and again with $R_{11} = R_0$ nominally,

$$|v_{op}| = |v_1|$$

and

$$\arg(v_{op}/v_1) = \pi - 2 \tan^{-1}(\omega CR)$$

It is assumed that a passive component, specified as a nominal value with an associated tolerance, takes on all values with equal probability throughout the tolerance band. For example, a resistance $R$, specified as $R_N$ nominally, can be expressed as $R = R_N(1 + \Delta)$, and the p.d.f., $p(\Delta)$, is shown in Figure 6.16 for $\pm 2\%$ tolerance. Such a uniform p.d.f. will yield more pessimistic results than one resembling a truncated Gaussian p.d.f., which is more likely to be met in practice.

The distribution of values taken on by the voltage gain parameter, $A_v$, of the amplifier is derived from measurements made on 40 specimens of the LM709C at 32 kHz. The results are shown in Figure 6.17 and, in the simulation, it is assumed that the values of $A_v$ taken on by different amplifiers are uniformly distributed throughout the region within the dotted rectangle.

6.5.2 Plane Wave Response of Processor

In this section, the response of a large number of networks to an incident plane wave in the absence of noise is investigated. This shows the limitation on the performance imposed by the components used in the network.

The error, between the interpolated output of a particular practical network and the output of an ideal network, would be expected to vary
Figure 6.18

(a) P.d.f. of error before the detector in response to a plane wave of unit amplitude.

\[ s = \frac{10}{32L} \]

(b) P.d.f. of error after the detector in response to a plane wave of unit amplitude.

\[ s = \frac{10}{32L} \]
with the angle of incidence of the wavefront. However, taking a large number of networks, there is no obvious reason for the error characteristics of the family of networks to change significantly with the angle of incidence. It was not possible to perform a sufficient number of tests to confirm this view, but a numerical simulation was run at five different angles of incidence chosen in an arbitrary manner. Taking the unambiguous range of angles, \(-\frac{1}{2L} < \theta < \frac{1}{2L}\), the angles of incidence corresponded to \(s = -\frac{12}{32L}, -\frac{7}{32L}, 0, \frac{3}{32L}, \) and \(\frac{10}{32L}\).

The same set of 1000 networks was used in each run, with all circuit parameters given values by a random number generator program. This is equivalent to using 144000 passive components and 36000 amplifiers. The results, for \(\pm 2\%\) and \(\pm 1\%\) component tolerance, obtained by using a plane wave input of unit amplitude and at an angle of incidence corresponding to \(s = \frac{10}{32L}\), are shown in Figures 5.18(a) and (b). From Figure 6.18(b), for \(\pm 2\%\) components, 95\% of the networks gave an error \(e_2 < 0.0186\), which corresponds to \(S_o/e_2 > 26.3\ dB\); and the average value over all networks, \(\bar{e}_2 = 0.0096\), corresponds to \(S_o/\bar{e}_2 = 29.2\ dB\). For \(\pm 1\%\) components, 95\% of the networks gave \(S_o/e_2 > 29.0\ dB\), and \(S_o/\bar{e}_2 = 31.5\ dB\).

There is a large discrepancy between the signal to error ratios before and after the detector, for instance \(S_o/\bar{e}_2 = 17.6\ dB\) for \(\pm 2\%\) components, compared with \(S_o/\bar{e}_2 = 29.2\ dB\). This is caused by the phase shift, associated with the amplifiers used in all the units in the network, which is contained in the argument of the complex valued term \(1/(1+\mu)\) in equations 6.67, 6.69, 6.71a, and 6.72a. The spread in values of \(1/(1+\mu)\) is investigated in section 6.5.3.1, and the phase shift imparted by a sum unit is \(-1.8^\circ\), typically, and \(-1.2^\circ\) for a difference unit or
phase shift unit. The net effect is to impart a phase lag, to each output of the network, which is typically $5^\circ + 6^\circ$. The length of the error vector between the ideal and actual outputs is used to calculate $\varepsilon_1$ (equation 6.9), and consequently $\varepsilon_1$ can be substantially larger than $\varepsilon_2$ because of this phase lag.

The results obtained for the signal to error ratio after the detector, for the 5 angles of incidence considered, are summarised in Table 6.1, for ±2% components.

<table>
<thead>
<tr>
<th>$\sin \theta$</th>
<th>$\frac{S_0}{\varepsilon_2}$ (dB)</th>
<th>$95%$ lower limit for $\frac{S_0}{\varepsilon_2}$ (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{10}{32L}$</td>
<td>29.2</td>
<td>26.3</td>
</tr>
<tr>
<td>$\frac{3}{32L}$</td>
<td>29.0</td>
<td>25.6</td>
</tr>
<tr>
<td>0</td>
<td>31.2</td>
<td>26.0</td>
</tr>
<tr>
<td>$-\frac{7}{32L}$</td>
<td>29.4</td>
<td>25.3</td>
</tr>
<tr>
<td>$-\frac{12}{32L}$</td>
<td>28.7</td>
<td>24.3</td>
</tr>
</tbody>
</table>

Table 6.1
Summary of error characteristics after the detector for 1000 networks with ±2% component tolerance

6.5.3 The Noise Response of the Processor

The noise voltage, $e_r(t)$, in output channel $r$ of the analogue processor has a zero mean value, and a variance, $\sigma_r^2$, which is related to the noise variance, $\sigma_n^2$, in each input channel by the expression,
\[ \sigma_r^2 = \sigma_r^2(t) = \sum_{k=0}^{n-1} |m_{rk}|^2 \] - (6.73)

where \( m_{rk} \) is the coefficient in row \( r \) and column \( k \) in matrix \( M \).

The outputs of the quadrature demodulators in each channel are sampled, and the real and imaginary sample sets applied to the reconstruction filter. Over many reconstructions,

\[ \bar{e}_1 = \frac{1}{n} \sum_{r=0}^{n-1} |e_r|^2 \] - (6.74)

where \[ |e_r|^2 = e_{rx}^2 + e_{ry}^2 \]

\[ = 2\sigma_r^2 \] - (6.75)

since \( e_{rx}^2 = e_{ry}^2 = \sigma_r^2 \).

Combining equations 6.75, 6.74 and 6.73 gives

\[ \bar{\varepsilon}_1 = \frac{2\sigma_r^2}{n} \sum_{r=0}^{n-1} \sum_{k=0}^{n-1} |m_{rk}|^2 \] - (6.76)

Equation 6.76 gives a value of \( \bar{\varepsilon}_1 \) for a particular processor. \( \bar{\varepsilon}_1 \) will will vary from one processor to another over a range which is related to the spread in values taken on by the coefficients \( m_{rk} \). A mean value, \( \overline{\varepsilon}_1 \), taken over many processors can be defined, and also a measure of the spread about the mean. From equation 6.76

\[ \overline{\varepsilon}_1 = \frac{2\sigma_r^2}{n} \sum_{r} \sum_{k} |m_{rk}|^2 \] - (6.77)

and the variance \( \sigma_{\text{net}}^2 \) about the mean value is
\[ \sigma_{net}^2 = \frac{(E_1 - E_1)^2}{(E_1 - E_1)^2} \]  
\[ \text{i.e. } \sigma_{net}^2 = \frac{4\sigma_n^8}{n^2} \left\{ \sum_r \sum_k \left( |m_{rk}|^2 - \frac{1}{n} \right) \right\} \]  
- (6.78)

Equations 6.77 and 6.79 are developed in Appendix H for an analogue processor of dimension 8, using passive components tolerated to ±2% or better.

6.5.3.1 Amplifier Gain Variations. The amplifier gain is the circuit parameter subject to most change in equations 6.67, 6.69, 6.71a, and 6.72a, to the extent that the \( a, b, c, d, \) and \( r \) parameters describing the behaviour of the different units can be viewed as the product of two essentially independent components. One component is the term \( 1/1+\mu \) (where \( \mu = \mu_s \) or \( \mu_d \)), whose variability from one unit to another can be considered to be a function of amplifier gain only; and the other component is a function of the passive components only. For example, in the equivalent of the \( \Sigma/\Delta \) unit, \( a = -(1/1+\mu_s)(R_3/R_1) \). External resistor ratios, such as \( R_3/R_1 \), can be adjusted to 0.1% accuracy if necessary, so it is quite possible for the performance of the network to be limited by the amplifiers used within it.

Allocating typical values to the resistive components in the expressions for \( \mu_s \) and \( \mu_d \), gives,

\[ \mu_s = \frac{3.14}{A_v} \]

and

\[ \mu_d = \frac{2.10}{A_v} \]
Figure 6.19

P.d.f.'s of variables associated with the amplifier in a Σ/Δ unit
If $A_v = a_x + j.a_y$, then

$$\frac{1}{1+\mu_s} = \frac{a_x + j.a_y}{(a_x + 3.14) + j(a_y)}$$

A similar expression for $1/(1+\mu_d)$ can be generated.

The variations in $|\frac{1}{1+\mu_s}|$ and $|\frac{1}{1+\mu_d}|$ were investigated numerically, using a random number generator program, and the results for 1000 samples are shown in Figure 6.19(a) and (b).

Putting $|\frac{1}{1+\mu_s}| = K_s(1 + \Delta_s)$

where $K_s = |\frac{1}{1+\mu_s}|$, and $\Delta_s$ is the fractional deviation from the mean, gives

$$K_s = 1.004$$

and $\overline{\Delta_s^2} = 6 \times 10^{-6}$

Likewise, if $|\frac{1}{1+\mu_d}| = K_D(1 + \Delta_D)$,

$$K_D = 1.003$$

and $\overline{\Delta_D^2} = 3 \times 10^{-6}$

Now, $\text{arg}(\frac{1}{1+\mu_s}) = \tan^{-1}\left[\frac{3.14a_y}{a_x^2 + 3.14a_x + a_y^2}\right]$ \hspace{2cm} (6.80)

The mean value of equation 6.80 can be deduced algebraically, using the method for evaluating the mean of a function of independent variables (42).

This gives the mean phase shift associated with $(1/1+\mu_s)$ as $-1.8^\circ$, and $-1.2^\circ$ for $(1/1+\mu_d)$. 
6.5.4 Performance Graph for an Analogue Processor

From equation H.22

\[ \tilde{\varepsilon}_1 = 16.5 \sigma_n^2 \]  \hspace{1cm} (6.81)

Equation 6.81 should be compared with the value of \( \varepsilon_1 = 16 \sigma_n^2 \) for an ideal processor, obtained from equation 6.40. Also, from equation H.28

\[ \sigma_{net}^2 = 4 \sigma_n^4 (48 \Delta_S^2 + 96 \Delta_D^2 + 76 \Delta_T^2) \]  \hspace{1cm} (6.82)

where \( \Delta_T \) defines the limits of the passive component tolerance band (= 0.02 for ±2% tolerance).

The principal use of equation 6.82, which relates a measure of the spread in \( \tilde{\varepsilon}_1 \) to the processor component variations, is that the spread in values of \( S_o/\tilde{\varepsilon}_2 \) at low values of \( S_i/N_i \) can be inferred. From Figure 6.2 the difference between the curves for \( S_o/\tilde{\varepsilon}_2 \) and \( S_o/\tilde{\varepsilon}_1 \) can be determined for an ideal processor, at any value of \( S_i/N_i \).

Results from a simulation of 10 different analogue processors show that this performance differential is maintained at low values of \( S_i/N_i \) (< 3 dB) where the error component due to the signal is small enough to be ignored.

For a family of processors having the same design tolerances, a representative graph of \( S_o/\tilde{\varepsilon}_2 \) against \( S_i/N_i \) can be drawn. This is done in Figure 6.20 for an eight input analogue network with all components tolerated to ±2%. Substituting values obtained in section 6.5.3.1 into equation 6.82 gives

\[ \sigma_{net} = 0.35 \sigma_n^2 \]  \hspace{1cm} (6.83)
Figure 6.20

Performance of analogue processor after the detector,

\[ s = \frac{10}{32L}, \pm 2\% \text{ component tolerance} \]
which is almost entirely due to the passive component variations.

Since the value of \( \sigma_{\text{net}} \) in equation 6.83 is almost entirely due to a large number of passive component variations, each with a uniform distribution, the resulting variation in \( \varepsilon_1 \) about its mean value \( \varepsilon_1 \) would be expected to follow an approximately Gaussian distribution. Consequently, 95\% of the networks would give a value of \( \varepsilon_1 \) within \( 2\sigma_{\text{net}} \) (approx. 0.2 dB) of the mean.

Using the results in equations 6.81 and 6.83, and also the performance differential from Figure 6.2, the asymptotic behaviour of the graph in Figure 6.20 can be drawn at low \( S_i/N_i \). The limiting performance at high \( S_i/N_i \) is obtained from Figure 6.18(b) for an angle of incidence corresponding to \( \theta = \frac{10}{32L} \). Ten different networks were simulated and a typical curve for one network is shown, along with the spread in \( S_o/\varepsilon_2 \) at \( S_i/N_i = 21 \text{ dB} \). All networks gave results within \( \pm 0.2 \text{ dB} \) of the calculated mean, for \( S_i/N_i = -6 \text{ dB} \), according to prediction. It is not feasible to simulate a large number of networks since the results for just one network require 20 minutes of computer time.

