RESOLUTION IMPROVEMENT IN ACOUSTIC HOLOGRAPHY USING APERTURE EXPANSION TECHNIQUES

Volume 1

by

RADWAN E. S. ABDEL-AAL

THESIS

Presented to the University of Strathclyde

Department of Electronic and Electrical Engineering

for the degree of

Doctor of Philosophy

June 1983

ABSTRACT

One of the main reasons for the restricted application of holography in acoustic imaging is the limited resolution due to the relatively small numerical aperture for a given physical aperture. There is also a lack of suitable acoustic area detectors which provide adequate sensitivity and spatial resolution and which can be extended over large areas. The number of points at which the hologram can be sampled is limited by time or cost considerations. A requirement therefore exists for investigating signal processing techniques for resolution improvement in acoustic holography. Fortunately, the availability of linear detectors in acoustics and the relatively small amount of available data make such techniques easier to implement, especially with the increased availability of fast, efficient, and cheap computers and signal processing devices.

This thesis treats the limited resolution in holography as one facet of the basic problem of diffraction and wavelength limitations on the resolution of imaging systems. Although a number of techniques have been reported in the literature for resolving beyond the diffraction limit both in optics and microwaves, these suffer from sensitivity to noise, the requirement for a priori information on the object or the need for long computation times. The thesis proposes a new method for resolution improvement by aperture expansion using the principle of analytic continuation. This method has the advantages of computation simplicity, versatility, and robustness against noise. In the proposed technique, the hologram function is modelled using the data at the available limited aperture and the model is used to predict new points outside this aperture. A number of predictive models are discussed together with a method for correcting the predicted data. The effect of disturbing noise is then considered. Simulation results are presented for both noise-free and noisy data when imaging single and multiple point objects and the extension of the technique to the imaging of continuous objects is discussed. Examples for doubling the size of an aperture in the presence of 30% relative noise are given. An experimental holographic imaging system is described which has been designed and implemented to allow verification of the proposed aperture expansion algorithm on realistic measured holograms.

CONTENTS

Page No

ABSTRACT i

CHAPTER 1 : INTRODUCTION

1.1 Acoustic Holography 1
1.2 Resolution Improvement in Acoustic Holography 4
1.3 Aims, Layout, and Contributions of the Thesis 6

1.3.1 Aims of the Thesis 6
1.3.2 Layout of the Thesis 6
1.3.3 Contributions of the Thesis 10

CHAPTER 2 : A REVIEW OF ACOUSTIC HOLOGRAPHY

2.1 Introduction 13
2.2 Historical Background 14
2.3 Principles of Holography 18

2.3.1 Imaging with Wavefront Reconstruction 18
2.3.2 Gabor's On-Axis Holography 21
2.3.3 Leith-Upatneiks Off-Axis Holography 22
2.3.4 Holographic Images 24

2.4 Hologram Acquisition Techniques 27

2.4.1 Some Aspects of Hologram Sampling 28

- The Sampling Frequency 28
- Effect of Detector Size on Hologram Sampling 32

2.4.2 Electronic Reference in Acoustic Holography 33

- Measurement of the Complex Field Hologram 34
- Simulation of a Plane Wave Acoustic Reference 34

2.4.3 Review of Basic Methods of Hologram Acquisition 36

- Liquid Surface (Static Ripple) Holography 37
- Solid-Surface (Dynamic Ripple), Laser Beam Scanning Technique 37
- Hologram Recording with the Ultrasonic Camera 38
- Hologram Sampling with Arrays 39
- Mechanical Scanning 40

2.5 Image Reconstruction Techniques 41

2.5.1 Optical Reconstruction 41
2.5.2 Computer Reconstruction 43

- The Fresnel Integral Reconstruction Algorithm 48
- Backward Wave Propagation (Frequency Domain) Reconstruction Algorithm 52

2.6 Design Considerations in Holographic Imaging Systems 57

2.6.1 Resolution 57
2.6.2 Range 59
2.6.3 Field of View 59
2.6.4 Image Parameters 60

Page No

2.7 Holography Versus Other Acoustic Imaging Techniques 62

2.7.1 Holography Versus B-Scan Techniques 64
2.7.2 Holography Versus Other Orthoscopic Imaging Techniques 66

. Holography Versus Lens Systems 67
. Holography Versus Beamforming Systems 68

2.8 Applications 70

2.8.1 Nondestructive Testing (NDT) 70
2.8.2 Underwater Imaging 72
2.8.3 Medical Diagnostics 73
2.8.4 Seismic Imaging 74

2.9 Limitations 76
2.10 Conclusion 80

CHAPTER 3 : TECHNIQUES FOR RESOLUTION IMPROVEMENT

3.1 Introduction 81
3.2 Aperture Limited Resolution in Acoustic Holography 83
3.3 Synthetic Aperture Holographic Techniques 86
3.4 Object Restoration Techniques 90
3.5 Signal Processing Techniques for Resolution Imrpovement in Acoustic Holography 99

3.5.1 Resolution Improvement Beyond the Diffraction Limit 99
3.5.2 Resolution Improvement Beyond the Wavelength Limit 103

3.6 Extrapolation and Data Fitting Techniques 107
3.7 Conclusion 111

CHAPTER 4 : EXPANSION OF HOLOGRAM APERTURES USING PREDICTION TECHNIQUES

4.1 Introduction 113
4.2 Analyticity of the Fresnel Hologram Function 116
4.3 Hologram Simulation and Expansion - Basic Definitions 122
4.4 The Polynomial, Space-Variant, Predictive Model 125

4.4.1 The Power Series Polynomial Model 125
4.4.2 The Effect of Object Range 128
4.4.3 The Fourier Series Model 129

4.5 Error Analysis 132
4.6 The Linear, Space-Invariant, Predictive Model 136
4.7 The Hybrid, Polynomial/Linear, Predictive Model 138
4.8 The Corrective Model 142

4.8.1 The Polynomial, Space-Variant, Corrective Model 143
4.8.2 The Hybrid, Polynomial/Linear, Corrective Model 144

Page No

4.9	The Effect of Aperture Expansion on Range Resolution	147
4.10	The Effect of Increased Object Complexity	149
	4.10.1 Effect of Increasing the Point Spacing in a 2-point Object	149
	4.10.2 Effect of Increasing the Number of Points Representing an Object of a Given Width	149
	4.10.3 Imaging of Quasi-Continuous Objects	153
4.11	Conclusion	154
CHAPTER 5 :	APERTURE EXPANSION IN THE PRESENCE OF DISTURBING NOISE	
5.1	Introduction	157
5.2	Effect of Noise on the Polynomial Predictive Model	161
	5.2.1 Effect of Noise and Distance Errors on the Measured Hologram Signal	161
	5.2.2 Effect of Noise and Distance Errors on the Model Coefficients	164
5.3	Sensitivity to Noise Versus Model Type and Composition	171
5.4	Stability of the Square and the Triangular Polynomial Models	173
5.5	The Triangular, Polynomial, Predictive Model	179
5.6	The Triangular, Polynomial, Corrective Model	181
5.7	Imaging of Multiple Points in the Presence of Noise	184
5.8	Optimum Size of Available Aperture and the Effects of Computation Accuracy	187
	5.8.1 Effects of Computation Accuracy on the Performance of the Square Polynomial Predictive Model	190
	5.8.2 Effects of Computation Accuracy on the Performance of the Triangular Polynomial Predictive Model	192
5.9	Range Resolution of Apertures Expanded in the Presence of Noise	194
5.10	Imaging of Objects with Increased Complexity in the Presence of Noise	195
5.11	Conclusion	196
CHAPTER 6 :	EXPERIMENTAL HOLOGRAPHIC IMAGING SYSTEM	
6.1	Introduction	198
6.2	The Microcomputer System	200
6.3	The Hologram Acquisition System	202
	6.3.1 Mechanical Scanning Arrangement	203
	6.3.2 Transmission and Range Gating Circuits	204
	6.3.3 Phase and Amplitude Acquisition Circuits	205
	6.3.4 Post-Reception Data Processing and Display	208
	6.3.5 Interface with the Electrostatic Picture Recorder	210
	6.3.6 GPIB Interface with the Tektronix 4051 Minicomputer	211

		Page No
6.4	The Colour Display System	212
6.4.1	General Description	212
6.4.2	Economic Constraints on the Display Memory	214
6.4.3	The Display Software	216
6.5	Software for Hologram Simulation and Image Reconstruction	219
6.5.1	General Description	219
6.5.2	The Two-Dimensional FFT Algorithm (DFT)	220
6.5.3	Mathematical Functions Routines in Floating Point Format	222
CHAPTER 7 : EXPERIMENTAL RESULTS		
7.1	Introduction	225
7.2	Experimental Set-up and the Problem of Hologram Tilt	226
7.2.1	Effect of Tilt on Hologram Recording	226
7.2.2	Effect of Tilt on Image Reconstruction	227
7.2.3	Compensating for the Hologram Tilt	229
7.3	Image Reconstruction from Measured Holograms	231
7.4	Expansion of Measured Hologram Apertures	234
7.5	Simulation and Reconstruction of Two-Dimensional Holograms	238
7.6	Conclusion	241
CHAPTER 8 : CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK		
8.1	Conclusions	243
8.2	Suggestions for Future Work	250
8.2.1	Suggested Future Work Related to Theoretical Aspects	250
	. Modelling Techniques	250
	. Investigating the Effects of Errors in the Hologram Data on Image Reconstruction	251
8.2.2	Suggested Future Work Related to Experimental Aspects	252
ACKNOWLEDGEMENTS		254
REFERENCES		255

CHAPTER 1
INTRODUCTION

1.1 Acoustic Holography

Holography is a coherent imaging technique based on the principle of wavefront reconstruction. Although the two mechanisms of interference and diffraction which form the basis of the technique have been understood for many years, it was only in 1948 that Gabor suggested their application to three-dimensional imaging in optics as an ingenious way of preserving the complex spatial distribution of a given wavefront, storing it, and using it for later reconstruction. The practical application of the technique on a wide scale had to wait until the advent of laser in the early sixties which provided the required coherent sources. The success achieved with optical holography has encouraged research in other regions of the electromagnetic spectrum and in acoustics which has become the main area of the application of holography outside the optical domain. The early techniques of acoustic holography were replicas of those used in optics, as in the case of the liquid surface method for example. However, operation at the acoustic frequencies offers a number of advantages which have simplified the imaging process considerably and led to a great diversity of the techniques used for hologram recording and image reconstruction. For example, the physical reference wave which is necessary in conventional holography can be replaced by an electronically simulated reference. Moreover, the process of recording the hologram as an interference pattern for optical reconstruction can even be avoided completely, with the image calculated by the computer from a set of linearly detected hologram data.

Although acoustic imaging was already a well-developed art at the advent of holography, a large amount of interest has been shown in the holographic approach to acoustic imaging. This is because the holographic concept enjoys a number of advantages over conventional acoustic imaging techniques. For example, holography offers a considerable simplification in the data collection stage of the imaging process since only the complex spatial distribution (as the amplitude and phase data or the real and imaginary components) at the hologram plane is required. In particular, the fact that the temporal aspects of the received acoustic signal are discarded at an early stage of the process and replaced by a set of DC values makes holography more advantageous compared to beamforming techniques which rely on time delay information for image reconstruction. Since the processing in such systems must be performed at the acoustic frequency, they are generally more complex, bulky and costly and consume larger amounts of power in comparison with equivalent holographic systems. Holography also has a very large depth of field since it does not rely on focusing during hologram acquisition. In that sense, a two-dimensional hologram is a compact record of a three-dimensional volume in the imaged space within which any plane can be brought into focus when the hologram is reconstructed. In conventional imaging techniques, the number of planes at which focusing can be achieved is often restricted by difficulties in focusing the received waves. Moreover, since in holography the computer can be used for image reconstruction, this reduces the amount of special purpose hardware required and allows for a considerable amount and a wide range of signal processing techniques to be performed to improve the image quality and to analyse, display, store, and compare the resulting images.

Acoustic holography has been used in a number of applications which include : microscopy, nondestructive testing, underwater viewing, seismic imaging, and medical diagnostics. However, the extent to which holography has been implemented in practical systems has been rather limited in relation to the large amount of interest shown in the technique and considering the potential advantages it offers. This can be partially attributed to a number of limitations due to the coherent nature of holography and to operating at the acoustic frequency. As a coherent imaging technique, holography suffers from the problem of speckle and other artefacts which reduce the intelligibility of the resulting images. The main problem caused by operating in the acoustic domain is the limited image resolution because of the relatively small numerical apertures for a given physical aperture due to the large wavelength. This thesis is mainly concerned with this problem and proposes a new technique for resolution improvement.

A number of other limitations in holography are associated with the conventional approach of recording the hologram as an interference pattern which is optically reconstructed. The time-lag involved in preparing the optical transparancy is often unacceptable in situations which require real-time imaging. Moreover, optically reconstructed images are degraded by the distortion due to the large disparity between the wavelengths used for recording and reconstruction and the interference from unwanted reconstruction products which are the result of using a spatial carrier reference wave at the recording stage. These problems can be largely overcome by measuring the complex hologram directly and using the computer to perform the linear transforms required for image reconstruction. With the increase in the availability of high speed, large capacity, and inexpensive computers and signal processing devices, the computer can play a major role in improving the performance of holography by solving the problems associated with the conventional approach, performing the signal processing required for alleviating more basic problems such as the limited resolution, and integrating the various aspects of the holographic imaging process.

1.2 Resolution Improvement in Acoustic Holography

The requirement for resolution improvement in acoustic holography is basically dictated by the limited size of the numerical aperture for a given physical aperture. This problem is aggravated even further by practical limitations which restrict the size of physical apertures in acoustics with the required sensitivity and spatial resolution and which can be afforded economically or scanned at a reasonable speed. Fortunately, the very cause of the problem, i.e. operating at the acoustic frequencies which are much lower compared to those in optics, makes it easier to implement techniques for overcoming it. Linear detectors, which are readily available in acoustics, are capable of sensing the complex acoustic field and therefore allow for a wider range of possibilities for manipulating the data for image improvement compared to the case of optics where only intensity detectors are available. The limited numerical aperture, a disadvantage in itself, helps make data processing for resolution improvement a more feasible proposition since the relatively small amount of data can be manipulated speedily and cost-effectively by mini or microcomputers.

The problem of extending the resolving power beyond that of a given physical aperture has been extensively treated in the fields of optics and microwaves. Since the propagation part of the imaging process can be approximated to that of Fourier transforming the object function, this problem amounts to that of extrapolating the spectrum of a given finite object beyond the limit set by the size of the physical aperture as implied in object restoration techniques in optics. In spectral analysis, the dual problem is encountered where it is required to estimate the spectrum of a band-limited signal by extrapolating the corresponding time function beyond a given time window. Noise in the measured data sets the limit on the amount of resolution improvement that can be achieved and methods vary widely in their robustness against noise and therefore in their effectiveness when used on realistic data. Although a number of techniques have been reported in the literature for resolution improvement in acoustic holography, work in this direction appears to be somewhat isolated from the mainstream of the relevant activities in the fields of optics and spectral analysis. In this thesis, the relevant techniques for resolution improvement in the various fields are described and the prospects for their application in holography discussed. A new method for resolution improvement beyond the diffraction limit through expanding the hologram aperture is then presented. This method has the advantage of being computationally simple and robust against noise.

1.3 Aims, Layout, and Contributions of the Thesis

1.3.1 Aims of the Thesis

This thesis reports the results of a research program carried out with the following aims in mind.

1) Investigating the limitations of acoustic holography in general together with the methods which could be used to overcome these limitations, with particular reference to the role of the computer and the use of signal processing techniques.

2) Investigating the problem of limited resolution in acoustic holography as it relates to the basic problems of diffraction and wavelength limitations on the resolution of other imaging techniques.

3) Investigating the techniques for resolution improvement beyond the diffraction limit in the fields of optics, microwaves, and acoustics.

4) The development of a technique for improving resolution in acoustic holography which is computationally simple, more versatile, and robust against noise.

5) Verifying the proposed method using both simulated and experimentally measured holograms.

1.3.2 Layout of the Thesis

Following this introductory chapter, Chapter 2 of the thesis reviews the history, principles, and techniques of holography. The methods for hologram acquisition and image reconstruction are discussed, with particular emphasis on aspects of hologram sampling and computer reconstruction due to their relevance to the simulation and experimental work reported in later chapters. In section 2.5.2. which deals with computer reconstruction, mathematical treatment is given for both the Fresnel integral and the backward wave (frequency domain) reconstruction algorithms. The latter method is used for image reconstruction from both simulated and measured data which are reported in Chapters 4,5, and 7. The main design parameters of an acoustic holographic imaging system; such as resolution, range, and field of view are discussed with emphasis on the way in which each of these parameters is influenced by other parameters and the various trade-offs involved. The advantages of holography as an imaging system are then considered and acoustic holography is compared with other acoustic imaging techniques including B-scan, lens systems, and beamforming systems. The main applications of acoustic holography are reviewed and the limitations of the technique discussed, together with ways in which the use of the computer and the implementation of signal processing techniques can help alleviate some of these limitations.

Chapter 3 is devoted to reviewing the techniques for resolution improvement in the fields of optics and microwaves, and in acoustic holography. This chapter starts with a brief analysis of the aperture-limited resolution in holography. The first group of techniques discussed are those related to aperture synthesis. The problem of resolution improvement beyond the Rayleigh diffraction limit of a given aperture in optics is discussed together with some historical background . A group of techniques for object restoration are then considered with emphasis on the relationship between this approach and resolution improvement in acoustic holography. For comparison between the various techniques, simplicity of the computations and robustness against noise have been prime considerations since these tend to limit the usefulness of such techniques in practical situations. Techniques for improving resolution in acoustic holography both beyond the diffraction limit and the wavelength limit are reviewed with emphasis on aperture expansion methods. Due to their relevance to the aperture expansion method described in the remainder of the thesis, the main data fitting and function extrapolation algorithms are breifly discussed.

Chapter 4 presents a description of the new aperture expansion technique and contains the simulation tests performed on noiseless hologram data together with the relevant error analyses. The hologram function for a finite object in the Fresnel zone is shown to be analytic and therefore can be extended in space. A number of basic polynomial, linear, and hybrid models which represent the hologram signal over the available limited aperture are described and the results obtained when the models are used to predict new points outside this aperture are given for the case of simple 1-point and 2-point objects. Examples showing resolution distances of 0.25 the Rayleigh limit in the case of 2-points are given. For simplicity, only line holograms are considered. An error analysis based on Taylor-series expansion is presented and a method for correcting the predicted data to improve on the prediction accuracy is suggested. Aperture expansion is shown to yield the expected improvement in range resolution. The technique is then extended to imaging more complex objects consisting of a number of discrete points which simulate quasi-continuous objects.

The performance of the aperture expansion algorithm in the presence of various forms of noise is considered in Chapter 5 together with the modifications required to the models described in Chapter 4 in order to achieve stability with noise. A statistical analsis of the effects of measurement errors and errors in the positioning of the hologram samples is presented. The stability of the polynomial model with noise is discussed and it is shown that this stability increases when the matrix describing the model takes a triangular form. The performance of the triangular polynomial model is then considered for the case of a 1-point object with and without noise. The stability with noise is achieved at the expense of some loss in prediction accuracy and a triangular corrective model is employed. Imaging of multiple-point objects is then discussed using the linear model which assumes the triangular form due to the windowing effect at the construction stage. The effect of the computation accuracy and the round-off errors due to the limited word length of the computer is discussed for both the triangular and the square matrix models.

In order to verify the effectiveness of the proposed technique when used on realistic measured holograms, an experimental holographic imaging system has been designed and constructed; this is described in Chapter 6. This employs a microcomputer system based on a TMS 9900 16-bit microprocessor for controlling the various aspects of the imaging process which include aperture scanning, data acquisition, image reconstruction, hologram and image display, in addition to interfacing to a minicomputer which performs aperture expansion. The hardware and software for measuring the amplitude and phase of the complex hologram are described. Various aspects of the hardware and software design of the display system employed are discussed, with emphasis on the techniques adopted for economizing on the cost of the display memory by reducing the speed requirements and the method used for simplifying and speeding up the accessing of this memory by the microprocessor. The microprocessor software for image reconstruction which includes a two-dimensional FFT algorithm is described.

Chapter 7 deals with the experimental results obtained using a single point object. The problem of hologram tilt is addressed because of its influence on the measured holograms and reconstructed images, together with methods for compensating for the tilt effect. The experimental layout which allows for the correction of hologram tilt is then described. Images reconstructed from the full length of the measured line holograms are presented and the factors contributing to errors in these images are discussed. Results for aperture expansion are given where the measured aperture or part of it is enlarged by a factor of two. Methods for improving on the quality of the images obtained from predicted data are discussed, including the smoothing of the measured hologram data to reduce the effects of errors. At the end of this chapter, a number of simulation results on two-dimensional holograms are presented to demonstrate some aspects of resolution in holography.

In Chapter 8 the final conclusions are made and suggestions for future work regarding both the theoretical and experimental aspects of the research are presented.

1.3.3 Contributions of the Thesis

It is believed that this thesis makes the following contributions to the field of resolution improvement in acoustic holography and to acoustic holographic imaging in general:

1) The thesis highlights the link between resolution improvement by aperture expansion in holography when imaging finite objects in the far field and the techniques of object restoration in optics and the extrapolation of band-limited time signals in spectral analysis. A common factor among all these techniques is that they attempt to extend the knowledge about the spectrum of the signal beyond the limit set by practical considerations which restrict the window over which the data is collected either in the space domain or time domain respectively. The three classes of techniques are based on the fact that the spectrum of a finite object, or the function with a band-limited spectrum, is analytic and can therefore be continued beyond a given region over which it is known. In spite of this striking similarity, methods adopted in optics or spectral analysis are rarely mentioned in the literature in the context of resolution improvement in acoustic holography; while it is believed that it would be beneficial to treat the problem in holography as one facet of the more universal problem and to make use of the techniques developed in the other fields.

2) The Fresnel hologram of a finite object is shown to be an analytic function. Moreover, analogy with the problem of extrapolating time-limited functions in spectral analysis indicates that the hologram function is analytic in space regardless of the object range given that the hologram spectrum is band-limited. This serves to remove the restrictions on the imaging geometry for analytic continuation to be applied for aperture expansion.

3) A new technique for resolution improvement in holography is proposed which is based on aperture expansion through analytic continuation. With this technique, a model is fitted to the hologram data within the available aperture and is then used to

predict the data at new points outside this aperture. This method is believed to have the following advantages compared to other techniques reported in the literature for resolution improvement in optics and in acoustic holography:

a) The method is computationally simple and is not inherently iterative. The model is constructed by solving a set of linear equations and prediction is performed through a recurrence relationship. Other iterative techniques require long computation times which restrict the use of such techniques for resolution improvement when near real-time operation is required.

b) The method does not require the knowledge or estimation of the extent of the finite object to be imaged. The requirement for the object extent to be known or estimated demands additional a priori information which can restrict the application of other techniques. Moreover, the effectiveness of such techniques depends on the accuracy with which this parameter is estimated and the algorithms used may fail if the true object extent is underestimated.

c) The algorithm operates directly on the hologram data without requirement for modifying this data or changing the acquisition procedure. In some other techniques, such modifications require additional a priori information on the object and may introduce errors in the processed data.

d) The algorithm is valid for objects in the Fresnel zone and therefore reduces the restrictions on the imaging range and the extent of both the object and the hologram aperture. This makes the method more versatile and allows for a wider range of imaging geometries.

e) The method is robust against noise and offers some flexibility in achieving a compromise between prediction accuracy and noise sensitivity by proper choice of the model used. Experimental results have indicated the effectiveness of the method with realistic holographic data.

f) Simulation tests performed on discrete quasi-continuous objects suggest that the technique would be suitable for use with continuous objects.

4) The experimental holographic imaging system which has been designed and constructed serves to demonstrate the capabilities of small minicomputers and microprocessor systems in achieving the desirable integration between the various aspects of the holographic imaging process together with the attendant signal processing for resolution improvement.

5) A novel technique for digitally measuring the phase of the hologram signal is described. The digital phase measurement allows shorter acoustic pulses to be used compared to the conventional techniques of measuring the real and imaginary components as DC levels using balanced mixers. The use of shorter pulses should improve the range resolution of holographic imaging systems.

CHAPTER 2

A REVIEW OF ACOUSTIC HOLOGRAPHY

2.1 Introduction

Since its advent as an off-shoot of optical holography in the late sixties, acoustic holography has been the subject of a considerable amount of research, especially during the first decade of its development. Although it started by following the same lines adopted in optical holography, it was not long before the advantages offered by working in the acoustical domain made their impact. This manifested itself in simplifying the processes involved and in the great diversity of techniques offered for hologram acquisition and image reconstruction.

In this chapter, following a short historical background, the basic principles of holography are reviewed together with the techniques used for recording and reconstructing acoustic holograms. Here, the word 'hologram' is used in a broader sense than just the photograph of a diffraction pattern caused by the interference between an object wave and a mutually coherent reference wave, rather the record of the complex amplitude of the field across the wavefront at a given plane which is sufficient to reconstruct a replica of the wavefront at other planes. Emphasis will be on the fundamental techniques rather than the details of systems and equipment. Because of its relevance to the work described in this thesis, the method of digital image reconstruction will be dealt with in more detail.

Main design aspects of holography as an acoustic imaging system are then considered. The relationships governing the various system parameters such as resolution, range, and field of view are derived and the factors and trade-offs affecting the system design briefly discussed. This is followed by a comparison between holography and other imaging techniques, such as B-scan, lens, and beam-forming systems. Towards the end of the chapter the various areas of applications for acoustic holography in the fields of nondestructive evaluation, underwater imaging, medical diagnostics, and seismic imaging are discussed. The limitations associated with acoustic holography are then considered together with prospects for overcoming some of these limitations.

2.2 Historical Background [1],[2],[3]

Attempts to use sound for imaging date back in history to the early decades of this century [4]-[6]. However, the past two decades have witnessed a spurt of new research activity in this field, stemming primarily from the progress made in optical holography and the advent of laser.

Holography was initiated as an interferometric technique for coherently recording the exact complex wavefront of an optical wave and subsequently reconstructing it using coherent light. As a method of imaging, holography is based on the principle wavefront reconstruction investigated originally by Bragg [7]. This principle allows the wavefront of a propagating wave to be reconstructed at any plane with the knowledge of the exact complex field of the wave at one plane. Optical detectors, such as photographic plates are of the square-law type and therefore sense only the light intensity with the phase information being lost in the process as in ordinary photography. It was Gabor [8] who first recognized that both the amplitude and phase information of an object wave can be stored in the fringe pattern obtained when the object wave is present simultaneously with a suitable coherent reference on a recording medium, even when that medium responds only to intensity. However, it was not until the advent of laser in the early sixties that the modern revolution in holography began. Laser provided the required light sources with coherence lengths long enough to make it easy to record and reconstruct holograms. In 1962 Leith and Upatneiks [9] have established the link between the techniques used in holography and the concepts of communication and information theory, suggesting a modification of the Gabor's original technique, which greatly generalized and improved the holographic process. In 1964 great popular interest was created when the same authors [10] demonstrated the three dimensional imaging capabilities of holography. The success in optical holography prompted efforts to investigate the use of holography in other parts of electromagnetic spectrum and with other forms of radiation.

The early experiments with acoustic holography were direct analogues of optical holography methods which employed a separate reference beam and used square-law area detectors. The first ultrasonic hologram [3] was recorded by P. Greguss [11] in 1965 using a photographic plate whose exposure depended upon chemical changes introduced by the impenging sound. This was followed by the now well-known method of liquid-surface holography [12], also known as the static ripple method. In this technique, the acoustic hologram is recorded as deformations on the surface of a liquid-air interface. Other intensity-sensitive acoustic detectors reported in the literature for use in holography include liquid crystals [13], thermoplastic films [14], and suspended particles [15].

Linear detectors, which are readily available in acoustics, have greater sensitivity than intensity detectors. They also offer the advantage of replacing the acoustic reference with an electronically simulated reference during the recording of the hologram. This simplifies the recording process and adds to its flexibility. Additionally, the relatively large wavelengths in acoustics have made it possible to sample the hologram at a finite number of discrete points instead of continuously recording it in space as in the case of area detectors. Thurstone [16] was the first to describe such a system in an elementary form in 1966 using mechanical scanning. Since then, this category of sampled holograms has featured prominantly in the literature.

In search for more speedy methods of scanning the hologram aperture, the Sokolov tube arrangement which had been in use for direct imaging for a number of years was employed [17]. Another method for rapidly forming acoustic holograms using a laser scanning technique was described by Korpel and Desmares [18] in 1969. This technique is known as the dynamic ripple or solid-surface method.

The relatively small numerical apertures in acoustic holography have also made it possible to use piezoelectric arrays to sample the hologram [19]. Arrays made up of other types of transducers have also been reported in the literature, these include the electrostatic [20] and the condenser microphone [21] types.

A number of techniques have been used for simulating the acoustic reference electronically in sampled holographic systems. These include the addition [22] or multiplication [23] with an electronic wave of the same frequency as that of the object wave, followed by low pass filtering and DC biasing. Other methods employ frequency offsetting a signal derived from the electronic signal used to insonify the object [24].

Although the linear detectors used in the methods described above detect the complex amplitude of the object wave, the end product after mixing with the simulated reference is a hologram transparency for optical reconstruction. With fast computers becoming available and with the discovery of the Cooley and Tukey algorithm [25] for the Fast Fourier Transform in 1965, it has become possible to use computers to reconstruct acoustic holograms. Computers can be used to reconstruct a digitized form of a hologram transparancy by simulating the diffraction of coherent radiation by the hologram in direct correspondance with the process of optical reconstruction. Alternatively, the computer can operate directly on the data corresponding to the measured complex amplitude of the object wave over the hologram plane. In this way the process of recording the hologram as an interference pattern, whether actual or simulated, is bypassed altogether.

In 1969, Sondhi [26] reconstructed the image of an object from measurements of the amplitude and phase at a discrete set of sampling points in a plane at some distance from the object. In the same year, Boyer et al [27] reported a similar technique using the Fresnel approximation. In the following year the same authors [28] described a more powerful technique of wavefront reconstruction by backward wave propagation in the frequency domain, thus removing the restrictions on the object range and aperture size which limited the application of the Fresnel method.

Acoustic holography has found application in areas such as seismic imaging [29], medical diagnostics [30], underwater viewing [19], [31], and nondestructive evaluation [32]. As the digital processing becomes faster and cheaper, the use of dedicated microcomputers and signal processing devices should reduce the problems of speed and cost associated with the data processing for digital image reconstruction and wider applications of the principles of holography can be expected.

Acoustic holography today represents several different processes and techniques which have played an important role in the field of acoustic imaging. Equally significant has been the influence of the holographic approach on the entire field of acoustic imaging in fertilizing and refining existing imaging concepts and techniques.

2.3 Principles of Holography

It is well known in physics that coherent light, or any form of monochromatic radiation, which passes through an aperture is fully described by the amplitude and phase distribution within that aperture. According to the principle of wavefront reconstruction, this information is sufficient to reconstruct the wavefront at other planes along the path of propagation of the incident wave. If the incident wave carries information about an object, it will be possible to obtain an image of the object. To perform this method of imaging in optics was not possible before the introduction of the concept of holography as a two-step process by Gabor in 1948 and the discovery of laser as a source of coherent light later in the early sixties.

2.3.1 Imaging With Wavefront Reconstruction

Gabor [8] showed that phase information of an arbitrary coherent optical wavefront can be recoded on square-law detectors, which are the only type of detectors available in optics, by converting this information into intensity variations for recording purposes. This is achieved by allowing the wave to interfere with a mutually coherent reference wave of known amplitude and phase. The intensity of the resulting interference pattern is recorded on a photographic film which Gabor called 'Hologram',i.e. 'total recording'. At the end of this recording step the hologram contains both the amplitude and phase information of the original wavefront.

In the second step, the original wavefront is reconstructed by illuminating the hologram by a coherent wave. The fringe pattern recorded on the hologram acts as a diffraction grating and causes the incident wave to be diffracted, thus generating new waves which include the original wavefront. If the original wavefront is obtained by transmission through, or reflection off, an illuminated object, the reconstructed wavefront furnishes a three-dimensional image of the object. The key mechanisms involved in holography are therefore interference in the recording step, and diffraction in the reconstruction step.

Fig. 2.1a depicts the recording process. Let the object wavefront at the hologram plane be given by:

$$a(x,y) = \hat{a}(x,y) \exp[-j\varphi(x,y)] \tag{2.1}$$

and the reference wavefront be:

$$A(x,y) = \hat{A}(x,y) \exp[-j\psi(x,y)] \tag{2.2}$$

where $\hat{a}(x,y)$, $\hat{A}(x,y)$ are the amplitude distributions, and $\varphi(x,y)$, $\psi(x,y)$ are the phase distributions of the object and reference waves over the recording plane.

The reference wave interferes with the object wave at the recording plane, the intensity of the sum of the two waves is given by:

$$\begin{aligned} I(x,y) &= |A(x,y) + a(x,y)|^2 \\ &= |A(x,y)|^2 + |a(x,y)|^2 + 2\hat{A}(x,y)\hat{a}(x,y) \\ &\quad \cos[\psi(x,y)-\varphi(x,y)] \end{aligned} \tag{2.3}$$

The first two terms are the intensities of the two interferring waves, while the third term contains their amplitudes and relative phase. The expansion in eqn. (2.3) can also be written in the form:

$$I(x,y) = |A|^2 + |a(x,y)|^2 + A^*a(x,y) + Aa^*(x,y) \tag{2.4}$$

where '*' indicates the complex conjugate in the space co-ordinates. The photographic film at the hologram plane provides a linear mapping of the intensity incident during exposure into amplitude transmittance $T(x,y)$ after the film is developed which is proportional to $I(x,y)$ in eqn. (2.4).

To reconstruct the object wavefront $a(x,y)$ the developed transparancy is illuminated by a coherent reconstruction wave $B(x,y)$. The light transmitted through the transparancy is given by:

$$\begin{aligned} B(x,y)\, T(x,y) &= |A|^2 B + |a|^2 B + A^*Ba + ABa^* \\ &= u_1 + u_2 + u_3 + u_4 \end{aligned} \tag{2.5}$$

where the space co-ordinates x,y were omitted for simplicity.

If the reference wavefront is assumed to have a uniform intensity over the recording surface, then $|A|^2$ will have no spatial variations and therefore causes no diffraction of the constructing wave B. The first term u_1 is therefore an attenuated version of the incident wave B(x,y) , referred to in optics as the zero order term. The second term is a non-image-forming term since the phase portion of the object wave, a, is missing. It represents a noise-like background. The prospect for reconstructing the original wavefront, a, lies therefore in the remaining two terms u_3 and u_4.

If the illuminating wave B is chosen to be an exact duplication of the reference wave A used in the recording process, i.e. B=A, then the third term in eqn. (2.5) becomes $|A|^2$ a(x,y) which is, up to a simple multiplicative constant, an exact replica of the original object wavefront as shown in Fig. 2.1b. For an observer at the right hand side of the hologram, the transmitted wave component u_3 as defined by eqn. (2.5) would appear to be diverging from the original object. Thus the term u_3 can be regarded as generating a virtual image of the object .

Similarly, if B is a spatial conjugate of the reference wave, i.e. B=A*, then the fourth component of the transmitted field, u_4 is proportional to the conjugate of the original object wavefront, Fig. 2.1c. This wavefront converges in the space to the right hand side of the hologram to form a real image of the object corresponding to an actual focusing of light.

It should be noted that in both cases of wavefront reconstruction, the field component of interest, i.e. u_3 for B=A or u_4 for B=A* , is accompanied by three additional unwanted components. Obviously, such components have to be separated in space from the required wavefront in order to remove any extraneous interference with the required image.

2.3.2 Gabor's on-axis holography

In the arrangement for holographic imaging first proposed by Gabor, colinear object and reference waves were used as illustrated in Fig. 2.2a. A highly transmittive object is illuminated by a collimated coherent wave. Because of the high average transmittance of the object it passes a strong uniform plane wave which acts as a reference together with a weak scattered wave which is generated by the small spatial variations in the object transmittance and therefore carries object information. In this case, therefore, the object wave is much smaller in amplitude than the reference wave, i.e. $a(x,y) \ll A$, and the object supplies the required reference wave through its average high transmittance.

The developed transparancy is illuminated by a normally incident plane wave of uniform amplitude B, Fig. 2.2b. The resulting transmitted field amplitude is proportional to:

$$
\begin{aligned}
BI(x,y) &= B|A|^2 + B|a(x,y)|^2 + BAa(x,y) + BAa^*(x,y) \\
&= u_1 + u_2 + u_3 + u_4
\end{aligned}
\tag{2.6}
$$

where use has been made of the fact that $A=A^*$ since A is normal to the hologram. Because both A and B are uniform, the third term u_3 represents a field component which is proportional to the original scattered wave $a(x,y)$. This wave creates a virtual image of the original object located at distance z from the hologram (Fig. 2.2b). This is the same distance between the actual object and the photographic film during recording. Similarly, the fourth term leads to the formation of a real image of the object at the same distance on the opposite side of the hologram transparancy.

The basic limitation of this type of hologram is that both images occupy the same space, around the hologram axis, accompanied by the zero order term u_1 and the noise term u_2 in eqn. (2.6). To reduce the effect of the noise term and prevent it from dominating the image components, the magnitude of the object wave must be much smaller than the reference wave. In practice this restricts the use of Gabor's hologram in many applications. Another serious limitation is the difficulty in separating the twin images generated. When one image is brought to focus, it is always accompanied by a defocused contribution from the other image which adds spurious details in the image.

2.3.3 Leith-Upatneiks' Off-Axis Holography

The most common type of holograms is the off-axis or offset-reference type described by Leith and Upatneiks [9]. This hologram avoids the restrictions on the nature of object transmittance and the problems of inseparability of the twin images and other reconstruction products which severely limit the use of the original Gabor hologram. This is achieved by introducing a separate and distinct reference wave during the recording process rather than depending on the object to transmit the reference. The reference wave is introduced at an angle which can be chosen large enough to separate the reconstruction products in space.

Fig. 2.3a shows an arrangement for recording an off-axis hologram. A collimated coherent light beam is split into two portions, one illuminating the object and the other deflected by a prism to fall on the recording film at an angle θ with the film normal. The tilted plane wave comprising the reference interferes with the object wave and the resulting amplitude distribution across the film is given by:

$$u(x,y) = A\exp(-j2\pi\alpha x) + a(x,y) \qquad (2.7)$$

where A is the uniform amplitude of the incident reference wave and α is the spatial frequency of the reference wave,

$$\alpha = \frac{\sin\theta}{\lambda} \quad \text{cycles/mm} \qquad (2.8)$$

and a(x,y) is the object wave, eqn. (2.1).

The intensity distribution across the film is given by:

$$I(x,y) = A^2+|a(x,y)|^2+A\,a(x,y)\exp(j2\pi\alpha x)+Aa^*(x,y)\exp(-j2\pi\alpha x) \qquad \ldots (2.9)$$

Eqn. (2.4) describing the on-axis hologram is a special case of eqn. (2.9) when $\theta=0$ and therefore $\alpha=0$.

Combining the last two terms of eqn. (2.9) yields:

$$I(x,y) = A^2 + |a(x,y)|^2 + 2\,A\hat{a}(x,y)\cos[2\pi\alpha x-\varphi(x,y)] \qquad (2.10)$$

The last term of eqn. (2.10) reveals that the amplitude and phase of the object wave have been recorded, respectively, as amplitude and phase variations of a spatial carrier of frequency α which is produced by the reference wave.

The reconstruction process is illustrated in Fig. 2.3b where the developed transparancy of the recorded film is illuminated by a normally incident uniform plane wave of amplitude B. Dropping unimportant constants related to film development and using eqn. (2.9), the field transmitted by the hologram has four distinct components:

$$\begin{aligned} u_1 &= BA^2 \\ u_2 &= B\,|a(x,y)|^2 \\ u_3 &= BA\,a(x,y)\,\exp(j\,2\pi\alpha x) \\ u_4 &= BA\,a^*(x,y)\,\exp(-j2\pi\alpha x) \end{aligned} \qquad (2.11)$$

The field component u_1 represents the zero order undiffracted plane wave travelling along the hologram axis. The second component u_2 is a noise term which carries some spatial variations and therefore would generate wave components at various angles with the optical axis. Component u_3 is proportional with the object wave, therefore it generates a virtual image of the object at distance z from the hologram. The linear exponential term in u_3 indicates that the image is deflected at an angle θ off the hologram axis, θ being the angle of incidence of the reference wave during recording. Similarly component u_4 generates a real image at the opposite side of the hologram and at the same distance from it, but deflected at an angle $-\theta$ from the axis.

Fig. 2.3b shows that if the angle θ is large enough the twin images will be separated from each other and from the remaining reconstruction products. Frequency domain analysis [33] shows that complete separation is achieved when the spatial frequency of the reference across the recording plane, α, satisfies

$$\alpha \geqslant 3\beta \qquad (2.12)$$

where β is the width of the spatial frequency spectrum of the object in cycles/mm. Consider the critical value of the carrier frequency at which the reconstruction products can just be separated. From eqns. (2.8) and (2.12) the corresponding minimum angle of incidence of the reference wave is given by:

$$\theta_{min} = \sin^{-1} 3\beta\lambda \qquad (2.13)$$

In this case, the maximum spatial frequency on the hologram is:

$$f_{max} = 4\beta \qquad (2.14)$$

Although off-axis holography succeeds in separating the twin images, this is achieved only at the expense of increasing the spatial frequency content of the hologram which is a direct consequence of using a tilted reference wave with a carrier frequency α as compared to zero carrier frequency in the case of on-axis holography. In acoustics, the spatial resolution during the recording of the hologram is often restricted by technical difficulties or due to time and cost considerations. Since the spatial frequency of the reference increases with the increase in the angle θ, the minimum value should be used which is just enough for image separation.

2.3.4 Holographic Images

The images obtained when the hologram is reconstructed retain all the three dimensional characteristics of the original scene. One particular effect that is readily noticed in the case of the virtual image is the parallax effect, where the observer can look behind objects in the foreground by simply changing his viewing position. The observer in this case reconstructs somewhat different areas of the hologram corresponding to different wavefronts scattered from the object during recording, and therefore obtains different views.

In general, holographic images suffer from a number of drawbacks [33] unique to coherent and non diffused illumination. One of the problems associated with coherent imaging in general is the speckle effect which is responsible for the granular texture of coherent images. Due to the roughness of most objects relative to the wavelength of the illuminating radiation and the high coherence of the illuminating source, constructive and destructive interference takes place between waves scattered by various points on the object thus leading to the grainy appearance of the image. The size of the individual speckles can be shown [34] to be roughly the same size of a single resolution cell on the object. This indicates that when a particular object is near the resolution limit of the imaging system, the speckle effect not only affects the appearance of the image but can also affect the resolution and add spurious details in the image. Because the speckle size is inversely proportional to the numerical aperture of the hologram, speckle is a more serious problem in acoustic holography due to the limited aperture because of the large wavelength.

Another problem which characterizes images obtained through reflection holography is that of specular reflections. Smooth objects, relative to the radiation wavelength, cause reflections from mirror like parts to dominate the hologram signal. When the hologram is reconstructed this causes highlights in the image. If the dynamic range of the recording medium is not large enough then spherical objects may appear as points and cylindrical ones as lines. Again, because of the large wavelength in acoustics, objects considered rough in optics are smooth relative to the acoustic wavelength. This makes the specular reflections more serious in acoustics. Although diffuse illumination can help remedy this problem, it tends to increase the speckle effect in the image. Another coherence effect in holographic images is that of edge riging [33] which causes the edges of objects in a coherent image to be less sharply defined in comparison with the case of images obtained by incoherent illumination. Of particular importance to the use of holography in acoustics is the magnification parameter [33] which relates the object dimensions to the dimensions of the virtual image. When the reconstruction wavelength is different from the recording wavelength, as in the case of reconstructing an acoustic hologram optically, the manification in the depth dimension is greater than in the lateral dimensions. This causes distortion in the resulting three-dimensional image.

2.4 Hologram Acquisition Techniques

In optics there is virtually one basic technique for acquiring the holographic information; by recording the hologram as an interference pattern between the object wave and a reference wave when they simultaneously impenge on the recording photographic plate. Similarly, optical reconstruction of the hologram transparancy is the most viable and suitable method for recovering the optical image. In acoustics, however, a number of factors have contributed to the great diversity in the techniques used in hologram acquisition and image reconstruction. These factors include:

1) The lack of suitable area detectors of the square-law type which could replace the photographic plate in acoustics with equally high sensitivity and resolution in detecting distributions of acoustic intensity.

2) The availability of linear detectors which are sensitive to the complex amplitude of the acoustic signal.

3) The feasibility of sampling the acoustic hologram at a finite number of points because of the smaller numerical apertures and lower spatial frequencies compared to optics.

4) The problems associated with optical reconstruction of acoustic images due to the use of different wavelengths in the 2 steps of the holographic process.

5) The feasibility, and advantages, of using the computer to reconstruct the image from holographic complex data in acoustics.

This led to a wide range of methods which achieve the basic principle of acoustic imaging by wavefront reconstruction. These methods vary from direct analogues of the optical techniques, such as the liquid surface method, to techniques where the computer calculates the image by operating on a set of data without even the need for a hologram to be recorded as an interference pattern. A large group of methods employ a reference wave simulated electronically.

All hologram acquisition techniques, with the exception of those using square-law detectors employ some form of spatial sampling and use an electronic reference in one form or another. Before reviewing the various techniques of hologram acquisition, some aspects of hologram sampling are discussed together with the use of electronic reference waves in acoustic holography.

2.4.1 Some Aspects of Hologram Sampling

The subject of sampled and scanned holograms is discussed in a number of publications [35], [36]. Because of the large wavelength, acoustic holograms can be sampled at a relatively large spacing. For reasons of speed, economy, and convenience, the sampling frequency should be kept to the minimum required for the proper acquisition of the hologram information. Additionally, the finite size of the sampling detector also affects the sampling of the hologram.

. The Sampling Frequency

Let $g(x,y)$ be a band-limited function describing the spatial hologram distribution in the xy plane. This function can be the result of interference of the object wave with a simulated or actual reference. Assume that the hologram is sampled over an infinite aperture in two dimensions at an infinite number of equally spaced points on an orthogonal lattice using an ideal point detector. Let Δx and Δy be the sample spacings in the x and y directions respectively. Since $g(x,y)$ is assumed to be band limited, its specturm, $G(f_x,f_y)$ is non zero over only a finite region of the frequency space extending over $\pm f_{x\,max}$ in the f_x direction and $\pm f_{y\,max}$ in the f_y direction. The spectrum of the sampled hologram , $G_s(f_x,f_y)$ is the result of a convolution in the frequency domain between the spectrum $G(f_x,f_y)$ and that of the array of sampling Delta functions located at the sampling points, hence

$$G_s(f_x,f_y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} G(f_x - \frac{n}{\Delta x}, f_y - \frac{m}{\Delta y}) \qquad (2.15)$$

This spectrum consists of a multiplicity of the spectrum $G(f_x,f_y)$ each centred about each point $(n/\Delta x, m/\Delta y)$; $n,m=-\infty,\ldots,\infty$ in the f_x,f_y

plane. The centres of the hologram function spectra are therefore separated by $1/\Delta x$ in the f_x direction and $1/\Delta y$ in the f_y directions. When the hologram is optically reconstructed, each of the spectra is capable of producing an image corresponding to the hologram function. Therefore, the effect of hologram sampling is that multiple images are produced. The physical explanation of this is that the sampling lattice has the same effect as a two-dimensional grating superimposed on the continuous hologram transparancy. When the transparancy is illuminated by coherent light for reconstruction, the light usually diffracted to form the image in the case of a continuous hologram is further diffracted by the lattice grating to form additional multiple images arranged orthogonally to the grating lines.

These multiple images would overlap in space if their spectra overlap in the frequency domain. To prevent this, the spacing between the centres of the spectra in eqn. (2.15) must be greater than or equal to the total spectrum width in that direction. The upper bounds for the sampling intervals are therefore

$$\Delta x \leqslant \frac{1}{2 f_{x\ max}} \quad \text{and} \quad \Delta y \leqslant \frac{1}{2 f_{y\ max}} \tag{2.16}$$

when these two conditions are satisfied, no aliasing of the spectra takes place and it is theoretically possible to recover the original function $g(x,y)$ exactly by passing the sampled function through a linear filter that transmits only the centre spectrum ($n=0$, $m=0$) of eqn. (2.15) without distortion while perfectly excluding all other terms. Eqn. (2.16) indicates that the samples should be spaced at half the period of the finest fringe to be recorded or less.

To determine the maximum spatial frequency over the hologram, consider the simple one-dimensional case illustrated in Fig. 2.4. Any object can be considered as a series of points and the maximum frequency over the hologram aperture will be that corresponding to an extreme point on the object. In Fig. 2.4 two points represent the object and the reference at co-ordinates (x_1,z_1) and (x_2,z_2)

respectively. Assuming that the two points radiate in phase, the phase of the resultant signal incident on the hologram at point (x,0) is given by:

$$\varphi(x) = K\,(r_1' - r_2') \tag{2.17}$$

where r_1' and r_2' are the distances between the point on the hologram and the object and reference points respectively and K is the wavenumber; $K = 2\pi/\lambda$ where λ is the wavelength.

$$r_1' = [z_1{}^2 + (x-x_1)^2]^{\frac{1}{2}} \tag{2.18}$$

$$r_2' = [z_2{}^2 + (x-x_2)^2]^{\frac{1}{2}} \tag{2.19}$$

f_x, the spatial frequency in cycles/mm at point x is given by:

$$f_x = \frac{1}{2\pi}\,\frac{\partial\varphi(x)}{\partial x}$$

$$= \frac{1}{\lambda}\left[\frac{\partial r_1'}{\partial x} - \frac{\partial r_2'}{\partial x}\right] \tag{2.20}$$

Using eqns (2.18) and (2.19), f_x is given by:

$$f_x = \frac{1}{\lambda}\left[\frac{x-x_1}{r_1'} - \frac{x-x_2}{r_2'}\right] \tag{2.21}$$

Applying the Fresnel approximation,

$$r_1' \approx r_1 \qquad \text{and} \qquad r_2' \approx r_2 \tag{2.22}$$

where r_1 and r_2 are the distances between the centre of the hologram and the object and reference points respectively, yields:

$$f_x = \frac{1}{\lambda}\left(\frac{1}{r_1} - \frac{1}{r_2}\right)x + \frac{1}{\lambda}\left(\frac{x_2}{r_2} - \frac{x_1}{r_1}\right)$$

$$= \frac{1}{\lambda}\left(\frac{1}{r_1} - \frac{1}{r_2}\right)x + \frac{1}{\lambda}(\sin\theta_2 + \sin\theta_1) \tag{2.23}$$

where θ_1 and θ_2 are the angles between r_1 and r_2 respectively and the z axis. If plane waves are assumed for both the object and reference radiation, then $r_1 = \infty$ and $r_2 = \infty$ in eqn. (2.23) and f_x

becomes constant over the hologram aperture at a value given by:

$$f_x = \frac{\sin\theta_1 + \sin\theta_2}{\lambda} \tag{2.24}$$

Eqn. (2.24) shows that the spatial frequency of the hologram increases as the angle between the object and reference waves increases. This indicates that the on-axis type of holography is more tolerant to large sampling spacing than the off-axis type.