6.6 The Error Introduced by a Non-Ideal Interpolating Filter

In the preceding sections, the performance of the digital and analogue processors has been examined under the assumption that an ideal low pass filter is used for interpolating between the output samples. A practical filter will allow some breakthrough of frequency components outside the baseband of the spectrum of the sample sequence to be filtered, and this will give an error in the output waveform.

The limiting performance of a particular filter can be obtained by
IV(f)

Figure 6.21
Amplitude spectrum at output of interpolating filter

\[ |s(t)| > |e(t)| \]

\[ |s(t)| < |e(t)| \]

Figure 6.22
Signal and error components in output of filter in response to plane wave input to processor
assuming that the DFT processor is ideal, and that the filter output is in the steady state condition. The filter which will be examined here is the 5 pole Chebyshev filter introduced in section 4.6.2.1. The filter characteristics within the passband are compensated according to Table 4.2 to give uniform amplitude shading and a linear phase progression across an array of eight elements.

6.6.1 Response to a Plane Wave Signal in the Absence of Noise

The amplitude spectrum, $|V(f)|$, of the waveform, $v(t)$, at the output of the filter is shown in Figure 6.21. The transfer function, $H(f)$, of the interpolating filter is defined as

$$H(f) = H_H(f) \cdot H_F(f)$$

where $H_H(f)$ and $H_F(f)$ are the transfer functions of the zero-order hold and the Chebyshev filter, respectively.

The spectrum, $S^s(f)$, of the input sequence, $s^s(t)$, is periodic with period $\frac{n}{T}$, and throughout the entire frequency range $V(f)$ is defined by

$$V\left(\frac{n-1}{2T} + \frac{ni}{T} - \frac{k}{T}\right) = q_k \frac{b_k}{\tau} \cdot H\left(\frac{n-1}{2T} + \frac{ni}{T} - \frac{k}{T}\right)$$

for $-\infty < i < \infty$, integer values

and $k = 0, 1, \ldots, n-1$

Equation 6.85 is an extension of equation 4.51, and the baseband is defined with $i = 0$.

From equation 6.7,

$$V(f) = S(f) + E(f)$$
where the error term, \( E(f) \), is defined by \( V(f) \) outside the baseband,

\[- \frac{n}{2T} < f < \frac{n}{2T} \].

From equation 6.8, applying Parseval's theorem,\(^1\)

\[
\varepsilon_1 = \sum_{i=-\infty}^{i=\infty} \sum_{k=0}^{n-1} q_k \frac{B_k}{T} |H(\frac{n-1}{2T} + \frac{ni}{T} - \frac{k}{T})|^2
\]  

For a unit plane wave, \(|q_k| = 1\), and \(|B_k|\) and \(|H(f)|\) are shown in Table 6.2 for the 5 pole Chebyshev filter with 0.5 dB ripple and cut off frequency \( f_c = \frac{7}{2T} \). Only the first two components adjacent to the passband are shown since additional terms are insignificant by comparison.

<table>
<thead>
<tr>
<th>( f )</th>
<th>(- \frac{11}{2T})</th>
<th>(- \frac{9}{2T})</th>
<th>( \frac{9}{2T})</th>
<th>( \frac{11}{2T})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(</td>
<td>B_k</td>
<td>)</td>
<td>1.220</td>
<td>1.484</td>
</tr>
<tr>
<td>( \frac{1}{T}</td>
<td>H(f)</td>
<td>)</td>
<td>0.013</td>
<td>0.078</td>
</tr>
<tr>
<td>( \frac{1}{T}</td>
<td>H(f)</td>
<td>\cdot</td>
<td>B_k</td>
<td>)</td>
</tr>
</tbody>
</table>

**Table 6.2**  
Spectral components of significance outside the passband of the interpolating filter

Substituting values into equation 6.87 gives

\[
\varepsilon_1 = 2(0.016^2 + 0.116^2)
\]

i.e. \( \varepsilon_1 = 0.0274 \)
Now \[ S_0 = \sum_{k=0}^{n-1} |q_k|^2 \]

\[ = 8 \]

\[ \therefore \frac{S_0}{\delta_1} = 24.65 \text{ dB} \]

The effect of this error can be seen in Figure 4.20 which provides a feel for the numbers involved. A numerical simulation was performed to determine the steady state response of the filter and detector for a plane wave input to the system. The results are:

\[ \frac{S_0}{\delta_1} = 24.65 \text{ dB} \]

and \[ \frac{S_0}{\delta_2} = 24.78 \text{ dB} . \]

The error before the detector checks with the calculation using equation 6.87. The slight improvement after the detector is accounted for by the nature of the error signal. The field in the aperture, due to an incident plane wave, exhibits conjugate symmetry about its value at the centre of the aperture (section 4.5.1). Combining this with the conjugate symmetry of the terms \( B_k H(n-1, \frac{n+1}{T} - \frac{k}{T}) \) in equation 6.85 means that the argument of the complex waveform, \( v(t) \), remains constant with time, and so the error component can only affect the modulus of \( v(t) \). Figure 6.22(a) and (b) shows the situation for \( |s(t)| > |e(t)| \) and \( |s(t)| < |e(t)| \).

Provided that \( |e(t)| < |s(t)| \), equation 6.88a is true
\[ |v(t)| - |s(t)|^2 = |e(t)|^2 \]  

otherwise relation 6.88b is possible, as in Figure 6.22(b)

\[ |v(t)| - |s(t)|^2 < |e(t)|^2 \]  

Considering the low level of error introduced by this particular filter, relation 6.88b is possible only during the time that \(|s(t)| < 0.03\) approximately (relative to \(|s(t)|_{\text{max}} = 1.0\)). Consequently, relation 6.88a holds most of the time, and \(c_2 = c_1\) from equations 6.8 and 6.13.

### 6.6.2 Response to a Plane Wave Signal Plus Noise

Now, there are two sources of error contained in \(v(t)\), due to

(i) the noise component in each spectral line within the passband, and

(ii) breakthrough outside the passband.

\[
E(i) = v_k \frac{B_k}{T} H \left( \frac{n-1}{2T} - \frac{k}{i} \right) + (v_k + q_k) \frac{B_k}{T} H \left( \frac{n-1}{2T} + \frac{n_i - k}{i} \right) - (6.89)
\]

for \(-\infty < i < \infty\), \(i \neq 0\)

\[
\text{and } k = 0, 1, ..., n-1
\]

Within the passband, \(\left| \frac{B_k}{T} H \left( \frac{n-1}{2T} - \frac{k}{i} \right) \right| = 1\),

\[
\therefore c_1 = \sum_{k=0}^{n-1} \left| v_k \right|^2 + \sum_{i=-\infty}^{\infty} \sum_{k=0}^{n-1} (v_k + q_k) \frac{B_k}{T} H \left( \frac{n-1}{2T} + \frac{n_i - k}{i} \right)^2 - (6.90)
\]

\[
\therefore \tilde{c}_1 = \sum_{k=0}^{n-1} \left| v_k \right|^2 + \sum_{i=-\infty}^{\infty} \sum_{k=0}^{n-1} \left( \left| v_k \right|^2 + \left| q_k \right|^2 \right) \frac{B_k}{T} H \left( \frac{n-1}{2T} + \frac{n_i - k}{i} \right)^2 - (6.91)
\]
Figure 6.23

Steady state performance of 5-pole Chebyshev filter, 0.5 dB ripple
since the signal and noise sample in each channel are assumed to be independent.

If the amplitude of the input signal is $A$, then $|q_k| = A$, and

$$S_o = nA^2$$  \hspace{1cm} (6.41)

Also,

$$|v_k|^2 = 2\sigma_n^2$$  \hspace{1cm} (6.39)

For an eight element array, and using the results in Table 6.2, equation 6.91 becomes

$$\bar{\varepsilon}_1 = 16\sigma_n^2 + 2\sigma_n^2 (0.0274) + A^2 (0.0274)$$

$$\therefore \frac{S_o}{\bar{\varepsilon}_1} = \frac{8A^2}{16.05\sigma_n^2 + 0.0274A^2}$$  \hspace{1cm} (6.92)

Using equations 6.92 and 6.43, the output signal to error ratio before the detector is plotted against the input signal to noise ratio in Figure 6.23. The curve for $S_o/\bar{\varepsilon}_2$ can be drawn approximately, using equation 6.93,

$$\frac{S_o}{\bar{\varepsilon}_2} = \frac{8A^2}{8\sigma_n^2 + 0.0274A^2}$$  \hspace{1cm} (6.93)

for $\frac{S_i}{N_i} > 12$ dB

Equation 6.93 is derived from the assumption that, at high signal to error ratios, the action of the detector is to remove the quadrature component of the error, leaving the in-phase component. This is comprised
principally of the in-phase noise component from within the passband and signal breakthrough from outside the passband.

6.6.3 Transient Effects

Because of the limited time available for interpolation, the filter output may not be in the steady state by the time it is required for display. A numerical simulation was used to determine the error in the filter output during the display period for the scheme suggested in section 5.2.6. A plane wave input in the absence of noise was used, with angles of incidence corresponding to \( s = \pm \frac{0.75}{2L} \) and \( 0 \) in the unambiguous range \(-\frac{1}{2L} < s < \frac{1}{2L}\). This would give a 'bright-up' on the display screen at the far right, far left, and centre. The results are shown in Table 6.3.

<table>
<thead>
<tr>
<th>( \sin \theta )</th>
<th>( \frac{S_O}{\varepsilon_1} ) (dB)</th>
<th>( \frac{S_O}{\varepsilon_2} ) (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-\frac{0.75}{2L})</td>
<td>22.5</td>
<td>22.8</td>
</tr>
<tr>
<td>(0)</td>
<td>23.0</td>
<td>23.4</td>
</tr>
<tr>
<td>(\frac{0.75}{2L})</td>
<td>23.6</td>
<td>24.1</td>
</tr>
</tbody>
</table>

Table 6.3

Transient error of practical filter

As would be expected, the display waveform for a target to the right of the screen has less error associated with it than that for a target to the left, because of the longer time interval for the transient response of the filter to die away.
6.7 Summary of Chapter 6

The average measures of performance discussed in this chapter attempt to quantify the behaviour of the principal parts of a sector scan receiver in terms of the variable design parameters. Although no assessment has been made of the performance of the combination of a practical DFT processor and a non-ideal interpolating filter, inspection of the appropriate graphs shows the relative error contributions of the DFT processor and the 5-pole Chebyshev filter. From Figures 6.20 and 6.23 it can be inferred that this particular filter is fairly well matched to the analogue network using 2% tolerance components, in that both produce approximately the same amount of error in the display waveform when taken individually. This would be a reasonable combination to use as a basis for future experimental work, although to err on the side of safety, the use of 1% components in the processor would cause the overall performance to be limited by the practical realisation of the filter.

Although a processor for an 8-element array has been made the subject of the investigation, the techniques are quite general and can easily be extended to a 16-element array processor.
CHAPTER 7

Test Results

7.1 Results of Tests on the Digital Processor

The test environment used for the initial evaluation of the receiver, described in section 5.2, was a water tank, measuring 3.2 m long by 2.0 m wide by 0.9 m deep. The transmitting array illuminated a sector between -30° and +30° to the normal to the receiving array. Each of the eight receiving transducers measured approximately λ wide by 5λ high at the operating frequency of 500 kHz, giving an overall array width of 2.4 cms (= 8λ).

Two flat plates, of dimensions 12.2 cms × 12.2 cms and 8.4 cms × 12.2 cms, were used as targets at a minimum range of 2.1 m. In each target position, the plates were orientated with the normal at the centre pointing directly towards the transmitting/receiving transducer head. This gave maximum returned signal strength, and the reflected wavefront can be considered to be plane over the length of the receiving array.

In the limited time available the tests were restricted to simple configurations, involving one or two targets in a single range cell, under essentially noise free conditions. Photographs were taken of the display, and the samples in each input channel to the processor were recorded. The word length of the A/D convertor and the digital processor was 11 bits.

The channel samples were processed subsequently by a computer program which simulated the operation of the digital processor, filter,
and detector, and graphical outputs were obtained of the display intensity modulation waveforms. Since the single channel detector was used in the equipment, each Figure (7.1 to 7.5) accompanying the photographs (Plates 7.1 to 7.5) consists of three parts:

(a) the single channel detector output used to produce the display,
(b) the output of the complex detector, shown in Figure 4.10, for comparison with (a),
(c) a reference output waveform corresponding to an artificial target normal to the array.

The scale interval in all figures is arranged to be the same as the large scale interval in the photographs, and this corresponds to \( s = 0.11 \). Over the range of \( s \), \( (\sin \theta) \), being used, the scale is approximately linear in \( \theta \), with \( 6.5^\circ \) per division.

The effect of different word lengths in the A/D convertor and digital processor, down to 4 bits, is shown in Figure 7.6.

7.1.1 Commentary

Only the phase characteristics of the interpolating filter were compensated at the outputs of the transducers, and so the filter amplitude characteristic caused the aperture to be shaded, as described in section 4.6.1. The effect of this is to produce the point source response shown in Figure 7.1(c), for example.