For a given inclination of the plane reference wave, the maximum spatial frequency will be determined by the wave components originating from points on the object furthest from the reference wave. If the furthest plane wave component makes an angle $\theta_{1\ max}$ with the hologram normal, then from eqns. (2.16) and (2.24) the maximum sample spacing allowable for proper sampling of the hologram is given by:

$$\Delta x_{max} = \frac{\lambda}{2(\sin\theta_{1\ max} + \sin\theta_2)} \tag{2.25}$$

For a given sample spacing Δx, the requirement for proper sampling of the hologram restricts the field of view to rays arriving within a maximum angle given by:

$$\theta_{1\ max} = \sin^{-1} \left[\frac{\lambda}{2\Delta x} - \sin\theta_2 \right] \tag{2.26}$$

A commonly used reference wave is that of a plane wave normally incident on the hologram plane. This also corresponds to the case when the complex object field is measured directly. In this case the phase is measured relative to the electronic signal used to generate the insonifying wave. This electronic reference therefore has a constant phase and hence simulates a plane wave normally incident on the hologram. For such a reference wave θ_2=0 and from eqns. (2.25) and (2.26), Δx_{max} and $\theta_{1\ max}$ are given by:

$$\Delta x_{max} = \frac{\lambda}{2 \sin\theta_{1\ max}} \quad (2.27)$$

$$\theta_{1\ max} = \sin^{-1} [\frac{\lambda}{2\Delta x}] \quad (2.28)$$

Eqns. (2.26) and (2.28) show that limitations on the sample spacing of the hologram restrict the field of view in a holographic imaging system.

. Effect of Detector Size on Hologram Sampling

In the above analysis it was assumed that the sampling is achieved using an ideal point detector. In practice, however, the detector will have a finite aperture over which the received signal is integrated and the resulting signal assigned to the sample value at the position of the aperture centre, i.e.

$$g'(p\Delta x,\ q\Delta y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) p(x-p\Delta x, y-q\Delta y)\ dxdy \quad (2.29)$$

where $g'(p\Delta x, q\Delta y)$ is the new value assigned for the sample at coordinates $(p\Delta x,\ q\Delta y)$ and p,q are integers defining the position of the centre of the detector aperture on the sampling lattice and $p(x-p\Delta x,\ y-q\Delta y)$ is the detector sensitivity function over its aperture, centred at the sample coordinates $(p\Delta x,\ q\Delta y)$.

Since the detector sensitivity can be assumed an even function, then from eqn. (2.29), the new function $g'(x,y)$ is related to $g(x,y)$ by:

$$g'(x,y) = g(x,y) * p(x,y) \quad (2.30)$$

where * denotes spatial covolution. The hologram signal at the sampling point is therefore given in the frequency domain by:

$$G'(f_x,f_y) = G(f_x,f_y) \cdot P(f_x,f_y) \quad (2.31)$$

where $P(f_x,f_y)$ is the spatial frequency spectrum of the detector sensitivity function $p(x,y)$.

Eqn. (2.31) shows that the hologram spectrum is shaped by the spectrum of the detector function before any effects of periodic sampling are introduced. If the detector aperture is wide enough, its frequency spectrum becomes narrow and some high frequency components of the hologram spectrum may be attenuated or cut off, thus leading to loss of information. This loss of spatial information is equivalent to a reduction in the size of the hologram aperture and therefore causes deterioration in resolution and loss of fine details in the image. Moreover, limitations on the maximum spatial frequency that can be sampled on the hologram reduce the field of view. On the other hand, the sensitivity of the detecting transducer increases with the increase in its aperture area. This makes the choice of the detector aperture size a trade-off between sensitivity on one hand and resoltuion and the field of view on the other.

Consider the case of a circular detector of diameter a with uniform sensitivity over its aperture. It can be shown [37] that the influence of the detector function becomes negligible if:

$$a \leqslant 1.22\, [\lambda_f]_{min} \qquad (2.32)$$

where $[\lambda_f]_{min}$ is the minimum fringe width on the hologram, i.e.

$$[\lambda_f]_{min} = 1/\,(f_{x\,max}^{2} + f_{y\,max}^{2})^{\frac{1}{2}} \qquad (2.33)$$

where $f_{x\,max}$ and $f_{y\,max}$ are the maximum spatial frequencies in the x and y directions respectively.

2.4.2 Electronic Reference in Acoustic Holography

Linear acoustic detectors, such as hydrophones and microphones, which are small enough to measure the local field amplitudes can be used to map out complex amplitude distributions across the hologram plane. The results can be used directly for computer reconstruction of the image, or an electronic reference signal can be added to synthesize a hologram for optical reconstruction. The electronic synthesis of the reference wave has simplified the hologram recording, increased the flexibility of the process, and led to a wide range of holographic methods and approaches. This section deals with the use of the electronic reference in the measurement of the complex field hologram and in the simulation of a plane wave acoustic reference.

. Measurement of the Complex Field Hologram

Sampling the complex acoustic field for reconstruction by the computer represents the most efficient sampling method for hologram acquisition. The phase of the received signal from the object wave is measured relative to the signal which generates the acoustic wave insonifying the object using a phase detector (Fig. 2.5a). The resulting phase together with the amplitude of the received signal at every point in the sampled hologram constitute the data inputs to the computer for image reconstruction [27],[28],[38].

In another method reported in the literature [39] - [41] the real and imaginary components constitute the data inputs to the computer. Fig. 2.5b illustrates a hardware implementation of this technique. The received signal enters two identical channels each consisting of a mixer, an integrator, and a digital to analog converter. The mixer in one of the channels multiplies the received signal by a component which is in phase with the insonifying signal. After the harmonic components in the product are filtered out, this channel produces a DC output which is proportional to the real part of the spatial complex field. Similarly, the imaginary part of the field is obtained through multiplication by a signal in phase quadrature with the insonifying signal in the other channel.

. Simulation of a Plane Wave Acoustic Reference

The above two methods for holographic data acquisition implicitly use a reference whose phase is constant over the hologram aperture. This is equivalent to a wave normally incident to the hologram. Consider the simulation of an electronic reference at a general inclination angle with the hologram plane. To simplify the analysis without affecting its generality the object is assumed to be illuminated by a plane wave normally incident to the hologram plane while the reference to be simulated is a plane wave incident at angle θ with the hologram normal. This recording configuration is similar to that shown in Fig. 2.3a, for which the hologram intensity is given in eqn. (2.10). From eqn. (2.8), the angle θ of the reference wave is related to the spatial carrier frequency α by

$$\sin\theta = \alpha\lambda \tag{2.34}$$

Consider a mechanically scanned hologram, Fig. 2.6, of a similar configuration to that of Fig. 2.3a with the exception that the physical reference wave is to be electronically simulated. The effective reference wave is shown dashed in Fig. 2.6. The object signal is linearly detected at a number of points spaced at distance Δx and is given by:

$$a(x,y) = \hat{a}\,(x,y)\,\sin[\omega t-\varphi(x,y)] \tag{2.35}$$

To simulate an inclined plane wave reference, a signal derived from the insonifying signal has its phase stepped by an increment $\Delta\varphi$ between the samples. The spatial rate at which the phase of the simulated signal is varied along the hologram scan is:

$$\beta = \frac{\Delta\varphi}{\Delta x} \tag{2.36}$$

and the simulated signal at point x on the scan has the form:

$$R = A\sin[\omega t-\beta x] \tag{2.37}$$

At every sample point along the scan the detected and the simulated signals are multiplied in a balanced mixer and the harmonic components filtered out by a low pass filter. The resulting DC term is given by:

$$I'(x,y) = \frac{1}{2}A\,\hat{a}(x,y)\cos[\beta x-\varphi(x,y)] \tag{2.38}$$

Comparing eqns. (2.38) and (2.10), and recalling that the third term in eqn. (2.10) is the only term responsible for the reconstruction of the twin images, it is obvious that the hologram signal obtained with the simulated reference in Fig. 2.6 is similar to that with a physical reference in Fig. 2.3a. From the comparison of eqns. (2.38) and (2.10) and taking eqn. (2.36) into account, eqn. (2.34) shows that the two reference waves will be equivalent when:

$$\sin\theta = \frac{\beta}{2\pi}\lambda$$

$$= \frac{\Delta\varphi}{2\pi}\frac{\lambda}{\Delta x} \tag{2.39}$$

For an on-axis reference $\theta=0$ and $\Delta\varphi=0$, while with $\Delta x=\lambda$ an off-axis reference at $\theta=30^{\circ}$ is simulated when $\Delta\varphi=\pi$.

This method for the electronic simulation of the reference wave has the advantage that the noise generating term $|a(x,y)|^2$, is absent from the hologram signal. This improves the quality of the resulting images and reduces the requirement on the sampling of the hologram, cf.sec. 2.3.3.

The simulation of the reference wave by an electrical signal whose phase is continuously varied in space is used when the hologram is mechanically scanned [42] or when an array of fixed transducers is used to sample the hologram [43]. Fast scanning methods such as electron beam scanning of the ultrasonic camera [24],[44] and the laser beam scanning [18] provide a more convenient means of introducing the electronic reference using a frequency offset. If the frequency of the insonifying signal is offset by $\Delta\omega$ before it is mixed with the object wave then the angle of the simulated reference is given by:

$$\sin\theta = \frac{\Delta\omega}{\nu_x}\frac{\lambda}{2\pi} \tag{2.40}$$

where ν_x is the scanning speed

2.4.3 Review of Basic Methods of Hologram Acquisition

In this section the basic methods for recording the hologram or acquiring the holographic data are briefly described. These include the liquid-surface method, the dynamic ripple method, the ultrasonic camera technique, and the sampling of the hologram using arrays or mechanical scanning.

. Liquid-Surface (Static Ripple) Holography

This was one of the earliest techniques for recording and reconstructing acoustic holograms in real time [12], [45]. Fig. 2.7 shows a simplified diagram for the arrangement used. Two immersed transducers generate the acoustic wave insonifying the object together with a mutually coherent reference wave. The object and reference waves interfere at the liquid-air interface causing deformations in the liquid surface which represent the resulting acoustic hologram.

The surface of the liquid can be photographed to form a transparancy for optical reconstruction. Real time reconstruction can also be achieved as shown in Fig. 2.7. A laser beam illuminates the surface ripples and, upon reflection from the surface, it is modulated by the surface configuration. The image resulting from the diffraction of the illuminating laser by the hologram fringes is detected by a TV camera and displayed.

One of the basic limitations of this method, apart from the much lower sensitivity compared to piezoelectric detectors, is its spatial frequency limitations. Detailed analysis [46] shows that the liquid surface acts as a low pass filter in the spatial frequency domain and therefore suppresses the higher frequency components in the interference pattern. This limits the angle of the reference beam that can be used and makes this method best suited for on-axis holography. By pulsing the acoustic waves it is possible to increase the spatial cut-off frequency of the transfer function of the liquid surface and to improve on the image quality in general [1].

. Solid-Surface (Dynamic Ripple), Laser Beam Scanning Technique

In this technique [18] the object wave impenges on a solid surface coated by a light reflecting layer. A laser beam scans the surface and picks up the local acoustic excitations as linear phase variations. These are transformed into amplitude variations in an electrical signal using a knife edge placed in front of a photodetector. The electrical signal is then mixed with an appropriate reference signal to produce a hologram in real time which is displayed and can be photographed for optical reconstruction. This method of hologram detection has better resolution and sensitivity compared to the liquid surface method and is therefore suitable for high frequency imaging.

. Hologram Recording with the Ultrasonic Camera

The ultrasonic camera (Solokov tube) was adapted by Fritzler et al [24] for use in recording holograms with both physical and electronic references. A schematic diagram for the apparatus used in the latter case is shown in Fig. 2.8. A quartz crystal located at the front window of a cathode ray tube is irradiated by the object wave. This induces voltages in the piezoelectric crystal whose magnitude at any point on its surface varies linearly with the instantaneous amplitude of the acoustic signal at that point. This voltage modulates the intensity of the secondary emission generated by the scanning electron beam. The resulting signal is amplified in the electron multiplier and in an RF amplifier before it is mixed with an electronic reference of the frequency offset type whose phase is locked onto that of the signal generator during the blanking period of each scanning frame in the vertical direction. The frequency offset $\Delta\omega$ controls the angle of the simulated plane reference wave with the horizontal as given by eqn. (2.40). The resulting hologram is displayed and may be photographed for optical reconstruction.

Although this method has the advantages of fast scanning and simplicity of operation, it suffers from a number of basic limitations which restrict its use in holography. The ultrasonic camera suffers from a limited field of view of approximately $\pm 10^{0}$ from the tube axis for operation in water [21] due to the impedence mismatch between water and the piezoelectric material. This creates severe limitations on the imaging geometry and the use of off-axis acoustic reference waves. Further limitations on the field of view at low operating frequencies arise from the large thickness of the piezoelectric plate which has to be used in order to maintain operation at resonance for optimum sensitivity since this limits the spatial sampling resolution [37]. At higher frequencies, the reduced thickness of the plate often restricts the size of the plate for considerations of mechanical strength and therefore limits the aperture area available.

. Hologram Sampling with Arrays

Piezoelectric arrays as linear area receivers have a number of advantages over the piezoelectric plate discussed in the previous section in connection with the ultrasonic camera. In the case of arrays, the mechanical strength is no longer a limiting factor, and therefore large apertures can be covered. The sampling spatial resolution of the array is no longer limited by the element thickness, but will be determined by the size and the spacing between the array elements. Moreover, because each element of the receiving array can be processed individually, its signal can be integrated over the full time of the signal duration, therefore increasing the sensitivity. Such arrays have been used in holography both in linear [39] and two-dimensional [41] configurations.

The cost of building two-dimensional piezoelectric arrays rises rapidly with the increase in the array size. To achieve the most efficient use of the number of array elements, most holographic array systems reported in the literature are used to detect the complex field hologram rather than the interference pattern with a real or simulated reference. This is achieved by measuring either the amplitude and phase or the real and imaginary parts as described in section 2.4.2 To further economize on the array size, synthetic aperture techniques [47], [48] have been used so that a combination of a sparse transmitting array and a small filled receiving array can sample a large aperture area. The majority of holographic array systems in existance are used for underwater imaging. Because of the large distances involved and therefore the large apertures required, arrays provide the most practical way for hologram sampling, especially in the hostile underwater environment.

Electrostatic receiving arrays have also been demonstrated as linear array detectors for hologram sampling [20],[21],[49]. The advantages of these arrays in their potential simplicity, especially their amenability to printing techniques. An array of N^2 elements can be fabricated using only 2N electrodes by printing a set of N line electrodes as rows on one side of a plastic film and another set as columns on the other side. Although this technique allows arrays of large size to be built at a relatively low cost, electrostatic arrays in general suffer from lower sensitivity and greater inter-element cross-talk compared to piezoelectric arrays.

. Mechanical Scanning

Mechanical scanning provides a convenient and economic way of sampling the hologram information. An appropriate detector such as a small hydrophone or a focused transducer scans the hologram plane, usually in a raster format, to detect the interference pattern between the object wave and an acoustic reference wave. More commonly, the reference wave is simulated electronically by mixing the received object signal with a suitable reference derived from the insonifying signal. The resulting interference pattern is photographically recorded and used for optical reconstruction. The hologram acquisition process is simplified when only the complex field of the object wave is measured. (see section 2.5.2).

Mechanical scanning of the hologram has the advantage of covering very large apertures, with the sampling resolution set by the detector size for most applications. This makes it capable of producing good quality holograms and images. This method also has the advantage of relatively simple and inexpensive apparatus, especially when the scanning control is performed by a microcomputer which later computes the image. A serious drawback of this technique is the long time required for scanning the hologram and the requirement for moving parts in the appratus. Where these two factors are not a serious objection, mechanical scanning technique has its place, as in applications such as on-shore nondestructive testing [50].

2.5 Image Reconstruction Techniques

In acoustic holography, the techniques used for image reconstruction are classified into two main categories: optical reconstruction and computer reconstruction. A number of techniques have also been reported in the literature which employ acoustic waves for hologram reconstruction. In one technique [51], the information from a recorded intensity hologram is converted to binary (two-state) form and etched into a metal plate. When this plate is insonified, it diffracts the acoustic waves and the resulting real image is detected and displayed. In another technique [52], the data corresponding to the intensity hologram is used to drive the elements of a two-dimensional array in phase. The real image formed at a distance from the array is detected using another receiver array.

2.5.1 Optical Reconstruction

The optical method is the natural choice for reconstructing acoustic holograms recorded on a photographic film and developed as an optical transparancy [53]. This technique is in use in a number of applications which include medical diagnostics [54], nondestructive testing [55], and seismic imaging [56]. The photographic recording is usually achieved by photographing the acoustic interference pattern generated on a liquid surface interface or displayed on the face of a CRT tube or the screen of a TV monitor. The acoustic hologram is photographically reduced to suit the optical system and the developed transparancy is used for reconstruction. The process of optically reconstructing an off-axis hologram has been described in section 2.3.3 and is shown schematically in Fig. 2.3b.

The optical method of image reconstruction has a number of advantages. The imaging is performed almost instantly once the hologram transparancy has been prepared. The ability of the optical technique for handling very large amounts of data at the speed of light is remarkable. In optical holography this method has the great advantage of presenting three-dimensional virtual images which have all the effects associated with seeing the object in real life. Unfortunately, when using optical waves to reconstruct a hologram recorded with a radiation whose wavelength differs considerably from the optical wavelength, which is the case in acoustic and microwave holography, the resulting image suffers from severe distortion which renders three-dimensional imaging in this case almost unpractical. This is caused by the large disparity between the magnification factors in both the depth and lateral directions [57].

The loss of the ability to view acoustic images in three dimensions using optical reconstruction is not a problem when imaging planar objects. Moreover, it is possible to view the image of three-dimensional objects at any desired plane by focusing the reconstruction optics at such a plane, and this is considered adequate for most practical applications. It should also be noted that even without the depth distortion problem, the three-dimensional acoustic images would not have the same realistic quality of the optical images [58]. In optical holography, one of the most intriguing features of three-dimensional images is the parallax effect which enables the observer to see 'hidden' parts of the object by moving his head sideways. This is achieved by utilizing different portions of the hologram aperture area. In acoustics however, the wavelength is large and the numerical aperture is too small to allow for such an effect.

In addition to the depth distortion, the wavelength mismatch together with the need to reduce the acoustic hologram in size cause spherical aberration in the reconstructed waves [59]. Image degradation is also caused by imperfections in the photographic and optical apparatus used for hologram recording and reconstruction [60].

Conventional optical reconstruction in which a photographic transparancy has to be prepared from the acoustic interference pattern has the main drawback of the long time delays associated with photographic processing which makes the technique unsuitable for real time imaging. In many applications, such as medical diagnostics and underwater imaging, it is desirable to have real time imaging capability from the linear-detection,fast scanning techniques for hologram recording which have the advantages of high resolution and high sensitivity compared with the liquid surface method. A number of systems [61], [62] have been proposed to obtain real time optical reconstruction from a scanned hologram using electro-optical area modulators.

2.5.2 Computer Reconstruction

When optical radiation is used for recording a hologram, it is convenient to do the reconstruction optically; therefore using a single photographic transparancy for the three tasks of detection, storage, and spatial modulation of light during reconstruction. However, when the original illumination is non-optical (e.g. acoustic or microwave radiation) this convenience is no longer available since detection is not usually satisfactory with photographic films. Of course the choice of optical methods for reconstructing optically generated holograms was not just a matter of convenience. The enormous amounts of data in optical holograms has meant that optical processing is almost the only practical way of performing the linear transforms required for image reconstruction efficiently and speedily. There is also the unsurpassed capability of optical reconstruction for presenting realistic three-dimensional images in a way which no other method of display can achieve. However, when considering long wavelength holography where the wavelength is many times the optical wavelength, the choice of optical reconstruction needs re-examination.

As indicated in section 2.5.1, because of the large wavelength mismatch in the recording and reconstruction stages, the possibility of obtaining good three-dimensional images optically from acoustic holograms is almost ruled out. On the other hand, because of the large wavelength, numerical apertures encountered in acoustic holography are much smaller for a given physical aperture than in optical holography and the required sampling frequencies are equally smaller. This means that the amount of data in an acoustic hologram is much smaller than in an optical hologram of the same aperture size, and is therefore well within the capability of present day mini or microcomputers and inexpensive storage devices. The discovery of the Fast Fourier Transform (FFT) algorithm [25] in 1965 for performing the discrete Fourier transform and the formulation of the reconstruction process in a manner amenable to the use of such algorithm [27],[28] has meant that the data processing required for image reconstruction can be performed, on these fairly modest amounts of data, with reasonable speed and efficiency.

In fact, computer reconstruction is not only feasible, but also has a number of advantages over optical reconstruction. The technique is much simpler because no optical equipment, including laser, are required. All the factors relating to the reconstruction optics and contributing to image degradation will be avoided. The method also offers great flexibility in the processing of data. This allows for compensating for image degration, making quantitative analysis on reconstructed images, and performing various filtering and image processing and enhancement techniques [63]-[70]. With digital reconstruction it would be possible to superimpose, linearly, images corresponding to different angles of illumination in order to reduce the effect of specular reflections on the image. The ability to manipulate hologram and image data plays an important role in synthetic aperture techniques [47], [48] which aim at increasing the effective aperture of the acoustic hologram.

There are basically two main techniques for image reconstruction using the computer. In the first technique the computer simulates the illumination of an intensity hologram recorded as an interference pattern and calculates the scattered wavefronts. The result is a typical holographic reconstruction which includes conjugate images, a zero order component, and a noise component if a noise terms exists in the hologram record. This method has been used to reconstruct both optical [60],[71]and acoustic [72],[73] holograms.

A more practical method which is more advantageous regarding hologram acquisition, image reconstruction, and image quality, is to sample the hologram information as the complex data of the object field without the need for recording the hologram as an interference pattern between the object field and a reference wave. The computer then operates on this complex data to calculate the wavefield at the object plane which represents the required image. To appreciate the relevance of this method one has to recall that holography is essentially a technique for imaging by wavefront reconstruction. The wavefront is uniquely defined by its complex field at the measurement plane. The hologram in the sense of an interference pattern was found to be a must in optics. This is not only because there is no other way of recording the complex field but also because it is the only means of reconstructing the image optically. However the introduction of a spatial carrier in the form of an on-axis reference wave introduces extraneous terms to the required image in the form of the conjugate image, the zero order,and noise terms. To separate the required image from the unwanted components in space, the concept of off-axis holography has been introduced. This, however, could be achieved only at the expense of increasing the spatial frequency content of the hologram.

In acoustics, linear detectors are readily available and the need for optical reconstruction is not as compelling as in the case for optical holography. The bypassing of the process of recording the acoustic hologram has a number of advantages. The limitations arising during the recording of the optical intensity hologram which cause image degradation would be avoided. The absence of a spatial carrier means that there will be no interference with the computed image from other extraneous terms.

Another advantage of particular importance in acoustics is the increased efficiency in sampling the hologram information. From eqn. (2.14),when the spatial carrier has the critical value given in eqn. (2.13), the highest frequency component in the hologram is 4β cycles/mm where β is the spatial bandwidth of the object wave. However, when the object wave is sampled directly without using a reference the highest frequency is only β cycles/mm. This shows that in this case a four fold increase in the sampling efficiency is achieved with the direct measurement of the complex object field. For example, if N points at spacing Δx are required to sample an interference hologram over a given area, they can be reduced to N/4 spaced at $4\Delta x$ to sample the complex field over the same area. If the N points are spread out with the new larger spacing they will cover four times the size of the original aperture, hence increasing the resolution by the same ratio. If a sampling detector of diameter a is used to sample the interference hologram,it can be replaced with a bigger, and hence more sensitive, detector with diameter 4a when sampling the complex field alone; without increasing the band-limiting of the sampled spectrum due to the spectrum of the detector aperture function, cf. sec. 2.4.1.

Such a saving in the sampling efficiency is not very important in optics because of the availability of area detectors with high resolution capability in the form of the photographic plate, together with efficient 'area' data processors in the form of optical image reconstruction. In acoustics, however, the area detectors available, such as the liquid surface, have resolution and sensitivity limitations and the hologram is often sampled at a discrete number of points. The sampling process is either time consuming (as in mechanical scanning) or costly (as in array sampling). Therefore an improvement in the sampling requirement is much more welcome in acoustics not only because of the improvement in the image quality and the advantages in the data acquisition phase but also in the reconstruction phase using the computer. This is because the reduction in the number of sampling points required reduces the computer storage requirements and speeds up the reconstruction process.

As an example for the advantages gained consider a two-dimensional array used to sample the hologram. A reduction in the number of elements in each direction by a factor of 4 reduces the array cost by a factor of 16. An increase in the sample spacing by the same factor of 4 makes the array easier to manufacture and reduces the inter-element cross-talk and therefore improves the overall system performance. Increasing the element size increases the sensitivity and also simplifies the manufacturing process especially at high frequencies. The storage requirement for the input data during the digital reconstruction is reduced by a factor of 16 and the speed of the reconstruction process is also increased.

The speed and the capacity for handling data in the case of computer reconstruction are undoubtedly limited compared with the optical method. However, this is not a serious limitation in the case of the relatively small amounts of data encountered in acoustic holography. Moreover, this problem is being alleviated by the increase in the speed and storage capacity of modern computers and the decrease in their cost together with the availability of dedicated signal processing devices. Other limitations arise from the effects of quantization of the hologram data into a number of discrete levels for digital processing by the computer. The chief effect of the quantization errors is a reduction in image contrast, although it is possible for false image details to be introduced [60].