Comparison of bearings taken from the centre of the display trace with the actual bearings of the targets is extremely good in all instances; the measurements are recorded under each photograph.
Figure 7.1
Intensity modulation waveforms
(Horizontal scale = 6.5°/division)
Plate 7.1

Horizontal scale = 6.5°/cm
1 target, bearing = 0°
Measured bearing from display = 0°
1. Single channel detector, intensity modulation waveform corresponding to Plate 7.2

2. Complex detector

3. Artificial target, bearing = 0°

Figure 7.2
Intensity modulation waveforms
(Horizontal scale = 6.5°/division)
Plate 7.2

Horizontal scale = 6.5°/cm

1 target, bearing = +18°

Measured bearing from display = +18°
Single channel detector, intensity modulation waveform corresponding to Plate 7.3

Complex detector

Artificial target, bearing = 0°

Figure 7.3

Intensity modulation waveforms
(Horizontal scale = 6.5°/division)
Plate 7.3

Horizontal scale = 6.5°/cm

1 target, bearing = -18°

Measured bearing from display = -18°
1.0

Iv(t) I

Single channel detector, intensity modulation waveform corresponding to Plate 7.4

1.0

|v(t)|

complex detector

1.0

|v(t)|

artificial target, bearing = 0°

Figure 7.4

Intensity modulation waveforms
(Horizontal scale = 6.5°/division)
Plate 7.4

Horizontal scale = 6.5°/cm
2 targets, bearings = +4°, +18°
Measured bearings from display = +5°, +19°
Single channel detector, intensity modulation waveform corresponding to Plate 7.5

(a) Single channel detector, intensity modulation waveform corresponding to Plate 7.5

(b) Complex detector

(c) Artificial target, bearing = 0°

Figure 7.5

Intensity modulation waveforms
(Horizontal scale = 6.50/division)
Plate 7.5

Horizontal scale = 6.5°/cm

2 targets, bearings = -13°, +18°

Measured bearings from display = -12°, +18°
\[ \beta = 11 \text{ bits} \]

\[ \beta = 8 \text{ bits} \]

\[ \beta = 7 \text{ bits} \]
Figure 7.6
Complex detector output corresponding to Figure 7.1(c), with variable word length $\beta$.

(a) $\beta = 6$ bits

(b) $\beta = 5$ bits

(c) $\beta = 4$ bits
One effect not visible in the photographs is the behaviour of the first sidelobe on the left of the main lobe in the single target situations shown in Figures 7.1, 7.2, and 7.3. There is insufficient data available to offer a positive explanation for this. One possible cause is the presence of d.c. offsets in the input channels before A/D conversion. It was found necessary to zero these offsets from time to time as they were subject to drift. Another possible cause is non-uniform illumination of the aperture - the target used was approximately 40\(\lambda\) wide, and consequently the uniformity of the field across the receiving aperture is very sensitive to target orientation. Incorrect amplitude and phase compensation in each channel will distort the intensity modulation waveform from its desired shape as in Figure 7.1(c), but this shape would not be expected to change with target position if the only effect of this was to impart a linear phase progression across the array. A combination of all three of these factors is the most likely explanation.

In the tests with two targets, the relative positions of the target plates had to be adjusted so that a reasonable display of both was obtained. This was necessary because of the single channel detector scheme being used. In a more realistic time varying situation, both targets would be displayed over a period of time. The graph in Figure 7.4(a) does not show the waveform used to generate the display in Plate 7.4, although the target bearings for both were the same. The orientation of one target plate was changed slightly to improve the display photograph after the samples used to generate Figure 7.4(a) had been recorded.

Figure 7.6 is intended only as an illustration of the effect of using too short a word length.
A full assessment of the behaviour of a digital processor in a noisy sea-going environment has yet to be made, but the results of preliminary tests in the water tank are encouraging.

7.2 Results of Tests on the Analogue Processor

A ready-made transducer head was not available for testing the analogue processor with target echoes, so the results obtained so far were derived from bench tests with an array simulator. This had eight sine wave outputs at 32 kHz, with provision for varying the phase difference between channels, so that the outputs of an 8-element receiving array in response to a plane wavefront from any bearing could be simulated. It is estimated that the accuracy in setting up this phase difference, and of subsequent phase measurements at the outputs of the processor, is within ±0.3°.

The unambiguous range, $-\frac{1}{2L} < s < \frac{1}{2L}$ was divided into 32 equal intervals in $s$, and the eight outputs of the analogue network measured at each of the 32 angles of incidence. These results were then compared with the computed outputs of an ideal network, for the same input level, and the quantities $S_0/\varepsilon_1$ and $S_0/\varepsilon_2$ calculated, assuming an ideal interpolating filter. The results of this operation are listed in Table 7.1. Wherever the angle of incidence coincided with the axis of one of the 8 receiving beam patterns, (looking on the DFT as a multiple beam-forming operation), the levels at the 7 remaining outputs were too low for reliable phase measurements to be taken, and so the quantity $S_0/\varepsilon_2$ could not be calculated.

The measured outputs of the analogue processor and the calculated outputs of an ideal processor are listed in Table 7.2, for one angle of
Table 7.1

Results of tests on analogue processor, 1% components

<table>
<thead>
<tr>
<th>( i (\sin \theta = \frac{i}{3L}) )</th>
<th>( \frac{S_0}{\varepsilon_1} ) (dB)</th>
<th>( \frac{S_0}{\varepsilon_2} ) (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-15</td>
<td>19.2</td>
<td>27.5</td>
</tr>
<tr>
<td>-14</td>
<td>19.4</td>
<td>29.4</td>
</tr>
<tr>
<td>-13</td>
<td>19.2</td>
<td>29.5</td>
</tr>
<tr>
<td>-12</td>
<td>19.4</td>
<td></td>
</tr>
<tr>
<td>-11</td>
<td>19.5</td>
<td>30.4</td>
</tr>
<tr>
<td>-10</td>
<td>20.1</td>
<td>32.1</td>
</tr>
<tr>
<td>-9</td>
<td>19.5</td>
<td>28.2</td>
</tr>
<tr>
<td>-8</td>
<td>19.7</td>
<td></td>
</tr>
<tr>
<td>-7</td>
<td>19.7</td>
<td>27.4</td>
</tr>
<tr>
<td>-6</td>
<td>20.3</td>
<td>31.5</td>
</tr>
<tr>
<td>-5</td>
<td>20.3</td>
<td>31.3</td>
</tr>
<tr>
<td>-4</td>
<td>20.4</td>
<td></td>
</tr>
<tr>
<td>-3</td>
<td>20.4</td>
<td>30.6</td>
</tr>
<tr>
<td>-2</td>
<td>20.1</td>
<td>29.9</td>
</tr>
<tr>
<td>-1</td>
<td>19.4</td>
<td>26.9</td>
</tr>
<tr>
<td>0</td>
<td>19.7</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>19.2</td>
<td>27.8</td>
</tr>
<tr>
<td>2</td>
<td>19.0</td>
<td>28.7</td>
</tr>
<tr>
<td>3</td>
<td>18.6</td>
<td>29.4</td>
</tr>
<tr>
<td>4</td>
<td>18.5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>19.0</td>
<td>30.1</td>
</tr>
<tr>
<td>6</td>
<td>19.5</td>
<td>29.0</td>
</tr>
<tr>
<td>7</td>
<td>19.3</td>
<td>27.4</td>
</tr>
<tr>
<td>8</td>
<td>18.7</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>18.4</td>
<td>26.8</td>
</tr>
<tr>
<td>10</td>
<td>19.3</td>
<td>30.2</td>
</tr>
<tr>
<td>11</td>
<td>19.0</td>
<td>29.6</td>
</tr>
<tr>
<td>12</td>
<td>19.1</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>19.2</td>
<td>29.2</td>
</tr>
<tr>
<td>14</td>
<td>19.5</td>
<td>29.2</td>
</tr>
<tr>
<td>15</td>
<td>19.3</td>
<td>27.1</td>
</tr>
<tr>
<td>16</td>
<td>19.1</td>
<td></td>
</tr>
</tbody>
</table>
incidence. The magnitudes are normalised to the maximum value of the ideal response.

<table>
<thead>
<tr>
<th>Output number</th>
<th>Output of analogue processor</th>
<th>Output of ideal processor</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>$</td>
<td>u_r</td>
</tr>
<tr>
<td></td>
<td>$\arg(u_r)$ ($^\circ$)</td>
<td>$\arg(s_r)$ ($^\circ$)</td>
</tr>
<tr>
<td>0</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>1</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>2</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>3</td>
<td>0.20</td>
<td>0.19</td>
</tr>
<tr>
<td>4</td>
<td>0.94</td>
<td>0.90</td>
</tr>
<tr>
<td>5</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td>6</td>
<td>0.15</td>
<td>0.14</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>0.10</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Table 7.2
Outputs of practical and ideal analogue processors in response to plane wave, $s = -\frac{15}{32L}$

7.2.1 Commentary

Little can be said about the results listed in Table 7.1 except that they are in good agreement with what might be expected from the numerical simulation results in section 6.5.2, allowing some degradation for setting up the inputs and measuring the outputs of the processor.

Table 7.2 illustrates the point made in section 6.5.2 that all the outputs of the analogue processor lag behind the outputs of the ideal processor by approximately $6^\circ$ because of the limited open-loop gain of the operational amplifiers at 32 kHz. Another point demonstrated is
that the argument of the interpolated output waveform will remain constant with time in response to a plane wave in the ideal case, (i.e. the real and imaginary parts of the complex waveform are related by a scale factor - sections 4.5.1 and 6.6.1), because of the 'spectrum shifting' phase shifts at the output of the DFT network. Observe the values of \( \arg(s_x) \), noting that \( 129.4^\circ + 180^\circ \equiv -50.6^\circ \).
CHAPTER 8

Conclusions

The principal conclusion of this investigation is that the methods proposed for generating a within-pulse sector scan display are both practicable and also have the potential for producing a low cost receiver suitable for use by the fishing industry.

The investigation was directed at an analogue processor and a digital processor for an 8-element array. If each transducer is of width $\lambda$, this would produce the effect of a receiving beam of width approximately $7^\circ$ scanned over a sector extending $\pm 30^\circ$ about the normal. This may provide adequate resolution in some applications, but the design techniques are completely general and there are no foreseeable problems in extending the number of receiving elements to 16.

The analogue processor is the cheaper to produce at the present time, with the discrete component cost being less than £25 for an eight element array. Using the 709 amplifier, the carrier frequency is restricted to less than 40 kHz, and the interpolation scheme should cope adequately with the shortest pulse length likely to be used at this frequency. Doubtless, higher performance differential amplifiers will be developed which will enable the frequency range to be extended if required, but the use of transformers to perform the sum and difference operation in a high frequency analogue processor may be worth investigation.

The digital processor is not restricted to a particular frequency range of operation because the received signals are demodulated before the processor. The shortest pulse length will be determined by either the
input multiplex rate or the interpolator, and a 200\(\mu\)s pulse length combined with a 16 element array is thought to be possible, although the cost of the processor will be high at present if multiple arithmetic units are necessary to produce this order of operating speed. It is within the design philosophy of large scale integration to produce a single chip containing a complete arithmetic and control unit for the processor, and this is a possibility for the future. Even so, it is unlikely that the digital processor could compete with the analogue processor on economic grounds within the latter's operating frequency range, at least for an array of 16 elements or less.

As for future work, the a.g.c. amplifier remains to be developed, but the problems of producing electronically variable gain over a wide bandwidth down to D.C. are not thought to be insuperable.

The performance measure of signal-to-error power ratio, used in Chapter 6, allowed a comparison to be made between the two types of processor and an ideal processor. A correlation remains to be made between this measure of performance and the performance as observed on the display screen.
REFERENCES


35. 'The Application of Linear Microcircuits', Volume 1, SGS-UK Ltd.


The factorisation procedure to be described requires that \( n \) should be expressible as the product of 2 factors,

\[
n = p \cdot q
\]  

- (A.1)

The general case, where \( n \) is the product of any number of factors, is dealt with by successive application of this technique.

Labelling the rows and columns, in the square DFT matrix \( N \), 0 to \( n-1 \), the element in row \( r \) and column \( k \) is defined by

\[
r_{r,k} = n^{r \cdot k}
\]  

- (A.3)

where

\[
n = \exp \left(-j\frac{2\pi}{n}\right)
\]  

- (A.4)

Since \( n = p \cdot q \), equation A.4 can be written as

\[
n = \exp \left(-j\frac{2\pi}{p \cdot q}\right)
\]  

- (A.5)

and putting

\[
p_0 = \exp \left(-j\frac{2\pi}{p}\right)
\]  

- (A.6)

equation A.5 becomes

\[
n = p^q
\]  

- (A.7)
To start the factorisation procedure, the elements of $N$ are written in terms of $p$ using equation A.7.

In any row $r$, and column $k = mq$, where $m = 0, 1, 2, \ldots, p-1$, the element is defined by

$$n_{r, mq} = \frac{rq}{q}$$

i.e. $n_{r, mq} = p^{rm}$

an integer power of $p$.

The next step is to rewrite $N$ as the sum of $q$ $n \times n$ matrices, $V_s$, where $s = 0, 1, 2, \ldots, q-1$.

$$N = \sum_{s=0}^{q-1} V_s$$

Column $s + mq$ in $V_s$ is identical to column $s + mq$ in $N$, all other elements in $V_s$ being zero. In any matrix, $V_s$, the element in row $r$, column $s + mq$, is

$$v_{r, s+mq} = \frac{r(s+mq)}{q}$$

$$= \frac{rs}{q} \cdot p^{rm}$$

$$= r^{rs} \cdot p^{rm}$$

using equation A.7. Equation A.10 shows that a factor $r^{rs}$ can be taken outside each row. Each matrix $V_s$ can then be expressed as a product.
\[ V_s = T_s U_s \] - (A.11)

where \( T_s \) is an \( n \times n \) diagonal matrix with

\[ t_{r,k} = \delta_{rk} \cdot n^{rs} \] - (A.12)

\[ \delta_{rk} = \begin{cases} 0 & (r \neq k) \\ 1 & (r = k) \end{cases} \]

and \( U_s \) is an \( n \times u \) matrix with

\[ u_{r,s+mq} = \frac{p^r}{m} \] - (A.13)

all other elements being zero. Equation A.13 shows that corresponding rows in all matrices, \( U_s \), contain the same elements, although in different columns defined by the integer \( s \).

To illustrate the procedure so far, put \( n = 6 \), \( p = 3 \), and \( q = 2 \). So, \( m = 0, 1, 2; s = 0, 1; \) and \( n = p^4 \)

\[
N = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 \\
1 & n & n^2 & n^3 & n^4 & n^5 \\
1 & n^2 & n^4 & n^6 & n^8 & n^{10} \\
1 & n^3 & n^6 & n^9 & n^{12} & n^{15} \\
1 & n^4 & n^8 & n^{12} & n^{16} & n^{20} \\
1 & n^5 & n^{10} & n^{15} & n^{20} & n^{25}
\end{bmatrix}
= \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 \\
1 & \frac{1}{p} & \frac{1}{p} & \frac{1}{p} & \frac{1}{p} & \frac{1}{p} \\
1 & \frac{1}{p^2} & \frac{1}{p^2} & \frac{1}{p^2} & \frac{1}{p^2} & \frac{1}{p^2} \\
1 & \frac{1}{p^3} & \frac{1}{p^3} & \frac{1}{p^3} & \frac{1}{p^3} & \frac{1}{p^3} \\
1 & \frac{1}{p^4} & \frac{1}{p^4} & \frac{1}{p^4} & \frac{1}{p^4} & \frac{1}{p^4} \\
1 & \frac{1}{p^5} & \frac{1}{p^5} & \frac{1}{p^5} & \frac{1}{p^5} & \frac{1}{p^5}
\end{bmatrix}
\] - (A.14)

From equation A.9

\[ N = \sum_{s=0}^{1} V_s \] - (A.15)
\[
\begin{bmatrix}
1 & 1 & 1 & \\
1 & p & p^2 & \\
1 & p^2 & p^4 & \\
1 & p^3 & p^6 & \\
1 & p^4 & p^8 & \\
1 & p^5 & p^{10} &
\end{bmatrix}
\]

\[= \begin{bmatrix}
1 & 1 & 1 & \\
p^{1/2} & p^{3/2} & p^{5/2} & \\
p & p^3 & p^5 & \\
p^{3/2} & p^{5/2} & p^{7/2} & \\
p^2 & p^6 & p^{10} & \\
p^{5/2} & p^{9/2} & p^{11/2} & \\
p & p^5 & p^{10} & \\
p^{5/2} & p^{9/2} & p^{11/2} & p^{15/2}
\end{bmatrix}
\]

\[\text{i.e. } N = \begin{bmatrix}
1 & 1 & 1 & \\
1 & p & p^2 & \\
1 & p^2 & p^4 & \\
1 & p^3 & p^6 & \\
1 & p^4 & p^8 & \\
1 & p^5 & p^{10} &
\end{bmatrix}
\]

\[+ \begin{bmatrix}
1 & \\
1 & p & p^2 & \\
1 & p^2 & p^4 & \\
1 & p^3 & p^6 & \\
1 & p^4 & p^8 & \\
1 & p^5 & p^{10} &
\end{bmatrix}
\]

From equations A.11, A.12, and A.13

\[N = \begin{bmatrix}
1 & 1 & 1 & \\
1 & p & p^2 & \\
1 & p^2 & p^4 & \\
1 & p^3 & p^6 & \\
1 & p^4 & p^8 & \\
1 & p^5 & p^{10} &
\end{bmatrix}
\]

\[+ \begin{bmatrix}
n & \\
n^2 & \\
n^3 & \\
n^4 & \\
n^5 &
\end{bmatrix}
\]

\[\begin{bmatrix}
1 & 1 & 1 & \\
1 & p & p^2 & \\
1 & p^2 & p^4 & \\
1 & p^3 & p^6 & \\
1 & p^4 & p^8 & \\
1 & p^5 & p^{10} &
\end{bmatrix}
\]

Continuing, now, from equation A.13, matrix \( U_s \) is partitioned into \( q \) submatrices of dimension \( p \times n \). Row \( l \) in each submatrix may be addressed by putting

\[r = l + ip\]

where \( l = 0, 1, 2, \ldots, p-1 \)

\[i = 0, 1, 2, \ldots, q-1\]

So that, from equation A.13

\[u_{l+ip,s+mq} = \begin{bmatrix} p^{(l+ip)m} \\
p^l \cdot (p^p)^{im} \\
p^l m
\end{bmatrix}\]

\[\text{i.e. } u_{l+ip,s+mq} = p^l m\]
since \( p^p = 1 \).

Equation A.19 states that corresponding rows in each sub-matrix in \( U_s \) are identical.

In keeping with the partitioning of \( U_s \), \( T_s \) is now partitioned into submatrices of dimension \( p \times p \), which are either null matrices or diagonal matrices, \( D_{si} \).

In each matrix, \( D_{si} \), the element in row \( \ell \) and column \( m \) is,

\[
d_{\ell,m} = \delta_{\ell,m} \cdot n^{(\ell + ip)s}
\]

from equations A.12 and A.18.

Putting \( q = \exp(-j2\pi/q) \) - (A.21)

and repeating \( n = \exp(-j2\pi/pq) \) - (A.5)

then \( n = q^p \) - (A.22)

\[
\therefore n^{(\ell + ip)s} = n^{-s} \cdot q^i
\]

Putting \( D_{s0} = D_s \), (i.e. \( i = 0 \)), it can be seen that, generally

\[
D_{si} = q^is \cdot D_s
\]

where an element in \( D_s \) is defined by

\[
d_{\ell,m} = \delta_{\ell,m} \cdot n^{-s}
\]

Equations A.24 and A.25 are obtained by combining equations A.20 and A.23.
The illustration can now be taken a stage further. Applying equations A.19 and A.24 to A.17,

\[
\begin{bmatrix}
1 & 1 & 1 \\
1 & p & p^2 \\
1 & p^2 & p^3 \\
1 & 1 & 1 \\
1 & p & p^2 \\
1 & p^2 & p^3
\end{bmatrix} = 1_N + D_1 \begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix} = 1_{N(1)} + D_1 \begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix} \quad - (A.26)
\]

where

\[
D_1 = \begin{bmatrix}
1 \\
N \\
N^2
\end{bmatrix} \quad - (A.27)
\]

from equation A.25.

The \( p \) columns containing non zero elements in matrix \( U_s \) are moved now so that they are 'condensed', still maintaining the same column order, into a block of \( p \) adjacent columns. This is done in a particular way so that the first \( p \) columns in the first matrix, \( U_0 \), are occupied; the second \( p \) columns in the second matrix, \( U_1 \), are occupied; and so on to the last matrix, \( U_{(q-1)} \), where the last group of \( p \) columns are occupied. This generates a set of matrices, \( W_s \), \((s = 0, 1, \ldots, q-1)\) which are post multiplied by a common 'reordering' matrix to maintain equality with matrices \( U_s \). In the reordering matrix, \( C \), the first \( p \) rows are chosen to reorder the columns of the first matrix, \( W_0 \); the second \( p \) rows are chosen to reorder the columns of the second matrix, \( W_1 \), and so on.
Thus \( U_s = W_s C \) \hspace{1cm} - (A.28)

Matrix \( W_s \) may be partitioned into submatrices of dimension \( p \times p \), and the element in row \( \ell \) and column \( m \) of each submatrix containing non-zero elements is given by

\[
W_{\ell+p,m+sp} = P^{\ell m} \hspace{1cm} - (A.29)
\]

The elements in matrix \( C \) are defined by

\[
c_{m+sp,s+mq} = 1 \hspace{1cm} - (A.30)
\]

\[m = 0, 1, 2, \ldots, p-1\]

\[s = 0, 1, 2, \ldots, q-1\]

with all other elements zero.

Now, a DFT of dimension \( p \) is defined by the matrix, \( P \), where

\[
P_{\ell,m} = P^{\ell m} \hspace{1cm} - (A.31)
\]

\[\ell = 0, 1, 2, \ldots, p-1\]

\[m = 0, 1, 2, \ldots, p-1\]

so that equation A.29 states that the submatrices in \( W_s \), which contain non-zero elements, each define a DFT of dimension \( p \times p \).

Combining equations A.9, A.11, and A.28 gives

\[
N = \sum_{s=0}^{q-1} T_s W_s C \hspace{1cm} - (A.32)
\]
or \[ N = B \cdot C \]  

where \[ B = \sum_{s=0}^{q-1} T_s W_s \]  

In the example, from equation A.26

\[
N = \begin{bmatrix}
1 & 1 & 1 & \ldots \\
1 & p & p^2 & \ldots \\
1 & p^2 & p^4 & \ldots \\
1 & 1 & 1 & \ldots \\
1 & p & p^2 & \ldots \\
1 & p^2 & p^3 & \ldots \\
1 & p^4 & p^5 & \ldots \\
1 & 1 & 1 & \ldots \\
1 & p & p^2 & \ldots \\
1 & p^2 & p^3 & \ldots \\
\end{bmatrix} + \begin{bmatrix}
D_1 & \ldots & \ldots & \ldots \\
\ldots & D_q & \ldots & \ldots \\
\ldots & \ldots & D_q & \ldots \\
\ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots \\
\end{bmatrix} = \begin{bmatrix}
1 & \ldots & \ldots & \ldots \\
1 & \ldots & \ldots & \ldots \\
1 & \ldots & \ldots & \ldots \\
1 & \ldots & \ldots & \ldots \\
1 & \ldots & \ldots & \ldots \\
1 & \ldots & \ldots & \ldots \\
1 & \ldots & \ldots & \ldots \\
1 & \ldots & \ldots & \ldots \\
1 & \ldots & \ldots & \ldots \\
1 & \ldots & \ldots & \ldots \\
\end{bmatrix}
\]  

Since it has been arranged that the matrices, \( W_s \), \( s = 0, 1, \ldots, q-1 \), each have different, non-overlapping, blocks of \( p \) columns occupied by non-zero elements, the summation in equation A.34 is straightforward. \( B \) is a matrix of dimension \( n \times n \) which may be partitioned into submatrices of dimension \( p \times p \). Adopting the following notation

\[
B = \begin{bmatrix}
B_{0,0} & B_{0,1} & \ldots & B_{0,q-1} \\
B_{1,0} & \ldots & \ldots & \ldots \\
B_{q-1,0} & \ldots & \ldots & \ldots \\
\end{bmatrix}
\]  

\[
B = \begin{bmatrix}
\vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots \\
\end{bmatrix}
\]  

\[
\downarrow \quad q \text{ submatrices}
\]

\[
\uparrow \quad p \text{ cols.}
\]
then \[ B_{i,s} = \text{cis} \cdot D_s \cdot P \] \hspace{1cm} \text{(A.37)}

In the example, from equation A.35
\[ N = \begin{bmatrix} P & D_1 \cdot P \\ \vdots & \vdots \\ P & D_1 \cdot P \end{bmatrix} \cdot C \] \hspace{1cm} \text{(A.38)}

where
\[ P = \begin{bmatrix} 1 & 1 & 1 \\ 1 & P & P^2 \\ 1 & P^2 & P^4 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 1 & \ldots & \ldots & \ldots \\ \ldots & 1 & \ldots & \ldots \\ \ldots & \ldots & 1 & \ldots \\ \ldots & \ldots & \ldots & 1 \end{bmatrix} \]

The effect of the manipulation so far has been to reorder the columns of \( N \) in a systematic manner, and to express each element in \( N \) as a product which leads directly to the following factorisation.

Equation A.37 shows that \( B \) may be expanded as the product of 3 matrices, of dimension \( n \times n \), to give
\[ B = N_Q \cdot R \cdot P_p \] \hspace{1cm} \text{(A.39)}

where \( P_p \) is a quasidiagonal matrix having \( q \) submatrices, \( P \), of dimension \( p \times p \), on the main diagonal, and zero elements elsewhere
\[ P_p(i,s) = \delta_{is} P \] \hspace{1cm} \text{(A.40)}

\( R \) is a quasidiagonal matrix containing submatrices of dimension \( p \times p \), where
\[ R_{i,s} = \delta_{is} D_s \]  

- (A.41) 

\[ N_q \] is partitioned into submatrices of dimension \( p \times p \), each of which is a diagonal matrix defined by

\[ N_q(i,s) = \delta_{is} \cdot I \]  

- (A.42)

where \( I \) is the identity matrix of dimension \( p \times p \).