One of the other drawbacks unique to computer reconstruction is that the focusing distance at which the image is to be computed must be known beforehand for use in the reconstruction algorithm. In optical reconstruction this problem does not arise when observing the virtual image of an object because the focus is found automatically by an eye-brain interaction. This is not a very serious limitation in acoustics as the object distance can be measured by some other means such as pulse echo technique. Holographic imaging can then be used to obtain an image of the object at such range or nearby ranges. Moreover, a number of techniques have been suggested to overcome this problem. One method [74] employs an automatic object recognition scheme in which the computer calculates a series of images at a number of planes in the depth region of interest and decides the most likely position of the object using an edge detection algorithm. More recently, Sepehr et al [75] have described an automatic focusing technique which does not require the knowledge of a value for the range as a data input. This technique, however, is applicable only for planar objects and is limited to images with low spatial frequency content.

There are mainly two methods for computer reconstruction of complex field holograms,namely the Fresnel integral algorithm and the backward wave propagation (frequency domain) algorithm. The two techniques are described below.

. The Fresnel Integral Reconstruction Algorithm [27], [76]

Referring to Fig. 2.9,the basic problem in imaging with wavefront reconstruction can be stated as follows : given a complex field distribution u(x,y) of the object wave arriving at the hologram xy plane determine the corresponding field distribution $u_o(x_o,y_o)$ scattered or generated at the object plane x_o,y_o which is parallel to the hologram plane at distance z from it. The distribution $u_o(x_o,y_o)$ is taken as a representation of the object distribution and therefore as the required image.

According to the Huygens-Fresnel principle, each point on the wavefront at the x_o,y_o plane acts as a secondary point source radiating a spherical wave. The field amplitude at any point (x,y) on a plane at distance z is obtained as the resultant of contributions from all such point sources and can be readily written in the form of the superposition integral [33] as:

$$u(x,y) = \iint_{-\infty}^{\infty} h(x,y,z\ ;\ x_o,y_o,o)\, u_o(x_o,y_o)\, dx_o\, dy_o \qquad (2.41)$$

where $h(x,y,z\ ;\ x_o,y_o,\ o)$ is the impulse response of the propagation process which denotes the contribution at point (x,y,z) on the hologram plane due to a Delta function input, i.e. a point source, at point (x_o,y_o,o) on the object plane. This is given as [33]:

$$h(x,y,z\ ;x_o,y_o,o) = \frac{1}{j\lambda}\,\frac{\exp jKr}{r}\,\cos(\bar{n},\bar{r}) \qquad (2.42)$$

where K is the wavenumber, $K = 2\pi/\lambda$, λ is the wavelength , r is the distance between points (x_o,y_o,o) and (x,y,z), $\bar{r}$ is the vector pointing from (x,y,z) to (x_o,y_o,o) and $\bar{n}$ is the vector normal to the object plane in the negative z direction. (see Fig. 2.9).

In most cases the distance z between the object and hologram planes is much greater than the maximum linear dimensions of interest in both planes. This allows the following approximations to be made:

(i) $\cos(\bar{n},\bar{r}) \approx 1$

(ii) $r \approx z$
in the denominator of eqn. (2.42)

(iii) $r = [z^2+(x-x_o)^2+(y-y_o)^2]^{\frac{1}{2}}$

$$\approx z[1+\frac{1}{2}(\frac{x-x_o}{z})^2 + \frac{1}{2}(\frac{y-y_o}{z})^2]$$

in the exponent of eqn. (2.42). (2.43)

The above assumption concerning r in the exponent is known as the Fresnel approximation and is valid as long as the distance z is large enough so that higher order terms in the binomial expansion of r leading to this approximation do not significantly change the value of the superposition integral in eqn. (2.41). To give an indication of the restrictions on z for the validity of the approximation, the phase change contributed by the next higher order term only will be much less than 1 radian if z satisfies:

$$z^3 >> \frac{\pi}{4\lambda}[(x-x_o)^2 + (y-y_o)^2]^2_{max} \qquad (2.44)$$

Applying the approximations in (2.43) and substituting the impulse function from eqn. (2.42) into eqn. (2.41) yields:

$$u(x,y) = \frac{\exp jKz}{j\lambda z}\iint_{-\infty}^{\infty} u_o(x_o,y_o)\exp\{\frac{jK}{2z}[(x-x_o)^2+(y-y_o)^2]\}\,dx_o dy_o \qquad ...(2.45)$$

Expanding the quadratic terms in the exponent and rearranging gives:

$$u(x,y) = \frac{\exp jKz}{j\lambda z}\exp[j\frac{K}{2z}(x^2+y^2)]\iint_{-\infty}^{\infty}\{u_o(x_o,y_o)\exp[j\frac{K}{2z}(x_o^2+y_o^2)]\}$$

$$\exp\{-j\,2\pi[(\frac{x}{\lambda z})x_o + (\frac{y}{\lambda z})y_o]\}\,dx_o\,dy_o \qquad (2.46)$$

If f_x and f_y are defined as:

$$f_x \triangleq \frac{x}{\lambda z} \quad , \quad f_y \triangleq \frac{y}{\lambda z} \tag{2.47}$$

then the expression within the integration signs takes the form of a Fourier transform evaluated at the frequencies f_x and f_y,

$$u(x,y) = \frac{\exp jKz}{j\lambda z} \exp[\frac{jK}{2z}(x^2+y^2)]F[u_o(x_o,y_o)\exp\frac{jK}{2z}(x_o^2+y_o^2)]$$

...(2.48)

where F indicates the Fourier transform evaluated at the spatial frequencies f_x and f_y defined in eqn. (2.47). This can be written in the form:

$$\frac{\exp(-jKz)}{(j\lambda z)^{-1}} u(x,y) \exp[\frac{-jK}{2z}(x^2+y^2)] = F[u_o(x_o,y_o)\exp\frac{jK}{2z}(x_o^2+y_o^2)]$$

...(2.49)

In imaging with wavefront reconstruction, the field distribution $u(x,y)$ is known and it is required to determine the image distribution $u(x_o,y_o)$. Taking the inverse Fourier transform (F^{-1}) of both sides in eqn. (2.49) and solving for $u(x_o,y_o)$ gives:

$$u_o(x_o,y_o) = \frac{\exp(-jKz)}{(j\lambda z)^{-1}} \exp\frac{-jK}{2z}(x_o^2+y_o^2)\, F^{-1}\{u(x,y)$$

$$\exp[\frac{-jK}{2z}(x^2+y^2)]\} \tag{2.50}$$

The term $\exp(-jKz)$ is a constant phase shift, $(j\lambda z)^{-1}$ is a constant and the phase shift $\exp[\frac{-jK}{2z}(x_o^2+y_o^2)]$ is a function of the coordinates in the image plane. Therefore all these terms will not cause any spatial variations in the image intensity and hence can be ignored, thus:

$$u_o(x_o,y_o) = F^{-1}\{u(x,y)\exp[\frac{-jK}{2z}(x^2+y^2)]\} \qquad (2.51)$$

This equation forms the basis for the Fresnel integral method of numerical reconstruction from the complex field distribution u(x,y) in the hologram plane. The reconstruction algorithm consists of the following steps:

1) Multiplying u(x,y) by a quadratic focusing factor $\exp[\frac{-jK}{2z}(x^2+y^2)]$ which is a function of the distance z at which the image is to be reconstructed.

2) Taking the inverse Fourier transform of the result to obtain the image complex distribution $u_o(x_o,y_o)$. The image intensity is given by $|u_o(x_o,y_o)|^2$.

Referring to Fig. 2.10, the hologram plane xy is sampled at a finite number of NxN points spaced at Δx in the x direction and Δy in the y direction. In view of eqn. (2.48), the hologram space is equivalent to a spatial frequency space where the frequency coordinates are related to the space coordinates by eqn. (2.47). The sample spacings in that equivalent frequency space are therefore given by:

$$\Delta f_x = \frac{\Delta x}{\lambda z} \qquad \Delta f_y = \frac{\Delta y}{\lambda z} \qquad (2.52)$$

in the f_x and f_y directions respectively.

Let p,q be the point indices in the x and y directions respectively in the hologram space and in the f_x and f_y directions respectively in the equivalent spatial frequency space. Multiplication by the focusing function in the equivalent frequency space is obtained as:

$$u'(p\Delta f_x,q\Delta f_y) = u(p\Delta x,q\Delta y)\exp\frac{-jK}{2z}[(p\Delta x)^2+(q\Delta y)^2]$$

$$p,q = 0,1,\ldots,N-1 \qquad (2.53)$$

The image field is calculated by obtaining the inverse Fourier transform of the function u' in eqn. (2.53). If r, ℓ are the point indices and $\Delta x_o, \Delta y_o$ the sample spacings in the x_o and y_o directions in the image plane, then the image complex field is given by:

$$u_o(r\Delta x_o, \ell\Delta y_o) = \frac{1}{N^2} \sum_{q=0}^{N-1} \sum_{p=0}^{N-1} u'(p\Delta f_x, q\Delta f_y) \exp\{+j2\pi[(p\Delta f_x)(r\Delta x_o)$$

$$+ (q\Delta f_y)(\ell\Delta y_o)]\}$$

$$r, \ell = 0,1,\ldots,N-1 \qquad (2.54)$$

The sample spacings in the image plane are:

$$\Delta x_o = \frac{1}{N\Delta f_x} \quad , \quad \Delta y_o = \frac{1}{N\Delta f_y} \qquad (2.55)$$

Substituting f_x and f_y from eqn. (2.52),

$$\Delta x_o = \frac{\lambda z}{N\Delta x} \quad , \quad \Delta y_o = \frac{\lambda z}{N\Delta y} \qquad (2.56)$$

The image intensity distribution $u_o'(r\Delta x_o, \ell\Delta y_o)$ is the square of the modulus of the result in eqn. (2.54).

. Backward Wave Propagation (Frequency Domain) Reconstruction Algorithm [29],[74],[76]

In the Fresnel integral method the diffraction equations have been manipulated so that the propoagation process, apart from a quadratic phase multiplicative factor, is represented by a Fourier transform which enables image reconstruction at a reasonable speed using the FFT algorithm, eqn. (2.51). However, the approximations made, eqn. (2.43), in order to arrive at this form restrict the use of this method to large distances compared to the aperture size, eqn. (2.75), thus making it inappropriate for near-field imaging and for high resolution, large numerical aperture imaging systems. These systems are required for many applications such as medical diagnostics where there is a delicate balance between attenuation and resolution which determines the operating range.

This restriction on the imaging range can be overcome by using the backward wave propagation method based on the frequency domain representation of the diffraction process. In the spatial frequency domain the propagation phenomenon can be regarded as a linear dispersive spatial filter with a finite spatial bandwidth [33]. Referring to Fig. 2.9; let $U_o(f_x,f_y)$ and $U(f_x,f_y)$ be the spatial Fourier transforms of the complex object and hologram distributions $u_o(x_o,y_o)$ and $u(x,y)$ respectively, i.e.

$$U_o(f_x,f_y) = \iint_{-\infty}^{\infty} u(x_o,y_o)\exp[-j2\pi(f_x x_o+f_y y_o)]\, dx_o dy_o \qquad (2.57)$$

$$U(f_x,f_y) = \iint_{-\infty}^{\infty} u(x,y)\exp[-j2\pi(f_x x+f_y y)]\, dxdy \qquad (2.58)$$

Propagation from the object plane to the hologram plane at distance z from the object plane can be expressed in the frequency domain by the relationship:

$$U(f_x,f_y) = U_o(f_x,f_y)\, H(f_x,f_y) \qquad (2.59)$$

where $H(f_x,f_y)$ is the propagation transfer function, given by [33]:

$$H(f_x,f_y) = \exp\{j \frac{2\pi z}{\lambda} [1-(\lambda f_x)^2-(\lambda f_y)^2]^{\frac{1}{2}}\} \qquad (2.60)$$

For spatial frequencies such that $(\lambda f_x)^2 + (\lambda f_y)^2 < 1$ the transfer function is a phase shift which depends on z, the propagation distance. For wave components having $(\lambda f_x)^2 + (\lambda f_y)^2 > 1$ the transfer function becomes a negative exponential which severely attenuates these waves as they propagate. These evanescent waves decay rapidly and they become negligible after propagating a few wavelengths. This makes the detection of such waves practically impossible at the hologram plane, particularly in optics where the wavelength is very short. Because of their high spatial frequencies, these waves carry information on fine details in the object which would be lost leading to some image degradation. The limiting case $(\lambda f_x)^2 + (\lambda f_y)^2 = 1$ represents the waves propagating normal to the z axis and hence contribute no net power flow in the z direction. The propagation transfer function can therefore be simplified as follows:

$$H(f_x,f_y) = \begin{cases} \exp\{j \frac{2\pi z}{\lambda}[1-(\lambda f_x)^2-(\lambda f_y)^2]^{\frac{1}{2}}\} & f_x^2+f_y^2 < \frac{1}{\lambda^2} \\ 0 & f_x^2+f_y^2 \geq \frac{1}{\lambda^2} \end{cases} \quad \ldots(2.61)$$

In the inverse (backward) propagation problem, the image spectrum can be obtained from the complex hologram spectrum by solving eqn. (2.59) for $U_o(f_x,f_y)$,

$$U_o(f_x,f_y) = H^{-1}(f_x,f_y)\ U(f_x,f_y) \quad (2.62)$$

Assuming that no evanecent waves are recorded at the hologram plane, the transfer function for the inverse propagation is given by:

$$H^{-1}(f_x,f_y) = \begin{cases} \exp\{-j \frac{2\pi z}{\lambda}[1-(\lambda f_x)^2-(\lambda f_y)^2]^{\frac{1}{2}}\} & f_x^2+f_y^2 < \frac{1}{\lambda^2} \\ 0 & f_x^2+f_y^2 \geq \frac{1}{\lambda^2} \end{cases} \quad \ldots(2.63)$$

The image in the space domain $u_o(x_o,y_o)$ is obtained by taking the inverse Fourier transform of the result obtained in eqn. (2.62).

The steps required to perform image reconstruction by backward wave propagation in the frequency domain are illustrated in Fig. 2.11 and can be summarized as follows:

1) Taking the Fourier transform of the complex hologram data

2) Multiplying the complex hologram spectrum by $H^{-1}(f_x,f_y)$

evaluated at the required reconstruction distance z. This step corresponds to focusing in the frequency domain.

3) Taking the inverse Fourier transform of the result to obtain the image complex field. The image intensity is the square of the modulus of this field.

Referring to Fig. 2.11, the hologram field $u(x,y)$ consists, in the discrete form, of NxN samples spaced at Δx, Δy in the x and y directions. Its complex spectrum $U(f_x,f_y)$ consists of the same number of samples spaced in the frequency domain at Δf_x, Δf_y in the f_x,f_y directions where:

$$\Delta f_x = \frac{1}{N\Delta x} \qquad , \qquad \Delta f_y = \frac{1}{N\Delta y} \qquad (2.64)$$

If p,q are the point indices in the hologram space in the x,y directions respectively and m,n the point indices in the frequency domain in the f_x, f_y directions respectively, then the discrete Fourier transform operation is given by:

$$U(m\Delta f_x,n\Delta f_y)= \sum_{q=o}^{N-1} \sum_{p=0}^{N-1} u(p\Delta x ,q\Delta y)\exp\{ -j2\pi[(p\Delta x)(m\Delta f_x)+(q\Delta y)(n\Delta f_y)]\}$$
$$m,n=0,1,\dots,N-1 \qquad (2.65)$$

Multiplying by the inverse propagation transfer function to obtain the image complex spectrum yields:

$$U_o(m\Delta f_x,n\Delta f_y) = \begin{cases} U(m\Delta f_x,n\Delta f_y)\exp\{ -j\frac{2\pi z}{\lambda}[1-(\lambda m\Delta f_x)^2-(\lambda n\Delta f_y)^2 \} & m^2(\Delta f_x)^2+n^2(\Delta f_y)^2 < \frac{1}{\lambda^2} \\ 0 & m^2(\Delta f_x)^2+n^2(\Delta f_y)^2 \geqslant \frac{1}{\lambda^2} \end{cases}$$

$$m,n=0,1,\dots\dots, N-1 \qquad (2.66)$$

The image field in the space domain is obtained by calculating the inverse Fourier transform of the result in eqn. (2.66),

$$u_o(r\Delta x_o, \ell\Delta y_o) = \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} U_o(m\Delta f_x, n\Delta f_y) \exp\{+j2\pi[(m\Delta f_x)(r\Delta x_o)$$

$$+(n\Delta f_y)(\ell\Delta y_o)]\} \quad r,\ell = 0,1,\ldots,N-1 \qquad (2.67)$$

where r,ℓ are the point indices in the image space in the x_o and y_o directions respectively. The image field therefore consists of NxN points spaced at Δx_o, Δy_o in the x_o and y_o directions, where

$$\Delta x_o = \frac{1}{N\Delta f_x} \quad , \quad \Delta y_o = \frac{1}{N\Delta f_y} \qquad (2.68)$$

Substituting for Δf_x and Δf_y from eqn. (2.64) yields,

$$\Delta x_o = \Delta x \quad , \quad \Delta y_o = \Delta y \qquad (2.69)$$

Comparing eqns. (2.69) and (2.56) shows that, contrary to the Fresnel integral method, the samples in the image plane are spaced at the same spacing in the hologram plane and therefore the dimensions of the image space do not vary with the reconstruction distance z.

Although the frequency domain method requires an extra FFT operation compared with the Fresnel integral method, there is no limit on the minimum diffraction distance used which makes this method useful for many imaging applications. Moreover, because the hologram and image spectra are available as intermediate steps in the reconstruction process, a wide range of frequency domain image processing techniques can be easily applied, e.g. [63]-[67]. These techniques include filtering, edge enhancement, deconvolution of the detector aperture function, and enhancement of signals corresponding to small objects in the image.

2.6 Design Considerations in Holographic Imaging Systems

The performance of a holographic imaging system is determined by a set of parameters such as resolution, field of view, range, and image quality. In this section these parameters are briefly discussed with emphasis on the way they are influenced by the design specifications such as the aperture size, the spatial sampling frequency, the pulse width, etc., and the various trade-offs involved. In order to narrow down the field of the discussion, reference will be made to the case of an underwater system which employs a sound transmitter to insonify the object and a filled receiving array of piezoelectric elements to sample the complex hologram. The image is reconstructed digitally using one of the algorithms discussed in section 2.5.2.

2.6.1 Resolution

An important parameter in any imaging system is the lateral resolution which determines the minimum distance between two points in the object plane beyond which the images of the two points cannot be resolved. Assume a hologram aperture of size D. For an object close to the hologram axis at a distance z from the hologram, the lateral resolution δ_ℓ is given by [35]:

$$\delta_\ell = \frac{\lambda z}{D} \qquad (2.70)$$

This shows that the lateral resolution increases with the increase in the numerical aperture of the hologram. A more detailed analysis of resolution is presented in section 3.2.

Another useful measure of resolution is that along the normal to the object plane. This range, or depth, resolution is a measure of the ability of the imaging system to resolve objects close in range. It also determines the depth of focus, over which parts of an object can be brought into focus for lateral imaging at one time without need for refocusing. According to one criterion [35] the range resolution is defined as the axial distance from the focal plane which causes 20% drop in the image intensity and is given by:

$$\delta_R = 2\lambda(\frac{z}{D})^2 \tag{2.71}$$

All parts of an object within $\pm\delta_R$ from the focal plane can be considered in focus at one time and therefore the depth of focus, d_f, is given by:

$$d_f = 4\lambda(\frac{z}{D})^2 \tag{2.72}$$

As in photography, there are conflicting requirements for increased depth of focus and increased lateral resolution. To be able to image sections of an object which has a large extension in the depth direction without the need for refocusing, a large depth of focus is required. This can be achieved by reducing the aperture size, D, at the expense of the lateral resolution and the resulting images will not be sharply defined.

To achieve the best image resolution, the size of the numerical aperture in an imaging system must be as large as possible. This will be usually limited by physical and economic constraints. Increasing the operating frequency increases the numerical aperture for a given physical size but may require extra array elements to match the increase in the spatial frequency content of the hologram. Moreover, increasing the frequency tends to limit the operating range due to increased attenuation. In a practical system the resolution may be limited by noise considerations rather than imaging criteria. For example in a pulsed underwater system, due to the increased attenuation at large ranges, the minimum target size which reflects enough energy to be detected at the receiving array with signal-to-noise ratio greater than unity exceeds the lateral resolution predicted by the Rayleigh criterion and therefore sets the limit for the lateral resolution of the system [77].

From eqns. (2.71) and (2.72), apart from very short ranges, the range resolution and the depth of focus will be unacceptably high. When acoustic pulses are used, the depth of focus will be limited to the thickness of the sector of the object volume which is illuminated by the pulse, i.e.

$$d_f' = c\tau \tag{2.73}$$

$$\delta_R' = c\,\tau/2 \tag{2.74}$$

where c is the speed of sound in the medium and τ is the pulse width. In practice, the minimum value of the pulse width will be limited by the receiver bandwidth and noise considerations. The maximum value is determined by the requirement to reduce the effects of positional stability and object motion on the accuracy of the hologram measurement.

2.6.2 Range

The minimum range of the system will be determined by reverberations, the possibility of saturation in the receiver circuits due to strong returns, and limitations imposed by the reconstruction algorithm employed, cf. sec. 2.5.2. The maximum range will be set by resolution requirements. This is affected by the aperture size, the operating frequency, the transmitter power and directivity, sensitivity and directivity of the array elements, target strength, system noise, pulse length, attenuation, operating frequency, and stability and motion requirements.

2.6.3 Field of View

The field of view is governed by the directivity of both the sound transmitter and the receiving array. It will be further limited by the angle of the reference beam, c.f. eqn. (2.26). For a given system the largest field of view is obtained when a reference beam normal to the hologram plane is used (or simulated). This is the reference implied when the complex hologram is measured relative to a reference signal with a constant phase. In this case the total field of view, $2\theta_{I\,max}$, is obtained from eqn. (2.28) as,

$$2\theta_{I\,max} = 2\sin^{-1}\left(\frac{\lambda}{2\Delta x}\right) \tag{2.75}$$

where λ is the wavelength and Δx is the spacing between the receiving array elements in the x direction. For equal element spacing in both the x and y directions this will be equal to the field of view in the y direction. For small values of $\theta_{1\ max}$ the field of view can be expressed approximately as:

$$2\theta_{1\ max} \approx \frac{\lambda}{\Delta x} \tag{2.76}$$

i.e. the total field of view in radians is approximately equal to the inverse of the element spacing in wavelengths.

The element spacing for a given aperture size will be limited by economic and physical constraints. Moreover, the value of the field of view predicted by eqn. (2.76) is reduced due to the finite size of the receiving elements which limits the maximum spatial frequency that can be sampled, cf. sec. 2.4.1. As the sensitivity of the receiving element increases with the increase in its size, this might lead to some trade-off between the field of view and resolution at large ranges in a noise limited system. Adjacent element cross-talk in the receiving array reduces the independence of the neighbouring processing channels and therefore has the same effect as increasing the area of the sampling detector.

2.6.4 Image Parameters

The number of independent elements in the image will be equal to the number of resolution elements within the field of view. Assuming a square receiving array of N by N elements with equal element spacing in both x and y directions, i.e. $\Delta x = \Delta y = \Delta$. Angular resolution δ_a in each direction is obtained approximately from eqn. (2.70) as δ_ℓ / z for small angles. Substituting the approximate value of $N\Delta$ for the aperture size D yields:

$$\delta_a = \frac{\lambda}{N\Delta} \tag{2.77}$$

From eqns. (2.76) and (2.77), the number of resolution elements within the field of view in each direction is:

$$N_p = \frac{2\theta_{1\,max}}{\delta_a}$$

$$= N \qquad (2.78)$$

The number of resolvable picture elements in the two-dimensional image is therefore approximately equal to N^2, the number of array elements.

Another parameter of particular interest in acoustic holography and acoustic imaging in general is the dynamic range in the image because of the specularity problem associated with the long wavelength and relative smoothness of the surfaces of most objects of interest, cf. sec. 2.3.4. Large dynamic range is usually required for the imaging device in order to show the non-specular returns of much lower amplitude which carry much information about the object. In a practical system, however, the dynamic range in the image is limited by the nonuniformities and the limited dynamic range of the hologram channel processors [78]. It is also influenced by the number of quantization levels used in digitizing the hologram information, cf. sec. 2.5.2.

Image signal-to-noise ratio is determined by the effect of both the spatial and temporal sources of noise in the imaging system. Temporal noise includes ambient acoustic noise and thermal noise in the system electronics. The effect of thermal noise can be reduced by increasing the pulse width although this reduces the range resolution, cf.eqn. (2.74). Spatial noise in the reconstructed image is mainly due to nonuniformities, both in amplitude and phase, in the gains of the channel processors during hologram acquisition. Since each image point is some linear combination of all hologram signal, cf. eqns. (2.54), (2.67), the random spatial noise in the hologram is reduced by a factor equal to the square root of the number of independent channels averaged in the linear combination [79].

2.7 Holography Versus Other Acoustic Imaging Techniques

When holography came into being, acoustic imaging was already a mature art with a long and successful history [4]-[6]. The problem of phase measurement in optics, for which holography offered an ingenious solution, never existed in acoustics because of the availability of linear detectors. Moreover, holographic three-dimensional image reconstruction from acoustic holograms is riddled with distortion problems which make it virtually unrealizable at the frequencies of interest in most applications. In spite of this, acoustic holography has aroused great interest and found a number of applications. It is interesting to note that the new technique has been well integrated with other techniques used for non-holographic imaging. For example section 2.4 illustrates how available scanning techniques have been used for hologram recording and acquisition. Sections 2.4 and 2.5 have shown that factors unique to acoustic radiation, such as the large wavelength and the availability of linear detectors,have made their impact in simplifying the holographic process and increasing its versatility, both in the acquisition and reconstruction phases.