In the example, from equation A.38

\[
\begin{bmatrix}
    D_1 & D_2 & \cdots & D_{p-1} \\
    q_1 & q_2 & \cdots & q_{p-1} \\
\end{bmatrix}
\begin{bmatrix}
    I & 0 & \cdots & 0 \\
    0 & I & \cdots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & I \\
\end{bmatrix}
\begin{bmatrix}
    I & 0 & \cdots & 0 \\
    0 & I & \cdots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & I \\
\end{bmatrix}
\begin{bmatrix}
    D_1 & D_2 & \cdots & D_{p-1} \\
    q_1 & q_2 & \cdots & q_{p-1} \\
\end{bmatrix}
\]

- (A.43)

where

\[
I = \begin{bmatrix}
1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & 1 \\
\end{bmatrix}
\]

The final step in the factorisation is to reorder the rows and columns of \( N_q \) to form a quasidiagonal matrix, \( Q_Q \), having \( p \) submatrices, \( Q \), of dimension \( q \times q \), on the main diagonal and zero elements elsewhere.

Then,

\[ N_Q = C \cdot Q_Q \cdot C^T \]  

- (A.44)

where \( Q_Q(\ell,m) = \delta_{\ell m} Q \)  

- (A.45)

\[ \ell = 0, 1, 2, \ldots, p-1 \]

\[ m = 0, 1, 2, \ldots, p-1 \]
Matrix $Q$ defines a DFT of dimension $q$, with the element in row $i$ and column $s$ given by

$$q_{i,s} = q^{i s}$$  \hspace{1cm} \text{(A.46)}$$

$$i = 0, 1, 2, \ldots, q-1$$

$$s = 0, 1, 2, \ldots, q-1$$

$C$ is the reordering matrix already defined in equation A.30, and $C^T$ is its transpose.

In the example, from equation A.43

$$N_Q = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & 1 & 1 & \ldots & 1 \\ . & . & . & \ldots & . \\ 1 & 1 & 1 & \ldots & 1 \\ 1 & 1 & 1 & \ldots & 1 \end{bmatrix}$$

and from equation A.44

$$N_Q = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & 1 & 1 & \ldots & 1 \\ . & . & . & \ldots & . \\ 1 & 1 & 1 & \ldots & 1 \\ 1 & 1 & 1 & \ldots & 1 \end{bmatrix}$$

$$N_Q = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & 1 & 1 & \ldots & 1 \\ . & . & . & \ldots & . \\ 1 & 1 & 1 & \ldots & 1 \\ 1 & 1 & 1 & \ldots & 1 \end{bmatrix}$$

In this case, $Q = \begin{bmatrix} 1 & 1 \\ 1 & q \end{bmatrix}$
Combining equations A.33, A.39, and A.44 gives

\[ N = C \cdot Q_Q \cdot c^T \cdot R \cdot P_p \cdot C \]  \quad - (A.47)

The number of rotation vectors in \( R \) is of some interest, and can be determined from equations A.41 and A.25. Within each diagonal matrix, \( D_\kappa \), of dimension \( p \times p \), there are \((p-1)\) rotations since \( d_{0,0} = \frac{n^0}{n} = 1 \). There are \( q \) such matrices within matrix \( R \). However, one of these, defined by \( R_{0,0} = D_0 \), contains no rotations according to equation A.25. Therefore, the total number of rotations in \( R \) is \((q-1)(p-1)\).

So, equation A.47 shows that a DFT of dimension \( n \) can be decomposed into \( p \) DFT's of dimension \( q \), plus \( q \) DFT's of dimension \( p \), plus \((p-1)(q-1)\) rotations (complex multiplications or phase shifts) in addition to those contained within the DFT's.
APPENDIX B

On the Reduction of Computing Effort by using the FFT

Let \( m_a \) = the number of additions in a DFT of dimension \( n \),

taking two quantities at a time.

\[ m_a = \text{the number of complex multiplications in a DFT of} \]

dimension \( n \),

then \( m_a = n(n-1) \) \( \cdots \) \( - (B.1) \)

and \( m_m = (n-1)^2 \) \( - (B.2) \)

Result B.2 arises because there are no complex multiplications con­
tained in the first row and first column of the DFT matrix.

It was shown in Appendix A that if \( n = p \cdot q \), the DFT of dimension \( n \) can be decomposed into \( p \) DFT's of dimension \( q \), plus \( q \) DFT's of
dimension \( p \), plus \( (p-1)(q-1) \) complex multiplications. Consequently,
if \( m_a^* \) and \( m_m^* \) refer to the numbers of additions and complex multipli­
cations in the DFT of dimension \( n \), after decomposition, then

\[ m_a^* = p \cdot q(q-1) + q \cdot p(p-1) \] \( \cdots \) \( - (B.3) \)

and \[ m_m^* = p(q-1)^2 + q(p-1)^2 + (p-1)(q-1) \] \( \cdots \) \( - (B.4) \)

If the decomposition is to be useful, then it would be expected that

\[ m_a^* < m_a \] \( \cdots \) \( - (B.5) \)

and \[ m_m^* < m_m \] \( \cdots \) \( - (B.6) \)
The conditions expressed in relations B.5 and B.6 taken together are more restrictive than is necessary, since a reduction in overall computation is required, and it may be possible to achieve this with only one of these conditions satisfied.

From equation B.3

\[ m'_a = pq(p + q - 2) \]  \hspace{1cm} - (B.7)

and, from equation B.1

\[ m_a = pq(pq - 1) \]  \hspace{1cm} - (B.8)

If expression B.5 is true, then from equations B.7 and B.8

\[ pq(p + q - 2) < pq(pq - 1) \]

i.e. \( (q - 1) < p(q - 1) \)  \hspace{1cm} - (B.9)

Since both \( p \) and \( q \) are integers greater than 1, expression B.9 is satisfied and expression B.5 is always true.

Expanding equation B.4

\[ m''_m = p(q^2 - 2q + 1) + q(p^2 - 2p + 1) + (pq - p - q + 1) \]

\[ \therefore m''_m = pq(p + q - 3) + 1 \]  \hspace{1cm} - (E.10)

Expanding equation B.2 and putting \( n = pq \),

\[ m_m = p^2q^2 - 2pq + 1 \]  \hspace{1cm} - (B.11)
If expression B.6 is true, then, from equations B.10 and B.11
\[ pq(p + q - 3) + 1 < p^2q^2 - 2pq + 1 \]
i.e. \[ p + q - 3 < pq - 2 \]
i.e. \[ (q - 1) < p(q - 1) \]
- (B.12)

Expression B.12 is identical to expression B.9, therefore expression B.6 is true.

To summarise, both the number of additions, taking two quantities at a time, and the total number of complex multiplications are reduced by the FFT factorisation of the DFT matrix.
Figure C.1

The first stage of a DFT decomposition
APPENDIX C

The Derivation of Expressions for the Number of Arithmetic Operations in the FFT

n is the product of \( t \) terms, so that

\[
n = \prod_{i=1}^{t} n_i \quad - (C.1)
\]

or

\[
n = n_1 \cdot n_2 \cdot n_3 \cdots n_{t-1} \cdot n_t. \quad - (C.2)
\]

In order to assess the total number of complex multiplications and additions in the FFT, the result obtained for just two factors is extended successively throughout the product expressed in equation C.2. This technique is implied by the introduction of nested parentheses into equation C.2 to give

\[
n = n_1 \cdot (n_2 \cdot (n_3 \cdot (\ldots (n_{t-1} \cdot n_t)\ldots))) \quad - (C.3)
\]

The first step is to consider the product of \( n_1 \) and the term inside the outermost set of parentheses (i.e. \( \frac{n}{n_1} \)). The operations involved, after performing the FFT decomposition, are shown schematically in Figure C.1. Then, each one of the DFT's of dimension \( \frac{n}{n_1} \) is decomposed in the same way, taking the product of \( n_2 \) and the term inside the second set of parentheses (i.e. \( \frac{n}{n_1 \cdot n_2} \)). The complete decomposition is built up in this fashion, and expressions will be derived which give the total number of complex multiplications and additions (taking two quantities at a time) required after the FFT decomposition.

In a DFT of dimension \( n_1 \), there are \((n_1 - 1)\) additions, taking
two quantities at a time, in each row of the matrix. There are $n_i$ rows, so the total number of additions is $n_i(n_i - 1)$. Now there are $\frac{n}{n_i}$ DFT's of dimension $n_i$ associated with each term in equation C.1. Therefore, the total number of additions, $m^*$, after the FFT decomposition is given by

$$m^* = \sum_{i=1}^{t} \frac{n}{n_i} \cdot n_i(n_i - 1)$$

or

$$m^* = \sum_{i=1}^{t} n_i(n_i - 1)$$

or

$$m^* = n\{-t + \sum_{i=1}^{t} n_i\}$$

- (C.4)

- (C.5)

The transforms, which make up the FFT, give rise to a total of $m_{mi}$ complex multiplications from within, in addition to $m_{me}$ multiplications which occur between the transforms. If $n_i$ is a prime number, there are $(n_i - 1)^2$ complex multiplications in a DFT of dimension $n_i$. However, when $n_i = 2$, the single complex multiplier is equal to $-1$, and the DFT of dimension 2 is defined by

$$N = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

- (C.6)

Equation C.6 shows that only sum and difference operations are involved, and in the following derivation a DFT of dimension 2 is counted as having zero internal multiplications. For this reason, factors of 2 are removed from equation C.1, so that in general,

$$n = 2^t \cdot \prod_{i=2+1}^{t} n_i$$

- (C.7)
The total number of complex multiplications, $m_m$, after the FFT decomposition is given by

$$m_m = m_{m_1} + m_{m_e}$$  \hspace{1cm} (C.8)

Now,

$$m_{m_1} = \prod_{i=\ell+1}^{t} \frac{n}{n_i} \cdot (n_1 - 1)^2$$  \hspace{1cm} (C.9)

arising from the transforms of dimension other than 2. Also,

$$m_{m_e} = (n_1 - 1)(\prod_{i=2}^{t} n_i - 1) + n_1(n_2 - 1)(\prod_{i=3}^{t} n_i - 1) + \ldots$$

$$+ n_1 n_2 \ldots n_{t-2}(n_{t-1} - 1)(n_t - 1)$$  \hspace{1cm} (C.10)

Equation C.10 arises from consideration of the successive decomposition procedure.

Expanding equation C.9 gives,

$$m_{m_1} = n\{2(\ell-t) + \frac{t}{\ell+1} \sum_{i=\ell+1}^{t} n_i + \frac{t}{\ell+1} \sum_{i=\ell+1}^{t} \frac{1}{n_i} \}$$  \hspace{1cm} (C.11)

A factor $n$ can be removed from each term in the summation C.10 to give

$$m_{m_e} = n\{(1 - \frac{1}{n_1})(1 - \frac{1}{t}) + \ldots + (1 - \frac{1}{n}) (1 - \frac{1}{t}) + \ldots$$

$$\prod_{i=2}^{t} n_i$$

$$+ (1 - \frac{1}{n_{t-1}})(1 - \frac{1}{n_t}) \}$$  \hspace{1cm} (C.12)

Expanding the products in equation C.12, the term $-\frac{1}{t} \prod_{i=j}^{t} n_i$
arising from the \((j - 1)\)th product cancels with the term
\[
\frac{1}{t} \prod_{i=j+1}^{n_j} n_i
\]

Thus,
\[
m_{me} = n \left\{ (1 - \frac{1}{n_1}) + (1 - \frac{1}{n_2}) + (1 - \frac{1}{n_3}) + \ldots + (1 - \frac{1}{n_j}) + \ldots + (1 - \frac{1}{n_{t-1}}) - \frac{1}{n_t} \right\}
\]

or
\[
m_{me} = n \left\{ \frac{1}{n} - \frac{1}{n_t} + \sum_{i=1}^{t-1} (1 - \frac{1}{n_i}) \right\}
\]

\[
m_{me} = n \left\{ (t - 1) + \frac{1}{n} - \sum_{i=1}^{t} \frac{1}{n_i} \right\}
\]

Combining equations C.11 and C.14,
\[
m_{m} = n \left\{ 2t - t - 1 + \sum_{i=k+1}^{t} n_i - \sum_{i=1}^{k} \frac{1}{n_i} \right\} + 1
\]

But \(\sum_{i=1}^{k} \frac{1}{n_i} = \frac{k}{2}\) since \(n_i = 2\) for \(1 \leq i \leq k\)

\[
m_{m} = n \left\{ \frac{3k}{2} - (t + 1) + \sum_{i=k+1}^{t} n_i \right\} + 1
\]
APPENDIX D

Copy of paper presented to British Acoustical Society

February 1970

February 1970.

A MICROELECTRONIC MULTIPLE BEAM FORMING NETWORK

A.J. Copping

Introduction

When an array of receiving transducers is illuminated by the echo from a reflecting object in the far field region of the array, the output of each transducer can be viewed as the real part of a phasor rotating at the carrier frequency. A phase difference related to the echo bearing will exist between the outputs of any adjacent pair of transducers, and due to the assumed plane wave nature of the echo, this phase difference will be constant across a linear array of equispaced elements, see figure 1.

Sonar receivers which measure the phase difference between adjacent elements using digital and analogue techniques have been described. Because these receivers reject amplitude information and rely only on the information obtainable from zero crossing positions, their performance is inferior to the more expensive within-pulse sector scanning sonar.

This paper describes an analogue and a digital technique for looking at an illuminated sector by creating a number of overlapping beams, the position of each of which is fixed in space, and which together cover the sector of interest. It is hoped that the use of microelectronic and thin film techniques will enable the production of an inexpensive sonar which does not suffer from the disadvantages which a 'phase only' sonar receiver has.

Background

Acoustic energy reflected back to an antenna will give rise to a continuous pressure distribution \( P(x) \) across the aperture. The outputs of an array of \( N \) transducers constitute a set of samples of \( P(x) \), from which the angular distribution \( f(0) \) of reflectors of acoustic energy in the far field region of the antenna is to be determined.

Only point samples need to be considered since the effect of the finite size of a transducer can be deduced from the pattern multiplication theorem for arrays of identical transducers.

Referring to figure 1(a), if a plane wave is arriving from an angle \( 0 \) to the normal, the phase difference \( \phi \) between element \( k \) and element \( 0 \) \( (k = 0, 1, 2, \ldots N - 1) \) is given by

\[
\phi = \frac{k \lambda}{2} \sin \theta; \quad \lambda = \text{wavelength}
\]
The contribution from a particular angle can be found by multiplying the array samples \( x_0, x_1, x_2, \ldots \) by an appropriate phase shifting factor and forming the sum

\[
f(\theta) = \sum_{k=0}^{N-1} x_k \exp(-j2\pi \frac{2k \sin \theta}{\lambda})
\]  

This function, although continuous, can be represented by \( N \) independent samples, spaced at intervals in \( \sin \theta \) of

\[
\Delta(\sin \theta) = \frac{\lambda}{N}.
\]

Thus \( f(\theta) \) can be expressed as a series of samples \( a_r \) where

\[
a_r = \sum_{k=0}^{N-1} x_k \exp(-j2\pi r \lambda \sin \theta) \quad \text{... (2)}
\]

and \( \sin \theta = \frac{r \lambda}{N} \quad (r = 0, 1, 2 \ldots N-1) \)

Expression (2) indicates the operation of taking the discrete Fourier transform (DFT) of the samples of the pressure distribution across the aperture.

Expanding (2) and putting \( w = \exp(-j2\pi r) \) gives

\[
\begin{align*}
a_0 &= x_0 + x_1 + x_2 + \ldots \ldots \\
a_1 &= x_0 + x_1 + x_2 + \ldots \ldots \\
a_2 &= x_0 + x_1 + x_2 + \ldots \ldots \\
&\vdots
\end{align*}
\]

The process of forming the samples \( a_r \) generates \( N \) beams whose directions are fixed in space relative to the array.

As an example, if the array consists of 8 elements spaced at intervals of \( \Delta \theta = \theta_0 \), 8 beams are formed at bearings defined by \( \sin \theta = 0, \frac{1}{N}, \frac{2}{N}, \frac{3}{N}, \frac{4}{N}, \frac{5}{N}, \frac{6}{N}, \frac{7}{N} \); this situation is illustrated in figure 2.

The DFT operation (3) is conveniently written in matrix notation

\[
A = W \cdot X
\]

where \( X \) is a column matrix having \( N \) elements which are the outputs of the array elements.

\( W \) is an \( N \times N \) square matrix, the element in the \( r \) th row and \( k \) th column is \( \exp(-j2\pi r k) \) \( r, k = 0, 1, 2 \ldots N-1 \).

\( A \) is a column matrix having \( N \) elements which are the outputs of the individual beams.

It has been shown that if \( N \) is not a prime number the matrix \( W \) may be factorized in such a way that the computational effort required to evaluate the DFT is considerably reduced. This procedure leads to the Fast Fourier Transform (FFT).

Only the case where \( N = 2^M \) will be considered here because this leads to a convenient result. Specifically, \( W \) may be written as the product of \( M \) factor matrices, each of order \( N \), but which have only two non-zero elements in each row

\[
W = W_1 \cdot W_2 \cdot W_3 \cdot \ldots \cdot W_M
\]

Rearrangement of the rows and columns of each factor matrix shows that the complete \( N \)-element DFT can be obtained
using $N \log_2 N$ 2-element transforms in conjunction with
$2 - N + N \log_2 N$ phase shifts.

The 2-element transform is defined by

$$W = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

This represents a device which takes the sum and difference of two inputs, and is particularly simple to implement in hardware form.

Table 1 illustrates the savings which may be achieved using the FFT factorization of the DFT for 4, 8, and 16 element arrays.

<table>
<thead>
<tr>
<th>No. of elements</th>
<th>DFT</th>
<th>FT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Additions and Subtractions</td>
<td>Multiplications</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>56</td>
<td>32</td>
</tr>
<tr>
<td>16</td>
<td>240</td>
<td>176</td>
</tr>
</tbody>
</table>

Table 1

As an example, consider the 4-element DFT

$$W = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & w & w^2 & w^3 \\ 1 & w^2 & 1 & w^3 \\ 1 & w^3 & w & 1 \end{bmatrix}$$  \[ w = \exp(-j\pi/2) \]

The individual elements have been written modulo 2.

The FFT factorization is

$$W = W_1 \cdot W_2 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & . & . \\ . & 1 & w & . \\ 1 & -1 & . & . \\ . & 1 & -w & . \end{bmatrix}$$

Rearranging the rows and columns of $W_1$ and $W_2$ gives


The operation $A = W.X$ is shown schematically in figure 3. Figure 3 shows how the spatial DFT can be implemented using digital and analogue hardware.

**Digital Realisation**

It is possible to use a general purpose digital computer to perform the FFT operation in a sonar receiver. However, in many situations, the greater speed of a special purpose FFT processor is required.
Referring to Figure 4, the outputs of the array are sampled, amplified, converted to digital form and stored. The FFT can now proceed in a number of ways depending on the processing time available.

The Sequential Processor

Each operation depicted by a block in Figure 3 is performed sequentially by a processor which has a single arithmetic unit. The time taken is that required to perform \( N \log_2 N \) complex additions and subtractions and \( 1 - N + N \log_2 N \) complex multiplications.

The Parallel Processor

The sequential processing time may be reduced by introducing as much parallel processing as is necessary. In the extreme case, one arithmetic unit could be provided for each block in Figure 3, and the effective computation time would be that required to perform just one 2-point transform.

The provision of several arithmetic units would be expensive at the present time, but the production of arithmetic units using large scale integration techniques may make such schemes more feasible in the future.

Analogous Realization

The schematic in Figure 3 can be implemented directly using analogue components to build a sum and difference network. It can be shown that any two port network having an output/input relationship expressible in the general form

\[
\begin{bmatrix}
  o_1 \\
o_2
\end{bmatrix} = \begin{bmatrix}
  e^{j\alpha_1} & 1 \\
e^{j\alpha_2} & 1
\end{bmatrix} \begin{bmatrix}
  i_1 \\
i_2
\end{bmatrix} \quad \ldots \quad (4)
\]

\[
= G \cdot \begin{bmatrix}
  i_1 \\
i_2
\end{bmatrix} \quad i_1, i_2 \text{ are inputs, } o_1, o_2 \text{ are outputs, } j^{\sqrt{2}}
\]

can be used as a basic building block in a multiple beam forming network for an array of \( N = 2^n \) elements.

The general expression (4) enables the phase shifts \( \alpha_1, \alpha_2 \) to be chosen independently to give a network which is convenient to fabricate.

For instance, if \( \alpha_1 = \pi, \alpha_2 = \alpha_3 = -\frac{\pi}{4} \), the network obtained is the hybrid coupler used in the Butler beam forming matrix at microwave frequencies.

One network of particular interest in sonar is defined by

\[
\begin{bmatrix}
  \alpha_1 = \pi, \alpha_2 = \alpha_3 = 0
\end{bmatrix}
\]

\[
i.e. \quad G = \begin{bmatrix}
  -1 & 1 \\
  1 & -1
\end{bmatrix} = \begin{bmatrix}
  -1 & -1
\end{bmatrix}
\]

since this may be fabricated using microelectronic operational amplifiers in conjunction with thin film resistors, see Figure 5.

The beam forming network is obtained by substitution into the network shown in Figure 3, using the relationship

\[
\begin{bmatrix}
  1 & 1 \\
  1 & -1
\end{bmatrix} = \begin{bmatrix}
  -1 & 0
\end{bmatrix}
\]

and then simplifying.
The equivalent multiple beam forming network using blocks G is shown in figure 6. If only the magnitudes of the beam outputs are required and their relative phases are unimportant, then the phase shifts of π radians in beam outputs a1 and a2 may be ignored.

The phase shifts required within the beam forming networks may be implemented using RC all-pass networks which can be fabricated using microelectronic amplifiers and thin film techniques, fig. 7.

Conclusion

Two schemes for processing the outputs of an array to produce multiple beams covering a sector have been described. Both are being developed at Loughborough University of Technology.

Through the use of low cost analogue and digital microelectronic circuits it is hoped that an inexpensive sector coverage sensor will soon become available.

References

8. COPPENS, A.J. 'Synthesis of Multiple-Beam Forming Networks.' Internal Memorandum, Dept. of Electrical Eng., Loughborough University.
Fig. 1
Array geometry and outputs of 4 elements

Fig. 2
Beams formed from 8 element array (1/2 spacing)

Fig. 3
FFT schematic
Fig. 4
Digital beamformer

Fig. 5
Basic analogue beamforming circuit

Fig. 6
Analogue beamformer
\[
\frac{\Delta \phi}{\Delta t} = 1
\]

where

\[
\tan \phi = \frac{-2\pi BR_1}{1 - \omega CR_1}
\]

and \(\omega = 2\pi \text{ frequency}\)

**Fig. 9**
Phase shift circuit
APPENDIX E

The Mean Square Value of an Ideal Reconstruction

An expression is derived, relating the mean square value of the output of the ideal reconstruction filter to the samples taken at the output of the DFT processor.

It is assumed that the number of array elements, \( n \), is equal to the number of inputs, and hence outputs, of the DFT network. Also, \( n \) is even, so that the period of the sequence at the input to the filter is \( 2T = 2\pi \). These conditions describe the prototype processor which was built.

Consider one cycle, \( e_1^*(t) \), of the sample sequence, \( e^*(t) \), at the input to the filter.

\[
e_1^*(t) = \sum_{r=0}^{2n-1} e_r \delta(t - rt)
\]  

where \( e_{r+n} = -e_r \).

Now

\[
e^*(t) = e_1^*(t) * \sum_{r=-\infty}^{\infty} \delta(t - r.2T)
\]  

and if

\( e^*(t) \leftrightarrow E^*(f) \) \hspace{1cm} (E.3)

and

\( e_1^*(t) \leftrightarrow E_1^*(f) \) \hspace{1cm} (E.4)

then from equation E.2

\[
E^*(f) = E_1^*(f) \cdot \frac{1}{2T} \sum_{k=-\infty}^{\infty} \delta(f - k) = \frac{1}{2T} \sum_{k=-\infty}^{\infty} \delta(f - \frac{k}{2T})
\]
The ideal reconstruction filter is specified in equation 4.35, repeated here,

\[
H(f) = \tau, \quad -(\frac{n-1}{2T}) - \Delta f < f < \left(\frac{n-1}{2T}\right) + \Delta f
\]

\[
= 0, \quad |f| > \left(\frac{n-1}{2T}\right) + \Delta f
\]

where \( \Delta f < \frac{1}{T} \).

Therefore, the spectrum, \( E(f) \), of the filter output, \( e(t) \), is given by

\[
E(f) = H(f) \cdot E_i^*(f)
\]

\[
= \frac{\tau}{2T} \sum_{k=-(n-1)}^{n-1} E_i^*(\frac{k}{2T}) \cdot \delta(f - \frac{k}{2T})
\]

i.e.

\[
E(f) = \frac{1}{2n} \sum_{k=-(n-1)}^{n-1} E_i^*(\frac{k}{2T}) \cdot \delta(f - \frac{k}{2T})
\]

\[\text{Equation E.7 is the Fourier series representation of } e(t) \).