Perhaps the main advantage of the holographic approach is the great simplicity and ease with which the acoustic information about an object field is recorded and processed to produce a visual image. An example for this advantage can be cited from the field of seismic imaging [80]. Before the application of holographic principles, wave equation migration techniques used for seismic exploration required that a full-time trace of each detector in the receiving array be recorded and stored. Holography replaces this requirement by the need to measure only two quantities; the amplitude and phase of the received wave relative to a reference at each detector. This achieves great savings in the amount of data to be stored and processed. In addition to reducing the time and cost of the imaging process as a whole, holography allows for a large variety of method for conditioning, analysing, interpreting, reconstructing and displaying this information. Holograms may be added, subtracted, filtered, correlated, and so on before being reconstructed. This capability allows one to take advantage of modern minicomputers and associated image processing hardware to increase the efficiency and effectiveness of seismic data processing.

Another advantage which is inherent in holography, by virtue of the wavefront reconstruction principle it embodies, is that a hologram is a compact two-dimensional record of a three-dimensional volume. The hologram is by definition a record of the complex field of the object wave at a given plane. According to the principle of wavefront reconstruction this is sufficient to reconstruct the object wave at any plane within the volume which contributed the signals recorded on the hologram plane.

It is important to note that this advantage, which is responsible for the three-dimensional property of the holographic imaging, is not eroded by the problem of depth distortion. In optical reconstruction, the depth distortion manifests itself when the viewing optics are weakly focused in order to view the virtual image of the whole of a three-dimensional object. However, three-dimensional properties of the image can still be observed by bringing into sharp focus various planes of the three-dimensional image so that each can be viewed alone without interference from out-of-focus planes. This is demonstrated for a wire object at an angle with the hologram plane [59] and for letter objects positioned at different ranges [57]. Simulation results described in section 7.5 demonstrate the same criterion for discrete point-objects.

The above advantage effectively means that holography has a very large depth of field which is not limited by the need to focus during the recording process. Focusing, in the sense of selecting a portion of the object to be imaged and excluding the rest of the object volume, is generally achieved in holography during the reconstruction phase. As the hologram contains all the information about the imaged volume, focusing is achieved by simply selecting the reconstruction plane through adjusting the reconstruction optics or inserting the required distance in the computer reconstruction algorithm.

A conventional imaging system, e.g. a microscope, is designed to have high lateral resolution and therefore the depth of focus is very limited. To image a large volume with such a system would require a sequence of continuous refocusing in order to cover the required volume. This is a tedious process especially when the object is continuously in motion.

Holography offers unique advantageous when imaging through aberrating media [33], especially when both the reference and object waves experience the same aberration. This is particularly advantageous in situations causing severe nonlinearity in the propagation of sound waves leading to amplitude and phase distortions in the propagating wavefronts [81].

Acoustic imaging techniques can be broadly divided, according to the image format, into two categories: B-scan, or sonar type, and orthoscopic techniques. In the first category, the image is presented as bearing, or lateral distance, versus depth which is plotted as a function of the time of propagation. In the second group the image is presented in a horizontal versus vertical format which is familiar in optical imaging. Although holography belongs to the second category, it will be briefly compared with the B-scan technique which is widely used in medical applications.

2.7.1 Holography versus B-Scan Techniques

In a B-scan imaging system, Fig. 2.12a, a weakly focused transducer with a long focal length performs a line scan over the volume to be imaged. At every point in the scan a narrow acoustic pulse is transmitted. The echos received by the same transducer are detected and their amplitudes used to modulate the intensity of a display which is scanned in synchronism with the transducer such that the position of the spot on the display screen corresponds to the position in time of the acoustic return causing it, and hence that of the corresponding point in the object.

Fig. 2.12b shows the corresponding arrangement using holography.

A strongly focused transducer with a short focal length scans a line hologram, at the focus plane of the transducer, to record the amplitude and phase of the received waves scattered by the object. It should be noted that the hologram can alternatively be sampled with a small point detector in the hologram plane.

The basic difference between B-scan and holography is that while in the B-scan each sample (time trace) obtained at every position of the scanning transducer is treated separately, holography synthesizes a large aperture by processing the hologram data obtained over the whole scan. Therefore, the B-scan transducer has to provide all the lateral resolution of the system. As the hologram aperture can be much greater than the size of the B-scan transducer, lateral resolution in holography is much greater than that of the B-scan.

Moreover, because adequate lateral resolution has to be maintained over the whole depth of field, the B-scan transducer must be weakly focused at a plane approximately halfway in the sector to be imaged, see Fig. 2.12 a. This further reduces the lateral resolution at shorter and longer ranges, while in holography resolution is improved at shorter ranges. These conflicting requirements for large depth of field and good lateral resolution in B-scan technique influence the design of the scanning transducer and the trade-off affects the system performance, while in the case of holography the transducer can be optimized for proper sampling of the hologram at a fixed plane.

In the depth direction, resolution is determined by the pulse width in the B-scan system. For very short pulses the effect of the receiver noise increases and could swamp returns from weak scattering regions. In holography, as indicated in section 2.6.1, the depth resolution at short ranges is determined by the aperture size. This means that the pulse width can be long enough to allow for the detection of weak targets.

Holography and B-scan techniques can perform complementary roles in acoustic imaging where B-scan offers good range resolution and holographic processing improves the lateral resolution. A number of techniques [82], [83] have been proposed which attempt to combine the two techniques by processing the received signal along the linear B-scans holographically.

2.7.2 Holography versus Other Orthoscopic Imaging Techniques

There are three fundamental approaches to acoustic imaging which are capable of presenting images that are similar in appearance to optical images; namely lens type, beamforming, and holographic techniques. The basic objective in all these imaging techniques is to 'undo' the effect of propagation shown in Fig. 13a, where for simplicity the object consists of two points P_1 and P_2. Through propagation, each point on the imaging plane receives the sum of two wavefronts; one from each point. The process of imaging simply requires this sum to be decomposed to its original components, so that one regains separately the amplitudes of the waves coming from P_1 and P_2.

Mathematically speaking, a linear transform of the individual wavefronts originating in the object plane produces their sums in the reception plane and therefore the imaging process attempts to do the inverse of this transform as shown in section 2.5.2. The process of reversing the effects of propagation is usually referred to as the spatial processing part of image formation. In addition, the imaging process must also include transduction to change from acoustical to electrical energy and detection to convert signals at the high acoustic frequencies to observable DC image signals. Imaging techniques differ primarily in the way the spatial processing is performed and the order in which the three operations of transduction, spatial processing, and detection are performed in time, Fig. 2.14.

In a lens imaging system a lens is used to perform the spatial processing required to form the acoustic image, Fig. 2.13b. This is followed by transduction of the acoustic energy into electrical signal at the acoustic frequency by a receiver array at the focal plane of the lens. The received signal is then converted to DC values representing the image intensity to be displayed, Fig. 2.14a.

A beamformer system attempts to perform the spatial processing function of the lens electronically. Therefore, transduction is performed first to convert the acoustic signals at the imaging plane into electrical signals at the acoustic frequency. The focusing effect of the lens is emulated by effectively introducing different time delays into the high frequency signals received by the various elements of the receiving array in the imaging plane. These are added together and the resulting signals detected to produce DC signals corresponding to the image for every point in the image field, Figs. 2.13c and 2.14b.

In holography, transduction is also the first step as in beamforming systems. Contrary to beamforming techniques however, the electrical signals at the acoustic frequency are converted directly to a set of DC values corresponding to the hologram, therefore detection is performed at an earlier stage. This is followed by spatial processing of the holographic DC signals in the form of hologram reconstruction, Fig. 2.14c.

Although the three techniques of acoustic imaging perform the same basic operations, the performance and the complexity of each technique depend to a large extent on the order in which these operations are performed. The practical realization of each technique suffers from a number of limitations due to the restricted aperture size of acoustic arrays, finite sampling of the wavefield, as well as amplifier noise, nonlinearities, and gain errors (both in amplitude and phase). The effect of such limitations on the image quality will be greatly influenced by the position of the acoustic array in the imaging process and the order of appearance of linear and non-linear operations. Following is a brief comparison between holography and both lens and beamforming systems.

. <u>Holography Versus Lens Systems</u>

The main advantage of lens acoustic imaging systems [84] is their simplicity and inherent broad-band nature. The acoustic array and processing electronics can be fairly simple since they only need to detect the intensity of the incident acoustic radiation at the image plane. On the other hand, acoustic lenses with adequate numerical apertures can be very bulky because of the large acoustic wavelength; especially at low frequencies and in applications such as underwater viewing. As the size of the lens increases, the focal lens increases which generally implies large size and bulkiness for the imaging system as a whole. Other problems include the difficulty in moving the lens or the receiving array to accommodate different focusing ranges, the limited depth of field, and the image distortion due to aberrations in the lens.

Comparison [85] between the effects of practical limitations on both lens type and holographic imaging system indicates a number of advantages for the holographic technique. These advantages stem mainly from the fact that transduction, electronic amplification, and detection in holography are performed before the linear spatial processing while in a lens system these are performed in the image plane. This causes noise spikes, non adjacent cross-talk and dead channels - a common defect in acoustic arrays - to be averaged out in the reconstruction process in holography and therefore cause only some reduction in image contrast while in a lens system these can give the appearance of false objects.

. Holography Versus Beamforming Systems

Perhaps the real competition facing holography, particularly in the field of underwater imaging and non-destructive testing, comes from beamforming imaging systems [86]. One of the basic limitations of all beamforming systems, however, is that the spatial processing is performed at the acoustic frequency, while in holography the temporal aspects of the received signals are discarded immediately upon the detection of the hologram information at a very early stage in the imaging process, Fig. 2.14 b,c. As the acoustic frequency increases, the cost, complexity, and power consumption of the signal processing in beamforming systems increases rapidly. Due to the same reason the signal processing in beamformers has to be performed in the proximity of the acoustic array used for the transduction. This increases the bulk and power consumption of the part of the system which has to be carried, for example, by the diver in nondestructive testing applications for offshore inspection. On the other hand, image reconstruction in holography can be separated from data collection. In addition to simplifying the portable part in such applications, this also allows the use of sophisticated computing at a later stage for image reconstruction, processing, analysis, and storage. Moreover, the ability to use general purpose computers reduces the need for special purpose hardware.

Another difficulty in beamforming imaging is that focusing at other than one or a very few intermediate ranges is difficult in most types of beamformers, since some form of variable time delay or phase shift is necessary with varying range [87]. In holography, however, because of the large depth of field, focusing at any plane within the imaged volume can be easily achieved by adjusting the reconstruction optics or inserting the required distance in the computer reconstruction algorithm. Moreover, while holography is suitable for near field imaging, most beamformers are geared to plane wave processing and therefore near field imaging causes considerable difficulties for such systems. A comparison [88] between holography and beamforming techniques shows that holography has a number of advantages in terms of signal-to-noise performance and processing complexity in many types of applications.

2.8 Applications

Acoustic holography has found applications in a number of fields which include nondestructive testing, underwater imaging, medical diagnostics,and seismic processing. Applications have also been reported in microscopy [89] and imaging of noise sources [90]. In this section the main features of each application are reviewed with emphasis on the advantages offered by holography compared to other techniques and the areas where improvements are required.

2.8.1 Nondestructive Testing (NDT)

This was one of the first areas where the use of holography has been investigated [91]. Compared to conventional pulse-echo techniques used in NDT, holography has the advantage of providing an image of the actual defect which allows more accurate sizing and quantitative evaluation to be made, in addition to better resolution and greater sensitivity. An experimental investigation [92] on actual defects in steel has been made in order to compare the results obtained from acoustic holography and conventional pulse- echo techniques with the actual shape and dimensions obtained by destructive analysis. It was found that the holography dimensions were closer to the true dimensions than were those determined by the other techniques .

Moreover, holographic measurements made at different frequencies were in much better agreement than were measurements made at different frequencies using the conventional methods. This is because variations in frequency or in the size of the transducer in a conventional system alter the radiation pattern of its narrow beam. Similarly, variations in frequency together with variations in the shape, orientation, and the surface roughness of the defect alters its effective scattering pattern. The response of the conventional system, therefore, varies over a wide range because of the interaction between these two beam profiles while in holography a wide beam is used for sampling the hologram and the results are less dependent on variations in the directivity patterns. In addition, the greater amount of data recorded in the hologram has an inherent capability of providing more information on defect shape and orientation which is more reliable than that obtained from conventional techniques, hence allowing for more accurate characterization.

Liquid surface transmission holography has been used for imaging relatively small movable objects which can be trans-illuminated [35]. However, the majority of holographic imaging systems for NDT employ some form of mechanical scanning in the reflection mode, particularly when the test piece is accessible from one side only. Both optical [55] and computer[93] reconstruction have been used in such systems. NDT holographic systems for underwater inspection [94] use piezoelectric arrays for sampling the hologram. Recently a system has been described [95] which combines microprocessor and colour graphics display technologies with the principle of acoustic holography to provide high speed, high resolution imaging of flaws using the frequency domain approach for image reconstruction.

Although the majority of NDT holographic techniques use the imaging feature of holography, application beyond imaging are possible. A number of techniques [96], [97] detect and estimate changes in the thickness of the test peice by observing the changes in the fringe spacing and counting the number of fringes on the hologram without reconstructing the image. In much the same way as in optical holography, interferometric techniques have been investigated in the acoustic domain [98] for deformation analysis and other applications for the investigation of internal defects which cannot be detected by similar optical techniques. In deformation analysis a double exposure acoustic hologram is made with deformation occuring between exposures. When reconstructing the hologram both the images of the object before and after deformation are obtained. These two images interfere and the resulting interference fringes mark the deformation.

The majority of the problems related to the use of holography in NDT are related to the complexity of sound propagation in solids [99] such as multiple reflections, mode conversion at boundaries, and the scattering nature of the medium especially in weld areas which are of interest. Because of the coherent nature of holography these effects tend to accentuate the problems of speckle and diffraction artefacts which interfere with the structural details of the images thus making their interpretation more difficult. These effects can be reduced if the degree of coherency of the insonifying source is slightly lowered by sweeping the frequency over a narrow bandwidth [100]. The problem of mode conversion can be reduced by insonifying the object normal to the surface. Time-gating can also be used to exclude shear waves as they are much slower than longitudinal waves and to discriminate against multiple reflections. However, because of the strongly scattering and inhomogenous nature of the propagation medium, there is a limit on the amount of improvement that can be achieved by time-gating.

2.8.2 Underwater Imaging

Holographic imaging techniques constitute a major part of the underwater imaging systems reported in the literature. In this area of application, holography offers a number of advantages over other competitive techniques which include lens and beamforming systems, cf. sec. 2.7.2. These advantages make holography more suitable for use in the hostile underwater environment since the holographic systems are more rugged and require no moving parts compared to lens systems. They are also lighter and consume less power compared to beamforming systems.

Because of the large ranges involved and the relatively low frequencies used, to reduce attenuation, large apertures are required in underwater applications to obtain adequate resolution. This makes arrays the most suitable candidate for hologram sampling since the relative motion between the hologram plane and the object and the varying nature of the environment make mechanical scanning unsuitable in most cases. To reduce the cost of arrays, underwater holographic systems usually measure the complex hologram data, cf. sec. 2.4.2. Systems reported during the last few years use filled arrays [78], [101] and the more economical and advanced technique of synthetic aperture [48], [102],[103] , see section 3.3.

In most cases, the underwater imaging system is attached to a moving platform or carried by a diver and therefore the system must be fairly insensitive to relative motion between the object and the target. For this reason parallel acquisition of the hologram data is preferred, where each element in the receiving array has its own channel processor so that the spatial distribution of the complex field over the hologram is obtained simultaneously. Pulsing the transmitted acoustic signal and gating the receivers helps in reducing the stability requirements, but this also affects the noise performance of the imaging system, cf. sec. 2.6.1. Another problem associated with underwater acoustic imaging in general is that of the specular reflections. The conventional solution of diffuse illumination using a number of widely spaced insonifying sources cannot be properly implemented due to the limited size of the imaging platform.

2.8.3 Medical Diagnostics

The majority of the applications of acoustic holography in medicine which are reported in the literature [30], [104], [105] employ the liquid surface technique, mainly because of its real-time imaging capability which is an important requirement in medical applications. The disadvantages of this technique include low sensitivity and poor spatial resolution in sampling the hologram, in addition to the inconvenience of immersing the parts to be imaged in water. Solid surface interface methods using laser scanning and optical reconstruction have also been utilized to investigate the use of holography for imaging various types of body tissues and organs and for obstetrics applications [54].

Scanning with the more sensitive piezoelectric detectors has been used to image certain parts of the body such as the eye [106] in the reflection mode. A contact scanner similar to the type used in B-scan techniques has also been reported in the literature [39]. This system uses a linear piezoelectric array which is mechanically scanned in the direction perpendicular to its length. The contact scanner overcomes the inconvenience of using a liquid bath and increases the resolution for the same hologram aperture.

In spite of the advantages of increased lateral resolution and better sensitivity in holography compared to B-scan techniques, the application of holography in the medical field is still rather limited. This can be attributed to the fact that the advantages offered do not match the increased processing complexity that holography entails compared to the well-established and relatively simple B-scan methods. The time delay required for optical reconstruction of the holographic images is a disadvantage. However, this can be overcome to a certain extent by using array sampling of the hologram together with fast computer reconstruction. A more serious problem associated with holographic medical imaging is caused by the coherent nature of the technique and the inhomogenious nature of the human body as a propagation medium. In addition to specular reflections, the image usefulness is degraded by the presence of a number of coherence products which tend to be confused with tissue texture and add spurious image details. This makes the images more difficult to interpret than in the case of imaging hard scatterers in a more homogenous medium such as in underwater applications.

2.8.4 Seismic Imaging

The application of holography to the field of seismic imaging offers a number of advantages. The data collection phase of the process is greatly simplified since the recording of time traces of the signal recieved at the array of geophones is simply replaced by the measurement of the amplitude and phase of the signal. Moreover, holography provides great flexibility in conditioning the original data in various ways to improve or enhance certain aspects of the reconstructed image, particularly when computer reconstruction is employed. This signal processing is of particular importance for the proper interpretation of images in seismic applications because of the complex nature of sound propagation in earth.

Seismic imaging systems reported in the literature utilize both one-dimensional [107] and two-dimensional [56] arrays and employ both optical and computer reconstruction [56]. Because of the large aperture required, due to the large ranges and the very low frequencies used, some forms of synthetic aperture techniques have been utilized [108]. A large number of data processing techniques have been employed in seismic holographic imaging to improve the quality of the images obtained and to extract more useful information from them [64], [80], [109].

2.9 Limitations

In spite of the large interest shown in holography and the diverse potential applications of the technique, the penetration of holography into operational systems both in military and commercial fields is still rather limited. This is partially due to the fact that before the advent of holography many conventional acoustic methods, such as the B-scan technique were already well established [110] and they continued to give reasonable performance with relatively simple equipment. Moreover, being an acoustic, coherent imaging system based on the principle of wavefront reconstruction; holography has a number of limitations attributed to each of these descriptions. A large number of these limitations, however, are associated with the conventional technique of recording the hologram as an interference pattern and reconstructing it optically. Such limitations, therefore, can be reduced to a large extend by measuring the complex hologram data and reconstructing it numerically using the computer. The use of the computer also opens the door for a wide range of signal processing techniques which improve image quality and resolution.

The most important drawback pertinent to all acoustic imaging methods, and not only to the holographic approach, arises from the small numerical apertures for a given physical size because of the large acoustic wavelength. This greatly reduces the information content on the imaging aperture and severely affects both lateral and range resolution. The significance of this limitation becomes clear when one realizes that working with an acoustic aperture of 300 wavelengths is equivalent to viewing objects in the visible world through a 0.2 mm pinhole. This problem is even further accentuated by the lack of suitable media for detecting the acoustic field with adequate sensitivity and spatial resolution and which can be extended over large areas, cf. sec. 2.4.3. Large physical apertures obtained using two-dimensional arrays are costly and mechanical scanning of the aperture area with one detector is time consuming. A number of methods have been suggested for increasing the effectiveness of a given physical aperture which include synthetic aperture techniques and algorithms for expanding the aperture or increasing the resolution. The problem of limited resolution in acoustic holography is the main concern of the work reported in this thesis.

Another problem related to the large acoustic wavelength is that of the specular reflections, cf. sec. 2.3.4. This is a more serious problem than in optics because ordinary objects are smoother relative to the acoustic wavelength and therefore act as mirrors to the insonifying waves. The specularity problem reduces the intelligibility of acoustic images especially when the dynamic range of the display device is limited, since only those parts producing large reflections could be seen. This problem can be reduced by using more than one source to insonify the object from different angles in order to provide more diffuse illumination. However, this is not always practical, as in the case of underwater applications. In addition, diffuse illumination tends to increase the speckle problem.

The coherent nature of holography gives rise to a number of image artefacts which degrade the appearance of the image, interfere with its structural details, and make it more difficult to interpret; especially in medical applications where they tend to be confused with other image details [54]. These artefacts result mainly from mutual interference of signals from one part of the object with those from another, and include speckle, edge ringing, and diffraction effects, cf. sec. 2.3.4. Since the size of the speckle is approximately equal to the size of the resolution cell of the imaging system, this problem is more severe in acoustics because of the small numerical apertures. Moreover, these effects also increase when the medium of propagation is inhomogenous and scattering and the object to be imaged has complex configuration leading to multiple reflections and mode conversions. For the same reason speckle tends to increase with diffuse illumination when attempting to remedy the problem of specular reflections. Speckle and other coherence artefacts are considered to be one of the most serious problems in the holographic approach to acoustic imaging. Since all these artefacts are products of coherency, they can be reduced by slightly reducing the coherency of the acoustic source by sweeping its frequency over a small bandwidth [100].

As a wavefront reconstruction technique, acoustic holography requires the recording of the amplitude and phase of the complex field at the hologram plane. The need for accurate phase measurement requires a degree of stability for the imaging set-up and constrains the relative movement between the object and the hologram plane which may be somewhat restrictive in certain applications such as in underwater imaging [77]. The stability requirement, however, is much more relaxed than in the case of optical holography because of the large acoustic wavelength. The situation is improved when the hologram data is sampled instantaneously using two-dimensional filled arrays with a separate channel processor for each array element in order to achieve parallel hologram acquisition with a single transmitted pulse. However, this is usually not economical with large apertures where some form of aperture synthesis is employed using multiple transmissions. Pulsing the acoustic signal and limiting reception to within a narrow receiver gate reduces the stability requirement but this tends to affect the noise performance of the imaging system and could therefore limit resolution at large ranges, cf. sec. 2.6.1.

The dependence of holography on phase information makes it also sensitive to phase distortions in the medium caused by nonlinear sound propagation at high intensities. However, in certain configurations the holographic technique can tolerate aberration in the medium which are produced by such phase distortions [33]. These configurations assume the use of a physical reference which experiences aberration in the same medium and therefore this advantage is not valid when the reference wave is simulated.

In its conventional form, acoustic holography shares a number of problems with optical holography which are related to both hologram recording the reconstruction. A large number of holographic acoustic imaging systems have adopted the technique of recording the hologram as an interference pattern with a reference which provides the spatial carrier and reconstructing such a hologram optically. As shown in previous sections, this technique is neither essential nor advantageous in acoustics since the hologram image can be reconstructed using the computer from the complex hologram data. The requirement for a spatial carrier introduces a conjugate image together with extraneous interference terms in the reconstructed image, cf. sec. 2.3.

Furthermore, the spatial carrier frequency introduced by the reference increases the spatial bandwidth of the hologram. Accordingly, the hologram bandwidth can be up to four times the object bandwidth, cf. eqn (2.14). Therefore, for the same image quality of a complex object this type of holography requires up to a four fold increase in the spatial resolution of the area detector compared to the case of a lens system. This limitation can be avoided by measuring the complex hologram data since in this case the data bandwidth does not exceed the object bandwidth.

Optical reconstruction introduces a considerable time delay in non real-time imaging techniques, and this can be disadvantage in many applications such as medical diagnostics. Additionally, the depth distortion caused by the large mismatch between the acoustic and the optical wavelengths, together with the small acoustic numerical apertures, severely limit the prospects for adequate three dimensional viewing of acoustic images, cf. sec. 2.5.2.

Both problems of long time delays and wavelength mismatch can be circumvented by going to computer reconstruction, although three-dimensional image displays generated by the computer are not as realistic as those obtained in optical holography. The efficiency and speed of computer reconstruction are optimized when the reconstruction algorithm is applied to the complex hologram data. As electronic data processing becomes faster and cheaper it is possible that near real-time operation can be achieved. Additionally, the same microcomputer used for image reconstruction can also be used as the controller for the hologram acquisition, therefore simplifying the system and reducing the need for specialized hardware. A large variety of signal and data processing techniques can also be employed to reduce the effect of other limitations and to improve the quality of the resulting images.

2.10 Conclusion

Holography offers a wide range of possibilities in acoustic imaging together with a number of advantages compared to other techniques in the field. With fast, inexpensive, and more powerful microcomputers and special signal processing and display devices becoming available, a number of the classical problems associated with the conventional type of acoustic holography can be overcome. The key for this improvement is the replacement of the hologram recording by the measurement of the complex object field and the optical reconstruction by computer reconstruction. In this way the problem of wavelength distortion and the time lag associated with optical reconstruction can be avoided, with the possibility of near real-time imaging using dedicated signal processors. The system hardware can be kept to a minimum with the computer performing all the control functions during hologram acquisition, reconstruction, and display.

The computer can also help in reducing the effect of more basic problem such as the limited numerical apertures, the specular reflections, and the speckle. A large number of signal processing methods can be used to perform synthetic aperture, aperture expansion, or object restoration techniques in order to improve the resolution obtainable from a given limited aperture. This would also reduce the effects of speckle on the image quality. Image processing techniques can reduce the specular appearance of acoustic images, help extract certain features from them, enhance weak returns, and perform focusing automatically. In short, combining holographic principles with the capabilities of modern microprocessors and signal processing devices should make it possible for holography to overcome a good deal of the problems it has been facing in its conventional form.