Applying Parseval's theorem to the periodic function, \( e(t) \), in equation E.7 gives

\[
\frac{1}{2T} \int_{0}^{2T} |e(t)|^2 \, dt = \left(\frac{1}{2n}\right)^2 \sum_{k=-(n-1)}^{n-1} |E_i^*(\frac{k}{2T})|^2
\]

\[\text{Equation E.8} \]
Now, from equations E.1 and E.4

\[ E_k^*(f) = \int_{-\infty}^{\infty} \{ \sum_{r=0}^{2n-1} e_r \delta(t - rT) \} \exp(-j2\pi ft) \, dt \]

i.e.

\[ E_k^*(f) = \sum_{r=0}^{2n-1} e_r \exp(-j2\pi f rT) \quad - (E.9) \]

\[ \therefore E_k^*(\frac{kT}{2T}) = \sum_{r=0}^{2n-1} e_r \exp(-j\frac{2\pi k r}{2n}) \quad - (E.10) \]

since \( \frac{T}{T} = \frac{1}{n} \)

Equation E.10 shows that \( E_k^*(\frac{kT}{2T}) \) is the \( k \)th coefficient in the DFT of the sequence \( e_r \), \( (r = 0, 1, \ldots, 2n-1) \) of dimension \( 2n \). Parseval's Theorem applied to the discrete Fourier Transform gives

\[ \sum_{r=0}^{2n-1} |e_r|^2 = \frac{1}{2n} \sum_{k=0}^{2n-1} |E_k^*(\frac{kT}{2T})|^2 \]

\[ = \frac{1}{2n} \sum_{k=-n}^{n-1} |E_k^*(\frac{kT}{2T})|^2 \quad - (E.11) \]

since \( E_k^*(\frac{kT}{2T}) = E_k^*(\frac{k-2n}{2T}) \).

Now in equation E.11, \( E_k^*(\frac{-n}{2T}) = 0 \), since all even harmonics are zero in a waveform possessing half wave symmetry. Therefore, equation E.11 becomes

\[ \sum_{r=0}^{2n-1} |e_r|^2 = \frac{1}{2n} \sum_{k=-n}^{n-1} |E_k^*(\frac{kT}{2T})|^2 \quad - (E.12) \]
Combining equations E.12 and E.8 gives

\[ \frac{1}{2T} \int_0^{2T} |e(t)|^2 \, dt = \frac{1}{2n} \sum_{r=0}^{2n-1} |e_r|^2 \]

\[ = \frac{1}{n} \sum_{r=0}^{n-1} |e_r|^2 \]

- (E.13)

since \( e_{r+n} = -e_r \).

In section 4.6.2.2, a method was discussed which enabled guard bands to be introduced into the spectrum, \( S^*(f) \), of the sample sequence \( s^*(t) \). This was because some of the components in \( S^*(f) \) were forced to be zero, assuming an ideal processor. If errors occur due to imperfections in the processor, then the corresponding components in the spectrum, \( E^*(f) \), of the error sequence, \( e^*(t) \), will not be zero in general. Retracing the development in this Appendix for a processor with \( n_N \) outputs rather than \( n \), \((n_N > n)\), will give the mean square of the complex error waveform after the ideal filter as \( \frac{1}{n_N} \sum_{r=0}^{n_N-1} |e_r|^2 \), provided that the ideal filter bandwidth is wide enough to include all \( n_N \) spectral lines in \( E(f) \) rather than just the \( n \) lines present in \( S(f) \).
**APPENDIX F**

**A Polynomial Representation of the Ideal Reconstruction Filter Output**

The sequence, \( u^*(t) \), applied to the input of the reconstruction filter is given by

\[
\begin{align*}
u^*(t) = u_1(t) \ast \sum_{r=-\infty}^{\infty} \delta(t - r.2T)
\end{align*}
\]

where \( u_1(t) \) is one cycle of length \( 2T \).

\[
\begin{align*}
u^*(f) &= u_1^*(f) \cdot \frac{1}{2T} \sum_{k=-\infty}^{\infty} \delta(f - k \cdot \frac{1}{2T})
\end{align*}
\]

i.e.

\[
\begin{align*}
u^*(f) &= \sum_{k=-\infty}^{\infty} u_1^* \left( \frac{k}{2T} \right) \delta(f - \frac{k}{2T})
\end{align*}
\]

where \( u_1^* \left( \frac{k}{2T} \right) = \frac{1}{2T} u_1 \left( \frac{k}{2T} \right) \)

The spectrum of the output \( v(t) \) of the ideal reconstruction filter is

\[
\begin{align*}
v(f) = H(f) \cdot u^*(f)
\end{align*}
\]

where

\[
\begin{align*}
H(f) &= \tau , -(n-1) \frac{2T}{2T} - \Delta f < f < (n-1) \frac{2T}{2T} + \Delta f
\end{align*}
\]

\[
\begin{align*}
= 0 , |f| > (n-1) \frac{2T}{2T} + \Delta f
\end{align*}
\]

and \( \Delta f < \frac{1}{T} \).

Substituting equations F.2 and 4.35 into equation F.3 gives
\[ V(f) = \tau \sum_{k=-(n-1)}^{(n-1)} u^* \left( \frac{k}{2T} \right) \delta (f - \frac{k}{2T}) \]  \hspace{1cm} \text{(F.4)}

\[ v(t) = \tau \sum_{k=-(n-1)}^{(n-1)} u^* \left( \frac{k}{2T} \right) \exp \left( j \cdot \frac{2\pi kt}{2T} \right) \]  \hspace{1cm} \text{(F.5)}

Taking account of the fact that

\[ u^* \left( \frac{k}{2T} \right) = 0 \text{ for } k = 0 \text{ or an even number}, \]

equation F.5 may be rewritten as

\[ v(t) = \tau \cdot \exp \left( j \frac{\pi t(n-1)}{T} \right) \sum_{k=0}^{n-1} u^* \left( \frac{n-1}{2T} - \frac{k}{T} \right) \exp \left( -j \frac{2\pi kt}{T} \right) \]  \hspace{1cm} \text{(F.6)}

or

\[ v(t) = \tau \cdot \exp \left( j \frac{\pi t(n-1)}{T} \right) \sum_{k=0}^{n-1} c_k z_t^{-k} \]  \hspace{1cm} \text{(F.7)}

where

\[ c_k = u^* \left( \frac{n-1}{2T} - \frac{k}{T} \right) \]

and

\[ z_t = \exp(j \frac{2\pi t}{T}) \]

Equation F.7 expresses the output of the ideal filter as a polynomial in the complex variable \( z_t \). This is equivalent to the Fourier series in equation F.5.
APPENDIX G

The Calculation of the Digital Processor Error

Processors having only a small number of inputs can be examined in
detail quite easily, although the procedure would become tedious if the
number of inputs exceeded 16, say. Even so, only one extra rank of
addition, subtraction, and complex multiplication operations are required
in the course of doubling the number of inputs, and the results developed
here for 8 inputs can be extended for 16 inputs and so on quite simply.

Starting with the mean and mean square values of the error due to
channel noise and quantization at each input to the processor, equations
6.49, 6.50, 6.55, 6.56, 6.58, and 6.59, are used to generate expressions
for the mean and mean square values of the error after each rank of pro-
cessor operations. The 'in-place' computing schematic diagram in
Figure 5.3 is used.

Tables G.1, G.2, and G.3, list the error properties before the
first, before the second, and after the third rank of Σ/Δ units. An
extension to 16 inputs would use the results in Table G.3 as a starting
point. Table G.4 lists the mean square error $|e_r|^2$ at each output of
the processor after the compensating phase shifts, and these values were
obtained by combining equations 6.56 and 6.59 as follows.

$$|e_r|^2 = e_{rx}^2 + e_{ry}^2$$

since $e_r = e_{rx} + j e_{ry}$
Using equations 6.56 and 6.59 gives,

\[
|e_r|^2 = b_x^2 \gamma_{ax}^2 + b_y^2 \gamma_{ay}^2 + 2\gamma_m^2 - 2b_x b_y \gamma_{ax} \gamma_{ay}
+ b_x^2 \gamma_{ax}^2 + b_y^2 \gamma_{ay}^2 + 2\gamma_m^2 + 2b_x b_y \gamma_{ax} \gamma_{ay}
\]

\[
|e_r|^2 = \gamma_{ax}^2 + \gamma_{ay}^2 + 4\gamma_m^2
\]

since \( b_x^2 + b_y^2 = 1 \) for the phase shift \( b_x + jb_y \).

<table>
<thead>
<tr>
<th>Store Location</th>
<th>Mean Error</th>
<th>Mean Square Error</th>
</tr>
</thead>
</table>
| All            | 0          | \( \sigma_{ref}^2 + 0.08\kappa^2 \)

Table G.1

Error Properties at Input to Processor
<table>
<thead>
<tr>
<th>Store Location</th>
<th>Mean Error</th>
<th>Mean Square Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 R I</td>
<td>$-0.25\kappa$</td>
<td>$0.5\sigma^2_{\text{ref}} + 0.17\kappa^2$</td>
</tr>
<tr>
<td>1 R I</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>2 R I</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>3 R I</td>
<td>$-0.25\kappa$</td>
<td>&quot;</td>
</tr>
<tr>
<td>4 R I</td>
<td>$-0.25\kappa$</td>
<td>&quot;</td>
</tr>
<tr>
<td>5 R I</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>6 R I</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>7 R I</td>
<td>$-0.25\kappa$</td>
<td>&quot;</td>
</tr>
</tbody>
</table>

**Table G.2**

Error Properties at Input to 2nd Rank of $\Sigma/A$ Operations

$R = \text{Real part of complex number.}$

$I = \text{Imaginary part of complex number.}$
<table>
<thead>
<tr>
<th>Store Location</th>
<th>Mean Error</th>
<th>Mean Square Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>R - 0.75k</td>
<td>0.125σ² ref + 0.68k²</td>
</tr>
<tr>
<td></td>
<td>I - 0.75k</td>
<td>0.125σ² ref + 0.68k²</td>
</tr>
<tr>
<td>1</td>
<td>R - 0.77k</td>
<td>0.125σ² ref + 0.75k²</td>
</tr>
<tr>
<td></td>
<td>I - 0.29k</td>
<td>0.125σ² ref + 0.24k²</td>
</tr>
<tr>
<td>2</td>
<td>R - 0.50k</td>
<td>0.125σ² ref + 0.37k²</td>
</tr>
<tr>
<td></td>
<td>I - 0.25k</td>
<td>0.125σ² ref + 0.18k²</td>
</tr>
<tr>
<td>3</td>
<td>R - 0.46k</td>
<td>0.125σ² ref + 0.38k²</td>
</tr>
<tr>
<td></td>
<td>I - 0.23k</td>
<td>0.125σ² ref + 0.22k²</td>
</tr>
<tr>
<td>4</td>
<td>R - 0.25k</td>
<td>0.125σ² ref + 0.18k²</td>
</tr>
<tr>
<td></td>
<td>I - 0.25k</td>
<td>0.125σ² ref + 0.18k²</td>
</tr>
<tr>
<td>5</td>
<td>R - 0.23k</td>
<td>0.125σ² ref + 0.22k²</td>
</tr>
<tr>
<td></td>
<td>I - 0.47k</td>
<td>0.125σ² ref + 0.38k²</td>
</tr>
<tr>
<td>6</td>
<td>R - 0.25k</td>
<td>0.125σ² ref + 0.18k²</td>
</tr>
<tr>
<td></td>
<td>I - 0.50k</td>
<td>0.125σ² ref + 0.37k²</td>
</tr>
<tr>
<td>7</td>
<td>R - 0.29k</td>
<td>0.125σ² ref + 0.24k²</td>
</tr>
<tr>
<td></td>
<td>I - 0.77k</td>
<td>0.125σ² ref + 0.75k²</td>
</tr>
</tbody>
</table>

Table G.3

Error Properties at Output of the 3rd Rank of E/Δ Operations

R = Real part of complex number.
I = Imaginary part of complex number.
| Store Location (r) | Mean Square Error $|e_r|^2$ |
|-------------------|-----------------|
| 0                 | $0.25\sigma^2_{ref} + 1.36\kappa^2$ |
| 1                 | $0.25\sigma^2_{ref} + 1.52\kappa^2$ |
| 2                 | $0.25\sigma^2_{ref} + 0.89\kappa^2$ |
| 3                 | $0.25\sigma^2_{ref} + 0.92\kappa^2$ |
| 4                 | $0.25\sigma^2_{ref} + 0.70\kappa^2$ |
| 5                 | $0.25\sigma^2_{ref} + 0.92\kappa^2$ |
| 6                 | $0.25\sigma^2_{ref} + 0.89\kappa^2$ |
| 7                 | $0.25\sigma^2_{ref} + 1.32\kappa^2$ |

$$\sum_{r=0}^{7} |e_r|^2 = 2.0\sigma^2_{ref} + 8.32\kappa^2$$

Table G.4

Mean Square Error at Output of Processor
8-input analogue processor
Component Tolerance Calculations on an Analogue DFT Network

The analogue network, with the operations shown within each block, is depicted in Figure H.1. The additional subscript '2', which was used in Chapter 3 to denote a unit having 2 inputs and outputs, has been omitted for clarity, and here,

\[ G_i = \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix} \]  \hspace{1cm} (H.1)

The operation of the network in Figure H.1 is described by the matrix, \( N \), whose coefficients (\( m_{rk} \) for \( r, k = 0, 1, \ldots, 7 \)) can be obtained by inspection of the figure. To determine \( m_{rk} \), multiply together all the operations in the path between output \( r \) and input \( k \). This procedure is illustrated for \( m_{11} \) in Figure H.2, giving

\[ m_{11} = r_6 \cdot b_{10} \cdot r_3 \cdot a_8 \cdot c_3 \]  \hspace{1cm} (H.2)

The modulus of each matrix coefficient is required in equations 6.77 and 6.79, repeated here for \( n = 8 \) as equations H.3 and H.4.

\[
\overline{e_1^2} = \frac{\sigma^2}{4} \sum_{r=0}^{7} \sum_{k=0}^{7} |m_{rk}|^2 \]  \hspace{1cm} (H.3)

\[
\sigma^2_{\text{net}} = \frac{\sigma^2}{16} \left\{ \sum_{r=0}^{7} \sum_{k=0}^{7} |m_{rk}|^2 - \left| \sum_{r=0}^{7} |m_{rk}|^2 \right|^2 \right\} \]  \hspace{1cm} (H.4)

Each term in the product defining \( m_{rk} \) can be expended as an expression involving the circuit elements by using the equations in section 6.5.1.
Figure H.2

Components in $m_{11}$
In any Σ/Δ unit, from equations 6.64 and 6.67,

\[ a = -\frac{1}{1 + \mu_s} \cdot \frac{R_3}{R_1} \]  

\( \mu_s \), defined in equation 6.68, is a function of the voltage gain, \( A_v \), of the amplifier and also of circuit resistance values which include \( R_3 \) and \( R_1 \), all of which are subject to change. However, the magnitude of the variations in \( A_v \), shown in Figure 6.17, are such that the effect of the resistance value variations on \( \mu_s \) are very small by comparison. Passive components will normally be tolerated to ±2% or better, and variations in the ratios \( R_3/R_1 \) and \( R_0/R_3 \) in equation 6.68 are small enough to be ignored. Consequently, \( (1/1 + \mu_s) \) will be considered to be independent of variations in \( R_3 \) and \( R_1 \) in equation H.5.

Each term in equation H.5 is expressed now as a mean value in conjunction with an associated deviation from the mean value. Thus,

\[ |a| = K_S (1 + \Delta_S) \cdot \frac{R_N (1 + \Delta_3)}{R_N (1 + \Delta_1)} \]  

where \( K_S = \left(\frac{1}{1 + \mu_s}\right) \), and nominally \( R_3 = R_1 = R_N \cdot \Delta_S \), \( \Delta_3 \), and \( \Delta_1 \) are fractional deviations from the mean of \( K_S \), \( R_3 \), and \( R_1 \), respectively.

The behaviour of \( \Delta_S \) is described in section 6.5.3.1, but it is sufficient here that \( |\Delta_S| < 0.02 \), and also \( |\Delta_3| \leq 0.02 \) and \( |\Delta_1| \leq 0.02 \). Therefore, to an adequate approximation, equation H.6 can be written as

\[ |a| = K_S (1 + \Delta_S) (1 + \Delta_3) (1 - \Delta_1) \]

or

\[ |a| = K_S (1 + \Delta_a) \]  

- (H.7)
where \( \Delta_a = \Delta_S + \Delta_3 - \Delta_1 \) \hspace{1cm} (H.8)

In a similar manner, expressions can be developed for \(|b|\), \(|c|\), and \(|d|\).

\[
|b| = K_S (1 + \Delta_b) \hspace{1cm} (H.9)
\]

where \( \Delta_b = \Delta_S + \Delta_3 - \Delta_2 \) \hspace{1cm} (H.10)

\[
|c| = K_D (1 + \Delta_c) \hspace{1cm} (H.11)
\]

where \( \Delta_c = \Delta_D + \frac{1}{2}(\Delta_7 - \Delta_5 + \Delta_8 - \Delta_6) \) \hspace{1cm} (H.12)

Here \( K_D = \frac{1}{1 + |\Delta_d|} \), and \( \Delta_D \) is the fractional deviation from the mean.

\[
|d| = K_D (1 + \Delta_d) \hspace{1cm} (H.13)
\]

where \( \Delta_d = \Delta_D + \Delta_7 - \Delta_5 \) \hspace{1cm} (H.14)

For the phase shift circuit, from equations 6.65, 6.71 and 6.72, it can be seen that \(|r|\) is a function of \( \omega_C R \). Taking the worst case for both circuits (\( \omega_C R \gg 1 \) in equation 6.71 and \( \omega_C R \ll 1 \) in equation 6.72) leads to a single expression for \(|r|\) as

\[
|r| = K_D (1 + \Delta_r) \hspace{1cm} (H.15)
\]

where \( \Delta_r = \Delta_D + \Delta_11 - \Delta_9 \) \hspace{1cm} (H.16)

All the terms in the product which constitutes the coefficient \( m_{rk} \) are independent of each other since they are contributed by different
units. So, taking \( m_{11} \) in equation H.2 as an example,

\[
|m_{11}|^2 = |r_6|^2 |b_{10}|^2 |r_9|^2 |a_8|^2 |c_3|^2 \quad - \text{(H.17)}
\]

\[
= K_D^2 (1 + \Delta r_6)^2 \cdot K_S^2 (1 + \Delta b_{10})^2 \ldots K_D^2 (1 + \Delta c_3)^2 \quad - \text{(H.18)}
\]

since \( |r_6| = K_D (1 + \Delta r_6) \) etc.

Equation H.18 becomes

\[
|m_{11}|^2 = K_S^4 K_D^6 (1 + \Delta r_6^2) (1 + \Delta b_{10}^2) (1 + \Delta r_9^2) (1 + \Delta a_8^2) (1 + \Delta c_3^2) \quad - \text{(H.19)}
\]

since \( \Delta r_6 = 0, \Delta b_{10} = 0 \), etc.

Equation H.19 can now be written in terms of \( \Delta r_6^2 \) and \( \Delta b_{10}^2 \) contributed by the amplifiers, and the mean square value of the fractional error in each passive component value. If all components have the same tolerance limits of \( \Delta T \), then \( \Delta r_6^2 = \Delta b_{10}^2 = \ldots = \Delta a_8^2 = \Delta c_3^2 = \frac{\Delta T^2}{3} \), for a uniform distribution of values within the limits.

From equations H.8, H.10, H.12, H.14, and H.16

\[
\Delta T^2 = \Delta S^2 + \Delta D^2 + \Delta r_6^2 + \Delta b_{10}^2 + \Delta r_9^2 + \Delta a_8^2 + \Delta c_3^2 + \frac{2 \Delta T^2}{3} \quad - \text{(a)}
\]

\[
\Delta T^2 = \Delta S^2 + \Delta b_{10}^2 + \Delta r_9^2 + \Delta D^2 + \Delta r_6^2 + \frac{2 \Delta T^2}{3} \quad - \text{(b)}
\]

\[
\Delta T^2 = \Delta D^2 + \Delta r_9^2 + \frac{1}{4} (\Delta r_6^2 + \Delta b_{10}^2 + \Delta c_3^2 + \Delta a_8^2 + \Delta r_9^2) + \frac{2 \Delta T^2}{3} \quad - \text{(c)}
\]

\[
= \Delta T^2 + \frac{\Delta T^2}{3}.
\]
\[
\Delta_d^2 = \Delta_D^2 + \Delta_7^2 + \Delta_5^2 = \frac{\Delta_D^2}{3} + \frac{2\Delta_T^2}{3} \quad \text{(d)}
\]
\[
\Delta_r^2 = \Delta_D^2 + \Delta_1^2 + \Delta_9^2 = \frac{\Delta_D^2}{3} + \frac{2\Delta_T^2}{3} \quad \text{(e)}
\]

From section 6.5.3.1, \(\Delta_S^2 = 6 \times 10^{-6}\), and \(\Delta_D^2 = 3 \times 10^{-6}\).

For ±2% component tolerance, \(\frac{2\Delta_T^2}{3} = 0.0003\)

\[
\therefore \quad \Delta_a^2 = \Delta_b^2 \approx 0.0003
\]
\[
\Delta_c^2 \approx 0.0001
\]
\[
\Delta_d^2 = \Delta_r^2 \approx 0.0003
\]

Substituting these values into equation H.19 gives

\[
|m_{11}|^2 \propto K_S^4 K_D^6
\]

to better than 0.2%.

Applying this procedure to all coefficients, \(m_{rk}\), and using equation H.3, gives

\[
\epsilon_1 = \frac{\sigma^2}{4} \{ 8 K_S^6 + 14 K_S^4 K_D^4 + 8 K_S^4 K_D^6 + 2 K_S^4 K_D^8 + 8 K_S^2 K_D^6 + 12 K_S^2 K_D^8 + 4 K_S^2 K_D^{10} + 2 K_D^8 + 4 K_D^{10} + 2 K_D^{12} \} \quad \text{- (H.21)}
\]

From section 6.5.3.1, \(K_S = 1.004\), and \(K_D = 1.003\).
Substitution of these values into equation H.21 gives,

\[ \frac{1}{\epsilon^2} = 16.5 \sigma_n^2 \quad - (H.22) \]

Returning to equation H.2,

\[ |m_{11}|^2 = |r_6|^2 |b_{10}|^2 |r_3|^2 |a_8|^2 |c_3|^2 \]

\[ \therefore |m_{11}|^2 = K_S^h K_D^e (1+\Delta_{r_6})^2 (1+\Delta_{b_{10}})^2 (1+\Delta_{r_3})^2 (1+\Delta_{a_8})^2 (1+\Delta_{c_3})^2 \quad - (H.23) \]

Combining equations H.23 and H.19,

\[ |m_{11}|^2 = |\overline{m_{11}}|^2 = K_S^h K_D^e \{2(\Delta_{r_6} + \Delta_{b_{10}} + \Delta_{r_3} + \Delta_{a_8} + \Delta_{c_3}) \} \quad - (H.24) \]

Higher order terms on the R.H.S. of equation H.24 are not worth retaining since they become insignificant after the squaring and averaging process indicated in equation H.4. At this point it is possible to make the approximation \( K_S^h K_D^e \approx 1 \), as with all similar terms associated with the other \( m_{rk} \). This approximation is not strictly necessary in order to proceed, but it does make the situation clearer since it avoids calculating the coefficients associated with each \( \Delta \) term when the summation in equation H.4 is generated from expressions similar to equation H.24 for all \( m_{rk} \). The error in \( \sigma_{net} \), caused by this approximation, is estimated to be less than 4%.

So, for any \( m_{rk} \), \( \{|m_{rk}|^2 - |\overline{m_{rk}}|^2\} = 2 \times (\text{the sum of all } \Delta \text{ terms associated with } m_{rk}) \). Now, any particular \( \Delta \) term will occur in more than one element \( m_{rk} \). For example, \( \Delta_{r_6} \) occurs in in occurs in all elements in row 1 in matrix \( M \). Inspection of Figure H.1 shows
that the terms \( a, b, c, \) and \( d \) associated with the \( \Sigma/\Delta \) unit \( G \), (equation H.1) will each be associated with four separate elements in matrix \( M \); and the term \( r \) from any one of the twelve phase shift units will be associated with eight separate elements in \( M \). Consequently, in the summation in equation H.4, the coefficient associated with any \( \Delta_a, \Delta_b, \Delta_c, \Delta_d \) is \( 2 \times 4 \), and that associated with any \( \Delta_r \) is \( 2 \times 8 \).

\[
\sum_{r=0}^{7} \sum_{k=0}^{7} \left[ |m_{rk}|^2 - \left| \frac{m_{rk}}{2} \right|^2 \right] = 8 \sum_{i=1}^{12} (\Delta_{ai} + \Delta_{bi} + \Delta_{ci} + \Delta_{di} + 2\Delta_{ri}) \quad - (H.25)
\]

After squaring and taking the mean of equation H.25, equation H.4 becomes

\[
\sigma^2_{\text{net}} = 4\sigma^4 \sum_{i=1}^{12} \left( \Delta_{ai}^2 + \Delta_{bi}^2 + \Delta_{ci}^2 + \Delta_{di}^2 + 4\Delta_{ri}^2 \right) + 2\Delta_{ai}\Delta_{bi} + 2\Delta_{ci}\Delta_{di} \quad - (H.26)
\]

The last two terms in equation H.26 occur because \( \Delta_{ai} \) and \( \Delta_{bi} \) contain common components, as do \( \Delta_{ci} \) and \( \Delta_{di} \). The mean of all other cross products is zero.

From equations H.8 and H.10,

\[
\overline{\Delta_a \Delta_b} = \overline{\Delta_a^2} + \overline{\Delta_b^2} = \frac{\Delta_a^2}{3} + \frac{\Delta_b^2}{3}
\]

and

\[
\overline{\Delta_c \Delta_d} = \overline{\Delta_c^2} + \frac{1}{4}(\overline{\Delta_c^2} + \overline{\Delta_d^2}) = \frac{\Delta_c^2}{7} + \frac{\Delta_d^2}{3}
\]

- (H.27)
Using equations H.20 and H.27, equation H.26 becomes

\[ \sigma_{\text{net}}^2 = 4 \sigma_n^4 \left( 48 \Delta_S^2 + 96 \Delta_D^2 + 76 \Delta_T^2 \right) \]  

- (H.28)