The work reported in this thesis is mainly concerned with the problem of limited resolution in acoustic holography. Chapter 3 reviews the techniques for resolution improvement in imaging systems in general. The new method described in Chapters 4 and 5 is based on expanding the hologram aperture using modelling and prediction techniques. The experimental implementation reported in Chapters 6 and 7 employ a microcomputer system for hologram acquisition, reconstruction and display.

CHAPTER 3

TECHNIQUES FOR RESOLUTION IMPROVEMENT

3.1 Introduction

All imaging systems are limited in resolution by the dimensions of the entrance pupil, the reocrding aperture, or the data window which govern the capture of information by the input stage of the system. However, the possibility of resolution beyond the diffraction limit has been recognised in optics for a number of years [33] and a number of techniques have been developed for this purpose [111], both in the optical, microwave, and acoustical domains.

In acoustical and long-wavelength holography in general the need for such techniques is even greater since numerical apertures are normally much smaller than those available in optics because of the large wavelength. In holography, the small numerical aperture not only limits resolution but also increases the speckle effect in the holographic image and adds to the problems associated with three-dimensional imaging, cf. sec. 2.9. The lack of suitable area detectors in acoustics has meant that sampling the hologram aperture with arrays or by mechanically scanning a single detector are the most suitable methods which provide the required sensitivity and spatial resolution. However, to cover large areas using such methods is either costly in the case of arrays or time consuming when mechanical scanning is used. In underwater applications, allowed scan times are often restricted by the stability of the imaging plateform, the speed of the objects to be imaged, and the fluctuating nature of the propagation medium. Additionally situations arise in the field of acoustic imaging where the available aperture is physically restricted; for example when imaging the inside of the chest through narrow spaces between the ribs.

On the other hand, there is a number of factors which allow a wide variety and a large amount of signal processing techniques to be employed in acoustics in order to improve resolution and image quality in general. The availability of linear detectors which are capable of sensing the complex amplitude of the acoustic signals allow more signal processing possibilities than in the case of optics where square-law detection is the general rule. Moreover, because of the relatively small amounts of data in the case of acoustics, large amounts of signal processing can be performed using general purpose computers and specialized processing hardware. This is particularly the case when using the complex hologram, since it corresponds to the minimum amount of data representing the hologram, cf. sec. 2.5.2. With the increasing use of the computer as the main component for system control and image reconstruction, such signal processing tasks can be carried out easily and cost effectively. It is expected that the revolution in microprocessor and signal processing devices will have a major impact on this field.

Techniques for resolution improvement can be classified into synthetic aperture methods and signal processing methods. Although synthetic aperture methods employ some techniques of signal processing, the above classification is useful as it differentiates between other important aspects of the two classes of techniques. Synthetic aperture methods are primarily based on modifying the way in which the data acquisition phase of the imaging process is performed so that a large aperture is covered using a relatively small number of transducers. The signal processing class of techniques operate on a given physical aperture on which the data acquisition has already been performed and attempt to improve the image resolution obtainable from such an aperture. However, some signal processing techniques might impose certain requirements on the imaging range or the detector sensitivity; see sec. 3.5.2. Following a formal derivation of the aperture limited resolution in holography, the techniques for resolution improvement in the fields of optics, microwaves, and acoustics are discussed with emphasis on those methods which are relevant to acoustic holography in general and the work described in this thesis in particular.

3.2 Aperture Limited Resolution in Acoustic Holography

As in the case of all imaging systems, both lateral and range resolution in holography are limited by the size of the hologram aperture. Consider a hologram aperture which extends ℓ_x in the x direction and ℓ_y in the y direction. A point object at (x_o,y_o) in the object plane at distance z from the hologram is represented by a Delta function $\delta(x_o,y_o)$. If the distance z is large enough for the far-field (Fraunhofer region) approximations to apply then the quadratic phase terms in eqn. (2.48) can be omitted. Ignoring constant amplitude and phase terms, the hologram of the point source is:

$$u(x,y) = F[\delta(x_o,y_o)] \tag{3.1}$$

where F denotes the Fourier transform evaluated at frequencies $f_x = \frac{x}{\lambda z}$ and $f_y = \frac{y}{\lambda z}$. Up to a multiplicative constant which is a function of the object range z, the hologram function over the aperture is therefore given by:

$$u(x,y) = 1 \quad ; \quad -\frac{\ell_x}{2} \leqslant x \leqslant \frac{\ell_x}{2} \;, \quad -\frac{\ell_y}{2} \leqslant y \leqslant \frac{\ell_y}{2} \tag{3.2}$$

In the equivalent frequency space this is:

$$u(f_x,f_y) = 1 \quad ; \quad -\frac{\ell_x}{2\lambda z} \leqslant f_x \leqslant \frac{\ell_x}{2\lambda z} \;, \quad -\frac{\ell_y}{2\lambda z} \leqslant f_y \leqslant \frac{\ell_y}{2\lambda z} \tag{3.3}$$

The image is obtained by the inverse Fourier transform of the hologram function, cf. eqn. (2.51).

$$u_o(x_o,y_o) = \int_{-\frac{\ell_y}{2\lambda z}}^{\frac{\ell_y}{2\lambda z}} \int_{-\frac{\ell_x}{2\lambda z}}^{\frac{\ell_x}{2\lambda z}} 1.\ \exp\{+j2\pi[x_of_x+y_of_y]\} \quad df_x\, df_y \qquad \ldots(3.4)$$

$$u_o(x_o,y_o) = \frac{\ell_x \ell_y}{(\lambda z)^2} \quad \text{sinc}\left(\frac{\ell_x x_o}{\lambda z}\right) \text{ sinc}\left(\frac{\ell_y y_o}{\lambda z}\right) \tag{3.5}$$

The image intensity is therefore:

$$I = I_o \ \text{sinc}^2 \left(\frac{\ell_x x_o}{\lambda z} \right) \ \text{sinc}^2 \left(\frac{\ell_y y_o}{\lambda z} \right) \qquad (3.6)$$

where I_o is the peak intensity. The intensity function has its first zeros in the x_o and y_o directions at a distance from the centre given by:

$$\delta x_o = \frac{\lambda z}{\ell_x} \quad , \quad \delta y_o = \frac{\lambda z}{\ell_y} \qquad (3.7)$$

The width of the image point is $2\delta x_o$ and $2\delta y_o$ in the two directions. The larger the dimensions of the hologram aperture the sharper the point image becomes and the better the resolution. Lateral resolution is usually defined as the minimum distance between two points in the object plane below which it is not possible to resolve their images. The most common resolution criterion in incoherent optics is the Rayleigh criterion which states that two incoherent points are barely resolved when the centre of the image of one source falls on the first zero of the image of the other . In this case the combined image of the two points has a central dip of about 19% of the maximum intensity. According to this criterion the resolution is the distance between the peak of the image of one point and its first zero, as given by eqn. (3.7). For a square aperture of size D the resolution is:

$$\delta_\ell = \frac{\lambda z}{D} \qquad (3.8)$$

in both x_o and y_o directions. With coherent radiation the situation is further complicated by the effect of relative phase between the two point sources [33]. When the illuminating source in a coherent system is a number of wavelengths in size the Rayleigh criterion can be useful as an approximation to the resolution of the system [35].

The integration limits in eqn. (3.4) indicate that the real limitation to resolution is caused by the fact that the hologram is measured only at a finite aperture relative to the wavelength λ and the object range z. The resolution therefore increases with the increase in the numerical aperture ($\frac{\ell_x}{\lambda}$ and $\frac{\ell_y}{\lambda}$) and with the decrease in the range z. It must be noted, however, that resolution as derived by the above analysis is a far field criterion. Most imaging systems operate at smaller ranges, in the near field or the Fresnel region. For a point object at such ranges the wavefront at the hologram plane is spherical rather than plane. In order to achieve the aperture limited resolution in eqn. (3.8) at such ranges the imaging system must compensate for the wave curvature, a process referred to as focusing. In a lens system this is achieved by the lens curvature. In the Fresnel region the phase of the incident wave can be approximated by a quadratic phase term. This is compensated for in the Fresnel integral method of image reconstruction by multiplying by the same exponential term with an opposite sign before performing the inverse linear transform, eqn. (2.51).

Since the resolution increases with the increase in the size of the numerical aperture, all techniques which aim at improving the resolution beyond this diffraction limit attempt to extend the effective aperture over which the hologram data is collected. In the case of imaging in the far field, the hologram is equivalent to the Fourier transform of the object, cf. eqn. (3.1). Therefore the extension of the hologram aperture is equivalent to extrapolating the object spectrum beyond the limit set by size of the physical aperture.

3.3 Synthetic Aperture Holographic Techniques

In its broad sense, the term "synthetic aperture" refers to the simulation of a filled array by the clever use of some smaller array. The price paid for the saving in the number of transducers used may include increased scan time or lower signal-to-noise ratios. Both of these effects may lead to some image degradation. The simplest form of synthetic apertures are those obtained by scanning a single transducer in a certain format to cover a large aperture as commonly used in holography, Fig. 3.2a. Although it is customary to scan the receiver, it has been shown [35] that similar results are obtained with the scanning of either the transmitter or the object, or with the simultaneous scanning of both the transmitter and the receiver.

This simple form is also the basis of synthetic aperture radar [112] and sonar [113] which use coherent radiation in a way that is very similar to holography. In a typical arrangement for a synthetic aperture radar, as the aircraft moves along a straight path, its radar continuously emits successive microwave pulses. The signal returns are detected along the path and mixed with a coherent reference and the result recorded on a photographic film. This photographic record is developed and reconstructed optically using a laser as in the case of conventional holography. Because of the coherent nature of the process, the returned echos appear to be received by a single antenna which has a long synthetic aperture equal to the distance travelled. This effective large antenna length provides the very high resolution required for such applications as aerial photography.

A number of methods have been employed to reduce the time required to scan a two-dimensional hologram aperture. A linear receiver array can be mechanically scanned in a direction perpendicular to its length [39], Fig. 3.1b. Approximately equivalent results were obtained using a fixed array of the type mentioned above and moving a broad-beam transmitter in a direction perpendicular to the array [107], Fig. 3.1c.

Greater saving in time can be achieved using two crossed linear arrays in the arrangement known as Well's cross [114] which simulates simultaneous scanning of both transmitters and receivers. This arrangement, Fig. 3.1d, uses two perpendicular arrays; one containing N transmitters and the other N receivers. The transmitter elements are fired sequentially and for each transmission the returns are received by all the receiver elements. For large ranges the signal received by the qth receiver from a transmission by the pth transmitter will be approximately equal to the signal received by the (p,q)th element of the equivalent filled receiver array of N^2 elements with a transmitter located at the point of intersection of the extensions of the two linear arrays. Because of the geometry employed, the resulting images suffer from conical aberration which is more pronounced at short ranges in addition to another form of aberration due to transmitter scanning. Moreover, this configuration uses an equal number of transmitters and receivers, which is not necessarily the optimum arrangement from the point of view of cost reduction since transmitter channels are generally more expensive than receiver channels.

Another approach to aperture synthesis is to mechanically scan a small two-dimensional filled array of receivers in order to cover the required large area, Fig. 3.1e. However, it is difficult in practice to displace such an array rapidly and accurately. Because of the transmitter/receiver duality [35], similar results are obtained if the receiver array is kept fixed and the transmitter is scanned by displacing it a distance equal to the size of the receiver array aperture in order to cover the area required as shown in Fig. 3.1f.

To reduce the data acquisition time even further and to dispense with mechanical motion altogether, practical systems using this principle [43] usually employ a sparse transmitter array with element spacing equal to the size of the filled receiver array, Fig. 3.1g. The transmitter elements are sequentially fired and the corresponding portion of the hologram data is received and stored. Due to the fact that scanning is effectively split between the transmitter and the receiver, a phase correction should be introduced to the data collected during each transmission. This correction is introduced into the subholograms obtained for each fired transmitter element. These subholograms are superimposed to obtain the synthesized hologram which is then reconstructed. Alternatively, equivalent phase shifts are introduced into the subimages reconstructed from the individual subholograms which are then superimposed to obtain the final imgage [47]. In practice, there will be some phase uncertainties between the contributions from the individual transmissions due to limited mechanical stability and environmental variations. These cause alignment errors in the synthesized aperture and lead to image degradation.

The disadvantage of sequential transmission is that large scan times are required which make it difficult to image moving objects and increases the stability requirements. A number of techniques have been proposed for parallel processing of the hologram information. In one approach [47], the transmitter elements are fired simultaneously but the signals driving various elements have different frequency or modulation coding so that the received signals can be decomposed into the individual subholograms. Another approach [48], [103] uses the transmitter array in a beamformer configuration where the phase of the signal fed to each transmitter element is selected so that the grating lobes of the transmitted beam are focused at a number of points in the object plane. The subimage corresponding to the data received by the small filled array is reconstructed. The phasing of the transmitter array is then changed so that the grating lobes are steered and focused a new set of points on the objects and the process is repeated until the whole image is obtained. Although this method achieves better tolerance to the problems of mechanical stability and object motion and to the phase variations due to the medium [48], this is achieved at the expense of increased complexity in the transmitter array circuitry.

The amount of improvement in resolution and reduction in cost in the synthetic aperture systems which employ a sparse transmitter array and a small filled receiver array is determined by the number of elements in the transmitter array. Generally, the size of the transmitter array will be limited in practice by the increase in acquisition time, the reduction in range resolution, and the increased complexity of either the beamforming circuits or the coding and decoding circuits when parallel acquisition is used.

Another way of saving in the number of transducers required to sample a given hologram aperture is to use nonuniform sampling [79], Fig. 3.1h. This technique is based on random sampling and achieves a considerable reduction in the number of array elements for the same effective aperture size without the effects of undersampling, especially when a priori information exists on how the hologram information is distributed over the aperture. In this case a sampling scheme should be employed which samples most in the most important regions [115]. One disadvantage of this 'thinning' of the array is that because of the reduction in the number of array elements, the spatial signal-to-noise ratio in the reconstructed image is also reduced, cf. sec. 2.6.4. It should also be noted that conventional reconstruction algorithms assume uniform sampling of the hologram.

3.4 Object Restoration Techniques

It has been recognised for some time now that diffraction does not impose an absolute limit on the resolving power of an imaging system. Toraldo di Francia [116] was the first to suggest in 1952 that the classical expression of 1.22 (λ/D) which has been always accepted as a theoretical resolution limit proves to be instead only a practical limit set by the system noise and the accuracy of the measuring instruments. For example, the image of two points, however close to one another, is mathematically different from that of one point. Therefore, the resolution limit which causes the two points to be mistaken for one point is actually imposed by the failure of practical image receptors to sense the difference between the images in the two cases. Using the principles of information theory Toraldo proposed a new definition for the resolution criterion as the number of independent data which can be found in the image. He also introduced the concept of the ambiguous image, according to which if two different objects have identical images they can not be resolved. Toraldo has indicated that the key to resolution beyond the diffraction limit lies in utilizing as much a priori information as possible about the object.

Previously, Schelkunoff [117] described the principles of super-gain antennas for microwaves. These results were transferred to optics by Toraldo who discussed a procedure for designing a super-resolving pupil in which improved resolution could be achieved over a limited field by modifying the pupil of a diffraction limited imaging system. However, for any substantial improvement in resolution, the tolerance on the pupil transmittance function would be severe.

When imaging an object in the far field, the propagation of the waves from the object to the imaging aperture approximates to a spatial Fourier transform and therefore the imaging system receives the spectrum of the object. Due to the limited aperture in diffraction limited systems, only a finite portion of the object spectrum is intercepted by the imaging aperture. The object spectrum is therefore truncated to a cut-off frequency f_c which is determined by the aperture dimensions relative to the wavelength, i.e. the numerical aperture, together with the object distance,cf. eqn. (3.3). During the imaging process, the inverse Fourier transform is performed on this truncated spectrum and the resulting image will have limited resolution due to the loss of information corresponding to the high frequency components outside the passband of the imaging system.

The same argument can be used to explain the finite resolution in diffraction limited systems using the concept of ambiguous image. Consider a set of different objects each having some fine details so that their true spectra extend beyond the cut-off frequency f_c of the imaging system. Since it is possible that some of these objects will have the portions of their true spectra within the passband of the system identical, their reconstructed images based on such an aperture will be identical and therefore cannot be resolved as two different scenes, Fig. 3.2. Accordingly, two points which are close enough may look like a 1-point object when imaged with a diffraction limited system. To be able to image beyond the diffraction limit and therefore resolve such ambiguities, it is necessary to effectively increase the aperture size in order to increase the cut-off frequency f_c or in other words extend the knowledge of the object spectrum beyond f_c.

Following up the idea of using a priori information as a key to super-resolution, Harris [118] has shown that a priori knowledge that the object is spatially bounded, i.e. finite in size, is sufficient to resolve completely the ambiguity in the object-image relationship. His work is based on a number of theories related to analytic functions [119] which indicate that the spectrum of a spatially bounded object is an analytic function. It is a characteristic of analytic functions that if the function is known exactly in an arbitrarily small, but finite, region of a certain plane then the entire function can be found uniquely by means of analytic continuation. This implies that a knowledge of the object spectrum within the passband of the imaging system can be used to determine the full spectrum of the object beyond the passband without any ambiguity with other spectra corresponding to other objects.

Since the true extended spectrum can be used to reconstruct the true object, infinite resolution can be achieved in principle for spatially bounded objects. In practice, however, the ability to continue the object spectrum beyond the cut-off frequency will be limited by noise and measurement accuracy since with these practical limitations it is not possible to know the object spectrum exactly over the passband. In general, the process of analytic continuation suffers from the problem of ill conditioning [120] where small errors at the input produce large errors at the output. The effect of noise and measurement errors depends on the type of analytic functions used. Unfortunately, many types of such functions require almost perfect precision of definition to make very small extensions to the spectrum.

Harris described an analytic continuation scheme based on the sampling theorem. Consider a finite object which is bounded in space by a rectangle extending over $\pm L_x$ and $\pm L_y$ in the x and y directions respectively. According to the Whittaker-Shannon sampling theorem the object spectrum $G(f_x,f_y)$ can be written in terms of its samples at $\left(\frac{n}{2L_x}, \frac{m}{2L_y}\right)$ as:

$$G(f_x,f_y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} G\left(\frac{n}{2L_x}, \frac{m}{2L_y}\right) \mathrm{sinc}\left[2L_x\left(f_x - \frac{n}{2L_x}\right)\right] \mathrm{sinc}\left[2L_y\left(f_y - \frac{m}{2L_y}\right)\right] \quad \ldots(3.9)$$

where sinc(x) denotes $\sin(\pi x)/\pi x$. In a practical system the values that can be readily measured at any point on the imaging aperture are those corresponding to $G(f_x,f_y)$ on the L.H.S. of eqn. (3.9). Due to the limited passband of the system, the sample values $G\left(\frac{n}{2L_x}, \frac{m}{2L_y}\right)$

can be found for only a few low integer values of (n,m). To extend the spectrum to larger integer values $n=\pm N$, $m=\pm M$, eqn. (3.9) can be approximated to:

$$G(f_x,f_y) \approx \sum_{m=-M}^{M} \sum_{n=-N}^{N} G(\frac{n}{2L_x},\frac{m}{2L_y}) \,\text{sinc}[2L_x(f_x-\frac{n}{2L_x})]\,\text{sinc}[2L_y(f_y-\frac{m}{2L_y})] \quad \ldots(3.10)$$

which gives a better representation of the spectrum in at least some finite region beyond the passband. To determine the sample values $G(\frac{n}{2L_x}, \frac{m}{2L_y})$ over the extended passband, the values of $G(f_x,f_y)$ are measured at any (2N+1)(2M+1) points within the actual passband. Substituting $G(f_x,f_y)$, f_x, and f_y in eqn. (3.10) produces (2N+1)(2M+1) linear equations inthe same number of unknowns , $G(\frac{n}{2L_x}, \frac{m}{2L_y})$. These can be solved to produce the required sample values which extend beyond the passband.

Harris applied this technique to image a numerically simulated two-point object. He achieved a resolution equal to 0.2 of the Rayleigh limit while ignoring the effect of noise. In practice, however, imperfections in the knowledge of the spectrum within the passband can be greatly magnified when the system of linear equations is solved for sample values outside the passband. This is because each sample value is expressed as a linear combination of all the measured values, each of which contribute some noise. It is estimated [111] that a three-fold improvement in resolution beyond the diffraction limit using this approach may require at least 30 db signal-to-noise ratio in the measured signal while the signal to noise amplitude ratio in the restored image can be as low as 0.7.

Barnes [121] adopted a more formal approach to the problem of object restoration in a diffraction limited imaging system. For the case of one-dimensional system, the imaging process is characterized by an integral of the form:

$$b(s) = \int g(s,x)\ a(x)\ dx \qquad (3.11)$$

where x and s are the coordinates in the object space and image space respectively, a(x) and b(s) are the complex amplitudes of the object and the image respectively, and g(s,x) is the imaging kernel. An object restoration scheme attempts to recover the object function a(x) in terms of the image function b(s) and the imaging kernel g(s,x).

For a finite object, so that both the object and the image functions are bounded by $-\alpha$ and α , and using a set of orthonormal eigenfunctions, the restored object is given as [120]:

$$a(x) = \sum_{n=0}^{\infty} \int_{-\alpha}^{\alpha} \frac{\sigma_n(x)\ \sigma_n(s)}{\lambda_n}\ b(s)\ ds \qquad (3.12)$$

where λ_n are the eigenvalues and $\sigma_n(x,s)$ are normalized prolate spheroidal wave eigenfunctions. When the number of terms in the summation in eqn. (3.12) is truncated to a finite value an approximate value of a(x) is obtained.

One disadvantage of this method is that it generates very strong sidelobe responses outside the area of interest in the object plane. In principle, this is not a serious problem as long as illumination is restricted to that area. In a practical implementation, however, this means that great care should be taken to exclude stray light from parts of the object space that are outside the region of allowed illumination. Moreover, Rushforth and Harris [122] have shown that due to the amplification of errors in the reconstructed function, the usefulness of this method for resolution improvement using realistically measured data is very limited.

Another technique for analytic continuation has been described by Biraud [123]. This is based on an additional a priori information about finite objects, namely that their radiance is a non-negative function. This method has been used to give resolution improvement of 5 times the Rayleigh limit at an amplitude signal-to-noise ratio of 36 in the input data.

The noise limitations are a common factor among all the restoration techniques mentioned above. A method designed to be more robust against noisy environments has been described by Gerchberg [124]. This is an iterative technique for continuing a given segment of the spectrum of a finite object. Ignoring the effects of noise, the known segment of the true spectrum is considered as the sum of the true wider spectrum and an error spectrum which is equal and opposite to the true spectrum outside the extent of the known segment, see Fig. 3.3. The method is based on reducing the energy of this error function which is implicit in the truncated spectrum. The technique discriminates against noise and distortion in the input data on the basis that, considered as additive functions, noise and distortion are generally not capable of being continued to yield a finite object while the true spectrum will.

Referring to Fig. 3.4, the known portion of the spectrum, $G_0(f)$, is first Fourier transformed to yield the diffraction limited object which is then corrected by setting all of the image points outside the known extent of the true object to zero. The image thus modified is Fourier transformed to generate a spectrum which now corresponds to a larger imaging aperture. This spectrum is corrected by substituting the values of the original measured spectrum only where they belong, while the new data is used outside this region. These steps are iterated until a criterion based on the estimated object energy outside the known extent of the true object is satisfied. The continued spectrum, $G_n(f)$, after the nth cycle of the iterative process is given by:

$$G_n(f) = G_0(f) + [G_{n-1}(f) * X \mathrm{sinc}(Xf)][1-\mathrm{rect}(f/F)] \quad (3.13)$$

where $G_{n-1}(f)$ is the spectrum at the end of the preceding iteration, X is the known or estimated extent of the object, and F is the extent of the known segment of the spectrum. In the above equation '*' denotes convolution in the spatial frequency domain. This algorithm has been shown to converge for finite objects when the object extent is known exactly or is overestimated since, in this case, the correction process both in the object and spectrum domains reduces the error energy in every cycle of the iteration.

It should be noted that this method together with all the object restoration techniques mentioned above require an estimation of the extent of the true object to be made. In this particular method and in the Harris technique, caution should be exercised to be sure that the value chosen does completely encompass the object. When the estimated value is greater than the extent of the true object the precision of both methods deteriorates even for the case of noiseless data. In an example given by Gerchberg [124], the energy of the residual error in the restored object increases from 10% when the exact object extent is used to 34% when the object extent is overestimated by a factor of 45%. Moreover, Gerchberg's technique fails if the extent of the object is underestimated.

A statistical approach to the problem of object restoration which is based on maximum entropy techniques has been proposed by Frieden [125] for the case of incoherent imaging. The method operates on the intensity distributions of both the image and the object which are assumed to have positive or zero values. The restored object is considered as the object distribution which is most likely to occur together with the set of noise values in the measured image data. Therefore, the method is based on maximizing the product of the number of ways in which the object and noise distributions can occur, subject to constraints of the measured image distribution and the given total intensity of the restored object. This corresponds to maximizing the entropy of both the object and the noise. The restoring formula is not intrinsically band limited and therefore resolution exceeding the Rayleigh limit can be achieved.

Let the object to be restored be represented by the intensity distribution:

$$O(x_j) = O_j \quad ; \quad j=1,2,\ldots,J \tag{3.14}$$

where x_j is the location of the jth pixel in the object space and

J is the total number of pixels in the restored object. The total object intensity, P_o, is assumed to be known,

$$P_o = \sum_{j=1}^{J} O_j \tag{3.15}$$

The measured values of the image intensity is known over a limited number of M pixels; M<J is further degraded by additive noise. For a linear system this is given by:

$$I_m = \sum_{j=1}^{J} O_j \, S(y_m,x_j) + n_m \; ; \; m=1,2,\ldots,M \tag{3.16}$$

where y_m is the location of the mth pixel in the image space, $S(y_m,x_j)$ is the impulse response of the imaging system and n_m is the noise contribution. Since the algorithm operates on non-negative quantities only, biased values N_m for the noise are used,

$$N_m = n_m + B \tag{3.17}$$

where B must be equal to or greater than the maximum negative value of the noise components. The restored object and the corresponding noise distribution are given by [125]:

$$\hat{O}_j = \exp[-1-\mu- \sum_{m=1}^{M} \lambda_m \, S(y_m,x_j)]; \; j=1,2,\ldots,J \tag{3.18}$$

and

$$\hat{N}_m = \exp[-1- \lambda_m/\rho] \; ; \; m=1,2,\ldots,M \tag{3.19}$$

where ρ represents the signal-to-noise uncertainty which is considered acceptable. The constants μ and λ_m ; $m=1,2,\ldots,M$ are obtained by substituting eqns. (3.18) and (3.19) into (3.15) and (3.16) and using eqn. (3.17). The resulting M+1 equations are highly non linear but can always be solved provided that the noise bias value, B, is sufficiently large. It should be noted that in addition to the measured image data and the a priori knowledge of the total object intensity, P_o, this method requires the two parameters ρ and B to be estimated. The actual values used for these parameters influence the restoration considerably [125]. This method has the advantage of being tolerant to noise for values up to 40% relative amplitude and of the lack of spurious oscillations or details in the restored image. It is interesting to note that although this method adopts a different approach in attacking the restoration problem, it achieves the same result of matching the spectrum of the restored object with that of the true object outside the optical passband.

More recently, a similar approach has been used to improve the resolution of B-scan images which is limited by the spread of the ultrasonic beam [126]. As in the case of Frieden's method, this technique is applied on the magnitude data only. Due to the non-linearity of the entropy function, an iterative scheme is employed for its constrained maximization which requires substantial computing time when the method is applied to two-dimensional images. The maximum entropy technique has also been employed by Yokota et al [127] in an active incoherent ultrasonic imaging system which achieves high resolution and overcomes the problem of speckle in coherent systems.

3.5 Signal Processing Techniques for Resolution Improvement in Acoustic Holography

A number of signal processing techniques have been developed for improving the resolution of a given hologram aperture beyond the diffraction limit set by the aperture size. One technique attempts to estimate the hologram in an extended area from the data obtained in a limited area in a manner which is similar to the methods used for object restoration in optics [128]. Another is based on increasing the effective numerical aperture of a given hologram area by artificially increasing the temporal frequency of the acoustic signal [129]. Another group of techniques increase the resolution in the near-field beyond the wavelength limit. This includes the use of a modified reconstruction formula as suggested by Williams et al [130], [131], or utilizing a priori information about the object [132].

3.5.1 Resolution Improvement Beyond the Diffraction Limit

As shown in section 3.2, the resolution of the holographic imaging system is defined by the limited size of its aperture. Sato et al [128] have described a sequential method for resolution improvement by extending a given hologram aperture. The aperture extension is treated as an estimation problem with the objective of determining an estimate of the hologram in an extended area from the knowledge of the hologram signal in a given smaller area. The estimation algorithm is performed on a modified version of the hologram signal which corresponds to the spatial spectrum of the object. Due to its relevance to the work described in this thesis, Sato's technique is summarized below.

Referring to Fig.3.5, assume a one dimensional object which is bounded in space by the limits $\pm X_o$. According to the sampling theorem, the object spectrum, and therefore the corresponding modified hologram signal, should be sampled at a frequency interval not exceeding Δf_o ; where Δf_o is given by:

$$\Delta f_o = \frac{1}{2X_o} \tag{3.20}$$

Assume that the maximum spatial frequency of the object is limited to $\pm N\ \Delta f_o$, beyond which it vanishes. In practice, the hologram is sampled at a frequency interval Δf which is less than Δf_o. Samples are taken at frequencies $m\Delta f$ where $m=0,\pm1,\dots,\pm M$, and $M \leqslant N$. The sampling theorem relates the hologram signal $H(m\Delta f)$ to the independent samples of the band-limited object spectrum $A\ (n\Delta f_o)$; $n=0,\pm1,\dots,\pm N$ by:

$$H(m\Delta f) = \sum_{n=-N}^{N} A(n\Delta f_o)\ \mathrm{sinc}[\frac{1}{\Delta f_o}\ (m\Delta f - n\Delta f_o)] \qquad (3.21)$$

$$m=0,\pm1,\dots,\pm M$$

where $\mathrm{sinc}(x) = \sin(\pi x)/\pi x$. The above equation is similar to eqn. (3.10) which corresponds to the case of a two-dimensional object. Eqn. (3.21) can be represented in the following vector form:

$$H_M = S\ A_N \qquad (3.22)$$

where H_M is the measured hologram vector and A_N is the overall vector representing the true hologram, or object spectrum, sampled at the critical frequency interval Δf_o ; see Fig. 3.5

$$H_M = [H(-M\Delta f),\dots,\ H(M\Delta f)]^T \qquad (3.23)$$

$$A_N = [A(-N\Delta f_o),\dots,\ A(-M\Delta f_o),\dots,A(M\Delta f_o),\dots A(N\Delta f_o)]^T \qquad \dots(3.24)$$

where T denotes vector transpose. S is a rectangular matrix having (2M+1) rows and (2N+1) columns, with its (i,j)th element given by:

$$S_{ij} = \mathrm{sinc}[\ \frac{1}{\Delta f_o}\ (i\Delta f - j\Delta f_o)] \qquad (3.25)$$

The objective is to obtain an estimate for the truncated vector A_M of the true hologram:

$$A_M = [A(-M\Delta f_o),\dots,\ A(M\Delta f_o)]^T \qquad (3.26)$$

The linear mean-square estimation problem is therefore formulated as follows : Given the hologram data H_M, derive a linear estimate H'_M for the truncated true hologram vector A_M:

$$H'_M = E\, H_M \tag{3.27}$$

where E is a square matrix of the (2M+1) dimension which has to be determined so as to minimize a criterion function J_M,

$$J_M = || H'_M - A_M ||^2 \tag{3.28}$$

where $||\ \ ||$ denotes the norm of the difference vector. The minimization is subject to the condition in eqn. (3.22). It can be shown [128] that the mean square estimator matrix E is given by:

$$E = S'^{-1} \tag{3.29}$$

where S' is a square matrix whose elements are defined by:

$$S'_{ij} = S_{ij} \qquad i,j=-M,..,M\ ;\ M \leqslant N \tag{3.30}$$

The estimated hologram vector H'_M will contain 2M+1 samples, i.e. the same number of samples in the measured hologram H_M. However, since H'_M is determined at a larger sampling interval Δf_o, it will extend over a larger area than the original hologram, hence improving resolution. The hologram expansion ratio γ is therefore given by:

$$\gamma = \frac{\Delta f_o}{\Delta f} \tag{3.31}$$

This shows that resolution improvement is achieved only when the hologram is oversampled.

The estimation process is performed sequentially and the image is obtained by reconstructing the hologram at every stage of the process. The technique has been demonstrated with simulated objects for expansion ratios ranging from 2 to 5. It has been shown that the reconstructed images converge to the required original object as the hologram data is increased.

The effect of disturbing noise in the measured hologram data is magnified through the inverse matrix S'^{-1} , whose norm increases rapidly with the increase in matrix size, therefore increasing the noise sensitivity. This sensitivity also increases with increasing the hologram expansion ratio. Simulation tests using noisy data have indicated that considerable degradation results in the reconstructed images due to noise and the technique fails to resolve the imaged objects satisfactorily even for 1% noise amplitudes. Sensitivity to noise can be reduced by proper choice of the sampling points during the hologram acquisition so that the row vectors of S' have the least dependence in order to decrease the norm of its inverse. Random sampling should be of considerable value in this respect. Using one technique to reduce noise sensitivity, this estimation method gives adequate results for noise amplitudes as large as 20% of the hologram data.

From the above discussion, it can be seen that this technique for hologram aperture extension is very similar to the object restoration methods discussed in section 3.4, particularly the technique by Harris [118]. In both cases, the object is assumed to be spatially bounded and both techniques effectively extrapolate the object spectrum beyond the portion measured by the imaging aperture. This shows the link between techniques aiming at expanding the hologram aperture in acoustics and those used for object restoration in optics. It should also be noted that the nature of the limitations of both classes of techniques due to the presence of noise are quite similar.

Another technique for resolution improvement by increasing the effectiveness of a given physical hologram area through increasing its equivalent numerical aperture has been proposed by Ikeda and Sato [129] . This is obtained through artificially increasing the temporal frequency of the measured hologram signal in order to get a signal which would be obtained if higher frequency acoustic waves were used. This is achieved computationally by nonlinear processing of the received hologram signal so that products having phase terms corresponding to the required higher frequencies are obtained. To prevent the formation of ghost images as a result of unwanted products of the nonlinear opertions, such terms are cancelled out before the computed hologram is reconstructed. The technique is suitable only for a limited number of point objects and it also requires a priori knowledge of the object for the signal processing to be performed. The effectiveness of the method has been demonstrated using computer simulation for objects having up to 3 points using noiseless data.

3.5.2 Resolution Improvement Beyond the Wavelength Limit

According to the $\lambda z/D$ resolution criterion, resolution less than the wavelength of the radiation, λ, is predicted for ranges, z, shorter than the aperture size D. However, the derivation of this criterion was based only on the limitations introduced by the diffraction effects due to the limited size of the hologram aperture, cf. sec. 3.2. The ability to resolve beyond the wavelength is limited by the effects of propagation of the waves from the object to the hologram plane. As shown in section 2.5.2, for spatial frequencies greater than or equal to $(1/\lambda)$ the propagation transfer function is represented by an attenuating factor,

$$H(f_x,f_y) = \begin{cases} \exp +j\,\frac{2\pi z}{\lambda}\,[1-(\lambda f_x)^2-(\lambda f_y)^2]^{\frac{1}{2}} & f_x^{\,2}+f_y^{\,2} < \frac{1}{\lambda^2} \\ \exp -\frac{2\pi z}{\lambda}\,[(\lambda f_x)^2+(\lambda f_y)^2-1]^{\frac{1}{2}} & f_x^{\,2}+f_y^{\,2} \geq \frac{1}{\lambda^2} \end{cases} \qquad \ldots(3.32)$$

Waves corresponding to such high frequencies, known as the evanescent waves, are heavily attenuated and have negligible amplitude after propagating a few wavelengths and therefore are not usually recorded on the hologram. The information carried by these waves about spatial fine details corresponding to the near-field of the object radiation are lost and the image reconstructed from the hologram will therefore be limited in resolution even if an infinite recording aperture were used.

In optical holography the wavelengths are so short that the hologram is necessarily recorded many wavelengths from the object with the result that the evanescent-wave components fall well below the noise level of the hologram recording medium. In acoustics, however, because of the larger wavelength, it is physically possible to record the hologram at distances which can be much smaller than the wavelength when low frequency sound is used.

If the evanescent-wave components are not assumed negligible in the recorded hologram, then the approximate transfer function of the inverse propagation given in eqn. (2.63) should be replaced by:

$$H^{-1}(f_x,f_y) = \begin{cases} \exp -j\,\frac{2\pi z}{\lambda}\,[1-(\lambda f_x)^2-(\lambda f_y)^2\,]^{\frac{1}{2}} & f_x{}^2+f_y{}^2 < \frac{1}{\lambda^2} \\ \exp\;\frac{2\pi z}{\lambda}\,[(\lambda f_x)^2+(\lambda f_y)^2\,-1]^{\frac{1}{2}} & f_x{}^2+f_y{}^2 \geqslant \frac{1}{\lambda^2} \end{cases} \qquad ...(3.33)$$

which should lead to improved resolution since the exponentially increasing amplitude term which becomes effective for frequencies $> 1/\lambda$ takes the strongly attenuated (but non zero) evanescent-wave components and restores them to their exact values in the image plane.

It should be noted that practical realization of eqn. (3.33) is possible only mathematically, since for frequencies $> (1/\lambda)$ this equation represents a wave with an ever increasing amplitude in the z direction. Physical reconstruction methods such as those using laser or acoustic radiation must use forward propagating waves which would be represented by $H^*(f_x,f_y)$ where * denotes the complex conjugate and $H(f_x,f_y)$ is given by eqn. (3.32). $H^*(f_x,f_y)$ agrees with $H^{-1}(f_x,f_y)$, eqn (3.33), for frequencies less than $(1/\lambda)$ but retains the exponentially decaying part for the higher frequency components. This further attenuates any such components which might have been recorded during the recording process. Therefore, super-resolution techniques based on eqn. (3.33) can only be implemented numerically.

One such technique has been proposed by William et al [130], [131] for imaging vibrating plates in the near-field at low frequencies. The effectiveness of the technique has been experimentally demonstrated [131] by imaging two small unbaffled speakers driven at a frequency of 220 Hz and separated by a distance of 0.13 λ and the hologram was measured at a range of 0.06 λ. In the image field calculated using the conventional reconstruction algorithm no details smaller than a wavelength could be observed. When $H^{-1}(f_x,f_y)$ given by eqn. (3.33) was used for the reconstruction, the two point sources representing the speakers were clearly resolved.

For general purpose applications this technique suffers from the requirement for near-field imaging and the use of low frequencies which are not practical in many situations. Additionally, the technique requires more precision in the measurement of the complex hologram signal and therefore needs better sensitivity and larger dynamic range for the detectors and the receiver circuits than in the case of conventional reconstruction.

Another technique for resolving beyond the wavelength limit has been reported by Sato et al [132] for the case of a passive imaging system used for the measurement of the wavefields on the surface of noise emitting objects. Resolution improvement is achieved in this case by utilizing a set of a priori information, on the wavefield such as the objects' shape or the smoothness of the field as well as the propagation law of the waves from the object to the signal detectors. This data is used to calculate the relationship between the objects' wavefield and the hologram signal. The relationship is expressed in a matrix form and the image reconstruction is carried out by solving for the object vector.

For a finite object, super-resolution can be achieved with this technique when large matrices are used and simulation results have indicated that resolution better than the wavelength is obtainable. However, the effect of noise becomes more serious with the increase in the matrix size. The performance with noise depends also on the model used to represent the object when the matrix relationship is constructed. For the object models discussed [132], the signal-to-noise requirement varies from 20 db to 180 db depending on the model type. This indicates that proper choice of the model is indespensible in order to get the required improvement in resolution using this method in practical situations.

3.6 Extrapolation and Data Fitting Techniques

The problem of hologram aperture expansion for increased resolution can be considered as that of extrapolating a band-limited function which is known over a finite region in space. A number of techniques have been proposed for solving this problem in the context of spectral analysis of time-limited signals. These techniques are based on the fact that a bandlimited function is analytic in the entire time axis [133]. By the duality of the Fourier transform, this is equivalent to the fact that the spectrum of a spatially bounded object is analytic, cf. sec. 3.4. Although it is possible to extrapolate the analytic function beyond the known portion through analytic continuation using Taylor-series expansion, this approach has not been favoured in general due to the inherent problem of ill-conditioning [120].

One method of extrapolating bandlimited functions is based on expansion into a series of prolate spheroidal functions [133]. These functions also form the basis of the object restoration technique described by Barnes [121], cf. sec. 3.4. Let g(t) be a time-limited function which is known in the region $-T \leqslant t \leqslant T$. The bandlimited extrapolation of the piece g(t) is given by:

$$f(t) = \sum_{k=0}^{\infty} a_k \, \varphi_k(t)$$

$$; \quad a_k = \int_{-T}^{T} g(t) \, \varphi_k(t) \, dt \qquad (3.34)$$

where $\varphi_k(t)$ are prolate spheroidal eigenfuncions. An approximate value of the extrapolated function is obtained by truncating the series in eqn. (3.34).

Another technique has been proposed by Papoulis [133] . This is based on the error reduction method by Gerchberg [124] described in section 3.4. As a method of spectral estimation, this technique attempts to extrapolate the signal beyond the time interval $-T \leqslant t \leqslant T$ over which the signal is known and therefore obtain a better estimate of the spectrum. This is achieved through an iterative process in which the error energy is reduced by alternatively setting the values of the estimated spectrum outside its known extent $-\sigma \leqslant f \leqslant \sigma$ to zero and setting the values of the estimated time function to the known values within the interval $-T \leqslant t \leqslant T$. This can be considered as the dual of the Gerchberg technique used in the space/spatial frequency domains, since in the case of the Papoulis method the signal is known to be finite in the frequency domain and extrapolation is performed in the time domain. Compared to the prolate spheroidal method, this technique has the advantage that it involves the use of the FFT algorithms which simplify and speed up the computations. Moreover, for large time-bandwidth products, $T\sigma$, the technique converges rapidly and therefore can be terminated early in order to reduce the computation time while this does not reduce the computations in the prolate method. However, being an iterative technique with large numbers of iterations required, especially for small $T\sigma$ products necessitates a large amount of computations and long execution times. Moreover, the method requires the bandwidth σ to be known or estimated, and the accuracy of the extrapolation depends on the accuracy of this estimation.

The aperture expansion technique described in the remaining chapters of this thesis is based on fitting a model for the hologram signal over the available aperture . The most common techniques for data fitting are briefly described here. Because of their simplicity, polynomials form the basis of the majority of data fitting algorithms. The main data fitting methods which use polynomials [134] are those which fit a polynomial model or polynomial splines to a set of given data points.

Let the data given at m points at which the independent variable is $x_1, x_2, \ldots, x_m$ be $y_1, y_2, \ldots, y_m$. The function representing the data based on a polynomial model of order n; $n \leqslant m$ has the general form

$$f(x) = b_1 \varphi_1(x) + b_2 \varphi_2(x) + \ldots b_n \varphi_n(x) \qquad (3.35)$$

where $\varphi_1(x), \varphi_2(x), \ldots, \varphi_n(x)$ are polynomial functions. The model coefficients $b_1, b_2, \ldots, b_n$ are obtained from the following observation equations at the m data points:

$$y_r = b_1\ \varphi_1(x_r)+b_2\ \varphi_2(x_r)+,\ldots,+b_n\ \varphi_n(x_r)$$

$$;\quad r = 1,2,\ldots,m \qquad (3.36)$$

Fitting a polynomial model is equivalent to solving this generally overdetermined system of equations subject to the minimization of an error criterion. Least square error is the most common criterion for realistic data with normally distributed errors. In this case, f(x) is chosen such that:

$$\sum_{r=1}^{m} |f(x_r) - y_r|^2 = \text{minimum} \qquad (3.37)$$

The order of the polynomial determines the closeness of the fit to the actual data points. The higher the order, the closer the function to the data but the more the model function is influenced by errors in the data. Smooth fits are obtained with lower order polynomials, although this could lead to the loss of some fine details in the data function. An interpolating polynomial is that which produces a fit that passes through all the given data points.

In spite of their simplicity, polynomial models in general suffer from ill-conditioning which makes the model sensitive to noise and round-off errors, especially for large model sizes. Spline fitting provides a more generalized form of a single polynomial which enables a wider variety of functions to be modelled and has the main advantage of numerical stability with large numbers of data points. Splines are a set of piecewise polynomials fitted together smoothly. In the general case , splines of order n(degree n-1) consist of N+n basis functions $\psi_{n,j}(x)$. The spline fit S(x) is given by:

$$S(x) = \sum_{j=1}^{N+n} c_j\ \psi_{n,j}(x) \qquad (3.38)$$

where N is the necessary number of knots within the data intervals with n additional external knots and c_j; j=1,2,...,N+n are the fit coefficients. Each basis function $\psi_{n,j}(x)$ is a basis spline (B-spline) of order n in x with knots at $x=\lambda_{j-n},\lambda_{j-n+1},\ldots,\lambda_j$ and is non zero (and positive) only over this interval, i.e.

$$\psi_{n,j}(x) \begin{cases} > 0 & \lambda_{j-n} \leqslant x \leqslant \lambda_j \\ = 0 & \lambda_j < x < \lambda_{j-n} \end{cases} \quad (3.39)$$

The non zero part of $\psi_{n,j}(x)$ consists of n polynomial pieces each of degree n-1. The B-splines are generated using a recurrence relationship [134]. Although splines are more advantageous from the point of view of versatility and numerical stability, they are generally more difficult to use for data fitting. For example, the selection of the knot points at which the polynomial pieces are joined should be made to suit the data. Moreover,this technique requires a larger amount of memory space to calculate the fit compared to other conventional polynomial methods. Since splines tend to be used primarily for data fitting within a given data interval, it is not clear if they are suitable for extrapolating a given function beyond a limited region over which the function is known.

3.7 Conclusion

In this chapter, a number of resolution improvement techniques developed for use in the fields of optics, microwaves, and acoustics have been discussed. The common factor among all these techniques is that they attempt to recover the object spectrum which has been degraded in the imaging process. Some of this degradation is caused by the propagation phenomenon which acts as a low pass filter with an effective cut-off frequency of $(1/\lambda)$. Further degradation is caused by the limited size of the imaging aperture which, in the case of coherent systems, introduces a sharp cut-off at a feequency proportional to $D/(\lambda z)$ where D is the aperture size and z is the range. Another common feature in most signal processing techniques for resolution improvement is the utilization of a priori information on the object; whether explicitly as in the technique by Sato et al [132] or implicitly as in the techniques which attempt to continue the object spectrum beyond the diffraction limit, both in optics [118] and in acoustics [128]. The implied assumption of a spatially bounded object, and therefore the possibility of extrapolating its spectrum, forms the basis of all such systems.

Holographic aperture extension techniques can be viewed as object restoration methods which attempt to gain more knowledge about the object spectrum beyond the limit set by the aperture size. They can also be considered as methods for extrapolating the band-limited hologram function in space. The key for the solution in both cases is the analyticity of the function to be extrapolated. Although analytic continuation using Taylor-series expansion offers a simple solution, it is generally sensitive to noise because of the ill-conditioning problem.

Other techniques which give better performance with noise have also their own limitations. The category of techniques which are based on maximum entropy methods operate on the intensity in an incoherent imaging system [125]. Since these techniques assume positive signals, they cannot be directly applied to the complex hologram data in a coherent system. Moreover the techniques require the knowledge or estimation of a number of critical parameters which influence the performance of the algorithm, cf. sec. 3.4. The same disadvantage exists with the Gerchberg [124] and Papoulis [133] techniques where the object extent, or the hologram bandwidth, must be known or estimated. The effectiveness of both methods depends on the accuracy of this estimation and both techniques fail if this parameter is underestimated. Additionally the principle of error energy reduction, on which both techniques are based, requires a large number of iterations in general with each iteration involving two Fourier transforms. This could lead to long execution times when the algorithm is applied to two-dimensional holograms of practical aperture sizes.

In order to take advantage of the simplicity of direct analytic continuation, the problem of noise sensitivity has to be solved. In the hologram estimation technique by Sato [128], the effect of noise is reduced by modifying the way in which the hologram data is collected. This may prove disadvantageous in a practical implementation of the technique, in addition to the fact that this method of hologram expansion is iterative. In the remaining part of this thesis an algorithm for hologram aperture expansion using analytic continuation is described. This is based on modelling the hologram signal over the given aperture and using the model so constructed to predict the signal outside this aperture. The sensitivity of the solution to noise is reduced by selecting a suitable form of the model without the need for modifying either of the acquisition or reconstruction phases of the holographic process.

CHAPTER 4

EXPANSION OF HOLOGRAM APERTURES USING PREDICTION TECHNIQUES

4.1 Introduction

In the previous chapter it has been shown that techniques for resolution improvement in holography by expanding, computationally, a given hologram aperture are similar to object restoration techniques in optics when the object is in the far field of the imaging aperture. Under this condition, the hologram aperture receives a truncated object spectrum due to the limited aperture size. For a spatially bounded object, the hologram function is therefore an analytic function and hence can be extended beyond the given aperture in the same manner the limited spectrum is continued in object restoration techniques.

However, in the majority of cases holographic imaging is performed at shorter ranges; within the Fresnel zone which extends roughly from D to D^2/λ where D is the aperture size. For such ranges the hologram is not identical to the object spectrum and the analyticity of the hologram function must be established if this criterion is to be used for continuing the hologram in this more general geometry.

In the estimation technique described by Sato [128], the object width is assumed much smaller than the distance between the object and the hologram plane. This is necessary for the superposition integral, eqn. (2.46),to be reduced to the object spectrum. Moreover, the hologram signal is further modified by compensating for phase and amplitude terms so that the measured hologram signal is directly proportional to the spectrum, cf. sec. 3.5.1. These terms contain the range of the object which should be known beforehand for the aperture expansion to be performed. Although the value of the object range is also required as an input to the reconstruction algorithm when the image is numerically reconstructed, the accuracy of the range measurement is not very critical for this application since the image can be reconstructed at a number of ranges and the best image selected. However, as an input to the estimation algorithm, errors in the value of the range will give rise to additional errors in the modified hologram signal and therefore leads to poorer performance in practical systems.

Most analytic continuation techniques utilize the principle of sampling theorem in order to obtain additional samples outside the object spectrum [118] or to estimate a new set of samples based on the minimization of a least mean square error criterion [128]. Due to the requirement for close sampling of the given aperture, the matrix representing the set of linear equations which describe the system usually has an inverse matrix with a very high norm [128], [132]. This leads to errors in the measured hologram data being greatly amplified and therefore limits the effectiveness of these methods when used with realistic data. The formulation of the processing algorithm is rather inflexible and the methods reported for improving the noise performance rely on modifying the data acquisition phase of the imaging process. This may be inconvenient as far as the practical measurement procedure is concerned and may require the use of unconventional reconstruction algorithms as in the case of random sampling suggested by Sato [128].

In this chapter a new method for expanding the hologram aperture using the principle of analytic continuation is presented [135], a copy of this reference is enclosed in Appendix D. The method operates on the hologram function directly without the need for attempts to reduce the problem of aperture expansion to that of continuing the object spectrum. This method is therefore suitable for both Fresnel and Fraunhofer holograms. Moreover, the new technique is not iterative. It takes advantage of the computational simplicity of the direct analytic continuation approach using polynomial and linear models. Compared to conventional analytic continuation techniques [118], [128] this method allows greater flexibility in combating the effects of disturbing noise which are a common feature in all analytic continuation techniques due to their ill-posed nature [120].

In this technique the hologram function at the available aperture is modelled by relating the signal at every sampling point to that at a number of the neighbouring points. The model parameters are obtained by solving the resulting set of equations. Analyticity is employed in the assumption that the model applies at least in a limited region outside the available aperture. The values of the signal at new sample points are predicted point by point by successively applying the model outside the given aperture.

The method avoids any modifications to the measured complex hologram data and therefore any uncertainties in the value of the object range do not contribute to the noise in the hologram data and the value of the object range is required only for image reconstruction. Additionally, the ability to choose the type of the prediction model and hence the structure of the system matrix makes the expansion algorithm more flexible in dealing with the noise problem; without affecting the procedures for data collection and image reconstruction. In practice, the choice of the prediction model will be the result of a trade-off between good prediction accuracy on the one hand and robustness against noise on the other.

In this chapter, the analyticity of the more general form of the hologram function in the Fresnel region is considered. A number of models for predicting the hologram signal outside a given available aperture are described for the case of simple objects consisting of one or two points and an error analysis based on Taylor-series expansion is presented. The method is then extended to the more complex case of quasi-continuous objects containing a large number of closely spaced points which, in the limit, simulate a continuous object. A technique for improving the prediction accuracy is also discussed. This is based on constructing a corrective model by comparing predicted data with corresponding data which is known to be true within the available aperture. This model is then used to correct predicted data outside the available aperture. The effect of aperture expansion on the range resolution is also considered.

The results described throughout this chapter are based on computer-simulated noiseless hologram data. The effects of noise in the hologram data are considered in Chapter 5 and the results with experimental data are described in Chapter 7. Only one-dimensional line holograms are considered for simplicity and the image reconstruction is performed digitally using the backward propagation (Frequency domain) algorithm described in section 2.5.2.

4.2 Analyticity of the Fresnel Hologram Function

Harris [118] has shown that the spectrum of a finite object is an analytic function. This forms the basis for object restoration techniques which aim at extrapolating the object spectrum beyond the passband of the far-field imaging system. Far-field holograms, in the Fraunhofer zone, correspond to the spatial frequency spectrum of the object and therefore the same principle can be utilized for extending the hologram aperture. However, for the more common holograms at shorter ranges, in the Fresnel zone, the relationship between the hologram and the object spectrum is less straightforward. In the following analysis a formal proof of the analyticity of Fresnel hologram is presented. This allows the principle of analytic continuation, and therefore the aperture expansion technique proposed in this chapter to be applied directly to the hologram in this region.

Referring to eqn. (2.48), the hologram signal for an object in the Fresnel zone is given by:

$$u(x,y) = \frac{\exp jKz}{j\lambda z} \exp \frac{jK}{2z}[(x^2+y^2)]F[u_o(x_o,y_o)\exp \frac{jK}{2z}(x_o{}^2+y_o{}^2)] \quad ...(4.1)$$

where K is the wavenumber, z the object range, (x_o,y_o) are the coordinates in the object plane and (x,y) are the coordinates in the hologram plane. $u_o(x_o,y_o)$ and u(x,y) are the object and hologram functions respectively and F denotes the spatial Fourier transform determined at the spatial frequencies:

$$f_x = \frac{x}{\lambda z} \quad , \quad f_y = \frac{y}{\lambda z} \quad (4.2)$$

Let f(x,y) be defined as:

$$f(x,y) \triangleq \frac{\exp jKz}{j\lambda z} \exp \frac{jK}{2z}(x^2+y^2) \quad (4.3)$$

Since the quadratic phase term in the R.H.S. of eqn. (4.3) can be expanded into a convergent power series for z > 0, f(x,y) is an analytic function [136]. Eqn. (4.1) can be written as:

$$u(x,y) = f(x,y)\, F[u_o(x_o,y_o) \exp \frac{jK}{2z} (x_o^2+y_o^2)] \qquad (4.4)$$

In the Fraunhofer region the range z is much larger than the lateral distances in the object domain, i.e. :

$$\frac{K}{2z} (x_o^2+y_o^2) << 1 \qquad (4.5)$$

Therefore, the Fourier transform in eqn. (4.4) reduces to that of the object alone, i.e. to the object spectrum. For a finite object this spectrum is analytic and since f(x,y) is analytic the hologram function is also analytic and therefore analytic continuation can be applied directly on u(x,y) rather than u(x,y)/f(x,y), [128].

Consider the Fourier transform in eqn. (4.4) in the more general case,

$$G(f_x,f_y) \triangleq F[u_o(x_o,y_o) \exp \frac{jK}{2z} (x_o^2+y_o^2)]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u_o(x_o,y_o) \exp \frac{jK}{2z} (x_o^2+y_o^2) \exp[-j2\pi(f_x x_o+f_y y_o)]\, dx_o\, dy_o \qquad \ldots(4.6)$$

Let the finite dimensions of the object be bounded by $\pm X_o/2$ in the x_o direction and $\pm Y_o/2$ in the y_o direction. The object may be completely described over the bounded region by the Fourier series pair:

$$u_o(x_o,y_o) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} G_{mn} \exp \{j2\pi[(\frac{mx_o}{X_o})+(\frac{ny_o}{Y_o})]\} \qquad (4.7)$$

and

$$G_{mn} = \frac{1}{X_o Y_o} \int_{-Y_o/2}^{Y_o/2} \int_{-X_o/2}^{X_o/2} u_o(x_o,y_o) \exp \{-j2\pi[(\frac{mx_o}{X_o})+(\frac{ny_o}{Y_o})]\}\, dx_o dy_o \qquad \ldots(4.8)$$

Substituting $u_o(x_o,y_o)$ from (4.7) into (4.6) and replacing the infinite integration limits in eqn. (4.6) with the spatial bounds of the object, $\pm X_o/2$ and $\pm Y_o/2$; (This is possible since the function of the bounded object vanishes outside this region),

$$G(f_x,f_y) = \int_{-Y_o/2}^{Y_o/2} \int_{-X_o/2}^{X_o/2} \left[\sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} G_{mn} \exp \{j2\pi[(\frac{mx_o}{X_o})+(\frac{ny_o}{Y_o})]\} \right.$$

$$\exp \frac{jK}{2z} (x_o^2+y_o^2)\exp[-j2\pi(f_x x_o+f_y y_o)]dx_o dy_o \qquad ...(4.9)$$

or

$$G(f_x,f_y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} G_{mn} \int_{-X_o/2}^{X_o/2} \exp j[2\pi(\frac{m}{X_o} - f_x)x_o + \frac{K}{2z} x_o^2] dx_o$$

$$\int_{-Y_o/2}^{Y_o/2} \exp j[2\pi(\frac{n}{Y_o} - f_y)y_o + \frac{K}{2z} y_o^2] dy_o \qquad ...(4.10)$$

Consider the integration:

$$I \triangleq \int_{-X_o/2}^{X_o/2} \exp j[2\pi(\frac{m}{X_o} - f_x)x_o + \frac{K}{2z} x_o^2] \; dx_o \qquad (4.11)$$

Let

$$\alpha = 2\pi (\frac{m}{X_o} - f_x) \qquad (4.12)$$

$$\beta = \frac{K}{2z} \qquad (4.13)$$

$$I = \int_{-X_o/2}^{X_o/2} \exp j[\alpha x_o + \beta x_o^2] \; dx_o$$

$$= \int_{-X_o/2}^{X_o/2} \exp j[\beta (x_o + \frac{\alpha}{2\beta})^2 - \frac{\alpha^2}{4\beta}] \; dx_o \qquad (4.14)$$

Let

$$u = x_o + \frac{\alpha}{2\beta}$$

$$u_1 = \frac{-X_o}{2} + \frac{\alpha}{2\beta}$$

$$u_2 = \frac{X_o}{2} + \frac{\alpha}{2\beta} \qquad (4.15)$$

Substituting into eqn. (4.14) yields:

$$I = \exp(-j \frac{\alpha^2}{4\beta}) \int_{u_1}^{u_2} \exp j\ (\beta u^2)\ du \qquad (4.16)$$

which gives:

$$I = \exp(\frac{-j\alpha^2}{4\beta}) [\int_{u_1}^{u_2} \cos(\beta u^2)\ du + j \int_{u_1}^{u_2} \sin(\beta u^2)\ du] \qquad (4.17)$$

The two integrals in eqn. (4.17) are similar to the Fresnel integrals [137]. Consider the first of these integrals,

$$I_1 \triangleq \int_{u_1}^{u_2} \cos(\beta u^2)\ du \qquad (4.18)$$

Expanding the cosine under the integral sign into a power series and integrating term by term yields [137]:

$$I_1 = [u - \frac{\beta^2}{2!5} u^5 + \frac{\beta^4}{4!9} u^9 - \ldots.]_{u_1}^{u_2}$$

$$= (u_2-u_1) - \frac{\beta^2}{2!5} (u_2^{\,5}-u_1^{\,5}) + \frac{\beta^4}{4!9} (u_2^{\,9}-u_1^{\,9}) - \ldots \qquad (4.19)$$

Similarly, the integration:

$$I_2 = \int_{u_1}^{u_2} \sin(\beta u^2)\ du \qquad (4.20)$$

can be expressed as:

$$I_2 = \frac{\beta}{1!3}[u_2^3-u_1^3] - \frac{\beta^3}{3!7}[u_2^7-u_1^7] + \frac{\beta^5}{5!11}[u_2^{11}-u_1^{11}] - .. \qquad (4.21)$$

Both series in eqns. (4.19) and (4.21) are convergent for all finite values of u_1 and u_2.

Following the same approach, the integral in y_o in eqn. (4.10) can be written as:

$$I' = \exp(-j\frac{\alpha'^2}{4\beta}) [\int_{v_1}^{v_2} \cos(\beta v)^2 \, dv + j\int_{v_1}^{v_2} \sin(\beta v^2) \, dv] \qquad (4.22)$$

where

$$\alpha' = 2\pi(\frac{n}{Y_o} - f_y)$$

$$v = y_o + \frac{\alpha'}{2\beta}$$

$$v_1 = \frac{-Y_o}{2} + \frac{\alpha'}{2\beta}$$

$$v_2 = \frac{Y_o}{2} + \frac{\alpha'}{2\beta} \qquad (4.23)$$

The two integrals in eqn. (4.22) have the same form of power series expansions in eqns. (4.19) and (4.21) with u_1 and u_2 replaced by v_1 and v_2 respectively. Substituting in eqns. (4.17) and (4.22) and then into (4.10) produces:

$$G(f_x,f_y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} G_{mn} \exp [\frac{-j}{4\beta}(\alpha^2+\alpha'^2)]$$

$$[(u_2-u_1)+j\frac{\beta}{1!3}(u_2^3-u_1^3)-\frac{\beta^2}{2!5}(u_2^5-u_1^5)-j\frac{\beta^3}{2!7}(u_2^7-u_1^7)$$

$$+\frac{\beta^4}{4!9}(u_2^9-u_1^9)+j\frac{\beta^5}{5!11}(u_2^{11}-u_1^{11})-....]$$

$$[(v_2-v_1)+j\frac{\beta}{1!3}(v_2^3-v_1^3)-\frac{\beta^2}{2!5}(v_2^5-v_1^5)-j\frac{\beta}{3!7}(v_2^7-v_1^7)$$

$$+\frac{\beta^4}{4!9}(v_2^9-v_1^9)+j\frac{\beta^5}{5!11}(v_2^{11}-v_1^{11})-....] \qquad (4.24)$$

Since $\exp[\frac{-j}{4\beta}(\alpha^2+\alpha'^2)]$ can be represented by a convergent power series in f_x and f_y then $G(f_x,f_y)$ is an analytic function and therefore the Fresnel hologram function $u(x,y)$ in eqn. (4.4) is analytic in x and y for values of z in the Fresnel zone. The analyticity of the Fresnel hologram function is useful in that the hologram aperture expansion can be performed, for the case of bounded objects, without the restrictions of operating in the Fraunhofer region. The aperture expansion algorithm can be applied directly on the measured complex hologram data without any modifications.

The above analysis presents a formal proof of the analyticity of the hologram function for an object in the Fresnel zone. A more general approach to investigating the analyticity of the hologram function for any imaging geometry would be to consider the hologram as a truncated space function whose spatial spectrum is band-limited. The band-limited nature of the hologram spectrum is implicit in the use of sampling techniques for the hologram acquisition and the low pass filtering imposed by the propagation process ensures that spectral components at frequencies greater than the cut-off frequency of $(1/\lambda)$ can be practically ignored. In this case, the hologram function is analogous to time-limited functions which are measured only during a finite time window and which have a band-limited spectrum. Since such functions are analytic in the time domain [133], this shows that the hologram function is analytic in space regardless of the object range, given that the hologram spectrum is band-limited.

4.3 Hologram Simulation and Expansion - Basic Definitions

Consider the one-dimensional holographic imaging configuration shown schematically in Fig. 4.1. For simplicity, the object is assumed to consist of a finite number of point sources which radiate acoustic energy coherently. The hologram is sampled at 2M uniformly spaced points in a limited region which represents the available aperture. The sampling points are spaced at a distance Δ wavelength and the distance between the object and the hologram is z wavelengths. The complex hologram signal H(k) at sampling point k on the available aperture is:

$$H(k) = \sum_{i=1}^{\ell} \frac{A_i}{r_i} \exp jKr_i$$

$$= \alpha_k + j\,\beta_k \qquad (4.25)$$

where ℓ is the number of points on the object, r_i is the distance between point k on the hologram and point i on the object, K is the wave number; $K = 2\pi/\lambda$, λ is the wavelength, and A_i is an amplitude factor. α_k and β_k are given by:

$$\alpha_k = \sum_{i=1}^{\ell} \frac{A_i}{r_i} \cos(K\,r_i) \qquad (4.26)$$

$$\beta_k = \sum_{i=1}^{\ell} \frac{A_i}{r_i} \sin(K\,r_i) \qquad (4.27)$$

Unless mentioned otherwise, points on the objects simulated in this chapter are assumed to be of equal radiation strength, i.e. A_i is a constant.

It is assumed that in practice the hologram is obtained by measuring the two quantities α_k and β_k at the sampling points. The measured quantities α'_k and β'_k will contain random components representing noise and can be simulated from the calculated values α_k and β_k as follows:

$$\alpha'_k = \alpha_k + \zeta_k \qquad (4.28)$$

$$\beta_k' = \beta_k + \xi_k \tag{4.29}$$

where ζ_k and ξ_k are white noise processes with zero mean and variance described by the signal-to-noise ratio. This method will be used to simulate noisy hologram data when the effects of disturbing noise are considered in Chapter 5.

The remainder of this chapter together with Chapter 5 are concerned with expanding the given available aperture by fitting a model to the hologram data on this part of the aperture and using the model to predict new additional points on each side of the given aperture in order to extend it from 2M points to 2N points, Fig. 4.1. Correction algorithms which improve the prediction accuracy are also described. The prediction of signals to the R.H.S. of the aperture uses a model which is based on the signals within the R.H. half of the available aperture. Similarly, the L.H. half is used to construct the model to be used for predicting new points to the left of the available aperture. The following definitions are made for the holograms at various parts of the aperture:

. H(k) ; $k \leq M$ is the hologram signal which is simulated or measured over the small available aperture. Plots related to this hologram are marked 'small' throughout the thesis.

. G(k) ; $k \leq N$ is the predicted hologram signal over the total expanded aperture. This consists of the simulated/measured signal over the available aperture and the predicted signals outside this aperture, i.e.:

$$G(k) = \begin{cases} H(k) & k \leq M \\ \text{Predicted signal at point k} & M < k \leq N \end{cases} \tag{4.30}$$

Plots related to this hologram are marked 'predicted'. When the predicted data is further corrected to improve the accuracy the plots corresponding to the corrected holograms are marked 'corrected'.

. P(k) ; $k \leq N$ is the true hologram that would be obtained if the actual aperture were to cover the expanded aperture. Plots related to this hologram are marked 'true'.

Let $$G(k) = g_1(k) + j\, g_2(k) \quad (4.31)$$

$$P(k) = p_1(k) + j\, p_2(k) \quad (4.32)$$

The percentage error in the magnitude of the predicted signal at point k is defined as:

$$E(k) \triangleq 100 \{[g_1(k) - p_1(k)]^2+[g_2(k) - p_2(k)]^2 \}^{\frac{1}{2}} / \{[p_1(k)]^2+[p_2(k)]^2\}^{\frac{1}{2}}$$

$$; \quad k = M+1, M+2, \ldots, N \quad (4.33)$$

A better definition for the prediction error is that which takes into account the effect of the error on the image reconstruction. This can be obtained by weighting the error defined in eqn. (4.33) by the ratio $|p(k)|/P_{max}$ where $|P(k)|$ is the modulus of the true signal at point k and P_{max} is the maximum of the modulus of the true hologram signal over the expanded aperture,

$$P_{max} \triangleq \max \{|P(k)|\} \quad ; k = 1,2,..,N \quad (4.34)$$

This leads to

$$E(k) = 100 \{ [g_1(k)-p_1(k)]^2+[g_2(k)-p_2(k)]^2\}^{\frac{1}{2}}/P_{max} \quad \ldots(4.35)$$

A single parameter which gives an overall indication of the prediction/correction errors over the aperture and their effect on the reconstructed image is the percentage mean error (M.E.) over the aperture. This is defined as:

$$M.E. = \{ [\underbrace{\sum_{k=1}^{N} E(k)}_{L.H.S.}] + [\underbrace{\sum_{k=1}^{N} E(k)}_{R.H.S.}] \}/2N \quad (4.36)$$

where E(k) is as defined in eqn. (4.35) and 2N is the total number of points in the expanded aperture. This parameter is useful in differentiating between situations giving rise to the same maximum prediction error $[E(k)]_{max}$ while the distribution of the errors over the aperture, and therefore their effect on the reconstructed images, is significantly different. The value of the mean error will be denoted M.E.1 for the predicted data and M.E.2 for the corrected data where applicable.

4.4 The Polynomial, Space-Variant, Predictive Model

The first hologram to be considered for aperture expansion is that of a single point object. In this case, a two-dimensional hologram is a single zone-plate corresponding to the point object. Fig. 4.2. shows the distributions of the real and imaginary parts of the hologram signal over the line hologram for a point object on the hologram axis at a distance of 100λ from a 64-point hologram. The sample spacing Δ is equal to 1λ. The figure indicates that variations in the hologram signal increase as the distance from the centre increases. This is due to the increase in the spatial rate of variations in the length of the vector joining the hologram point to the object point which determines both the phase and the amplitude of the signal. The signals at two neighbouring points on the hologram can therefore be related by a spatial function which is determined by their position relative to the centre.

4.4.1 The Power Series Polynomial Model

Referring to Fig. 4.1, assume a total available aperture containing 2M points and a total expanded aperture containing 2N points. Consider the R.H. side of the aperture, the signal at point (k) in the available aperture is expressed in terms of the signal at the preceding point (k-1) as:

$$H(k) = \psi(k) \; H(k-1) \tag{4.37}$$

where H(k) and H(k-1) are the complex hologram signals and ψ(k) is a complex function of the distance d_k from point k to the centre of the hologram, see Fig. 4.1. Polynomial expansion in d_k provides a simple and generalized form of representing the function ψ(k) over the aperture. Therefore, ψ(k) is assumed to have the following truncated expansion:

$$\psi(k) = a_1+a_2(d_k)^1 + a_3(d_k)^2 + \dots\dots + a_L(d_k)^{L-1}$$

$$= \sum_{i=1}^{L} a_i(d_k)^{i-1} \tag{4.38}$$

where a_i; $i=1,2,\ldots,L$ are complex coefficients and L is the number of terms in the polynomial.

Writing eqn. (4.37) for all but the first point in the R.H. half of the available aperture, i.e. for $k=2,3,\ldots,M$ and substituting for $\psi(k)$ from eqn. (4.38) produces the following set of M-1 linear equations in L unknowns which can be solved for the model coefficients $a_1,a_2,\ldots a_L$:

$$\begin{bmatrix} H(2) \\ H(3) \\ \cdot \\ H(M) \end{bmatrix} = \begin{bmatrix} H(1) & H(1)d_1 & H(1){d_1}^2 & \cdot & H(1)d_1^{L-1} \\ H(2) & H(2)d_2 & H(2){d_2}^2 & \cdot & H(1)d_2^{L-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ H(M-1) & H(M-1)d_{M-1} & H(M-1)d_{M-1}^2 & \cdot & H(M-1)d_{M-1}^{L-1} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ a_L \end{bmatrix} \qquad \ldots(4.39)$$

For an exactly determined system of equations, the number of terms in the polynomial, and therefore the number of model coefficients will be equal to the number of equations, i.e.:

$$L = M-1 \qquad (4.40)$$

To use the model for predicting new points, it is assumed that the relationship in eqn. (4.37) is also valid outside the available aperture, with $\psi(k)$ now defined by the model coefficients $a_1,a_2,\ldots,a_L$ obtained by solving the matrix equation (4.39). Therefore, the signal G(M+1) at the first point in the aperture extension can be predicted from the signal H(M) at the last point in the available aperture using:

$$\begin{aligned} G(M+1) &= \psi(M+1)\, H(M) \\ &= [\sum_{i=1}^{L} a_i\, (d_{M+1})^{i-1}]\, H(M) \end{aligned} \qquad (4.41)$$

Similarly, G(M+2) is obtained from G(M+1) , and so on until the last point in the expanded aperture G(N) is determined. In general $G(\ell)$ is obtained by:

$$G(\ell) = \left[\sum_{i=1}^{L} a_i \, d_\ell^{i-1} \right] G(\ell-1) \qquad (4.42)$$

The process of constructing the model and using it for prediction is repeated for the L.H. side of the aperture with the direction of prediction now from right to left; Fig. 4.1.

To demonstrate the effectiveness of this predictive model, it was used to expand a simulated hologram aperture of 16 points (M=8), spaced at 1λ to achieve a four-fold increase in its size (N=32). The object is a single point at the centre at a distance of 100λ from the hologram. The matrix equation (4.39), was solved using a utility program on the 4051 Tektronix minicomputer which employs the Gauss-Jordan elimination method. Appendix A contains a listing of an interactive computer program in Basic which simulates and expands the hologram of a point object and calculates the resulting prediction errors.

Fig. 4.3a shows the percentage prediction error, as defined by eqn. (4.35), over the aperture. The error increases slowly near the available aperture but since every predicted point is determined from the preceding point, the error accumulates as more new points are predicted. Therefore, the error increases rapidly towards the edges of the expanded aperture as shown in the figure.

Fig. 4.3b shows the intensities of images reconstructed from the small (available) aperture, H(k), the expanded predicted aperture G(k), and the extended true aperture , P(k) , cf. sec. 4.3. The image intensity is much lower in the case of the small aperture due to the relatively small number of hologram samples which contribute to the image formation compared to the case of the expanded apertures. Comparison between the image quality and resolution in the 3 cases can be better made when the images are plotted with their intensities normalized to the same peak value, these are shown in Fig. 4.3c. The images obtained from both the true and the predicted holograms are in good agreement in spite of the 13% prediction error in the hologram signal at the edge of the expanded aperture. Both images exhibit a reduction in the width of the point image compared to the image from the small aperture which indicates an improvement in resolution achieved by the four-fold increase in the aperture size.

Fig. 4.4 shows the prediction error and the normalized images for a point object which is offset from the centre. The prediction error is not symmetrical in both halves of the aperture due to the assymmetry in the hologram function over the aperture. The aperture expansion yields the expected improvement in resolution as in the case of the point at the centre.

4.4.2 The Effect of Object Range

The process of constructing the predictive model and using it for estimating the signal at new points was repeated with the same hologram parameters as in Fig. 4.3 for a point object at a shorter range, the results are shown in Fig. 4.5. Comparison of the two figures indicates that the prediction errors are greater in the case of the shorter range. The image reconstructed from the predicted aperture in the case of the short range is somewhat more spread and the sidelobe level is increased. At shorter ranges, the spatial frequency content of the hologram is greater and therefore there will be larger variations in the hologram signal from one point on the aperture to the next. The hologram function is therefore less smooth at shorter ranges which makes the accurate prediction more difficult with a given model order, L, and the prediction error increases.

In Fig. 4.6 the maximum prediction error over the expanded aperture is plotted versus the object range when doubling the size of a total available aperture of 8 points; M=4. The figure shows that very accurate prediction can be obtained at large ranges. However, at such large ranges greater expansion ratios will be required to achieve the amount of resolution improvement required from a given aperture. Since the errors increase rapidly as more new points are predicted, this will tend to limit the advantage of working at large ranges. The optimum operating range for the expansion algorithm will be a compromise between relatively larger initial prediction errors but smaller expansion ratios required at short range and smaller initial errors but larger expansion ratios needed at large ranges.

4.4.3 The Fourier Series Model

Several types of polynomial expansions have been used to represent the spatial function $\psi(k)$ in eqn. (4.37) which describe the predictive model. A number of orthogonal polynomials have been used including the Chebyshev, Legendre, and Hermite types. These three polynomials gave almost identical results to those obtained with the power series expansion in eqn. (4.38). However, a Fourier series expansion for the spatial function $\psi(k)$ has a number of interesting characteristics and can yield better prediction accuracy under optimum conditions.

For an exactly determined system, the spatial Fourier series expansion of $\psi(k)$ is:

$$\psi(k) = \sum_{i=1}^{L} a_i \exp j[(i-1)\omega_o d_k] \tag{4.43}$$

where a_i ; $i=1,2,\ldots,L$ are the model coefficients , d_k is the distance from point k to the centre of the aperture, ω_o is the fundamental radial spatial frequency of the Fourier series, and L is the number of terms in the model as defined by eqn. (4.40). The fundamental frequency ω_o is related to the total expanded aperture by:

$$\omega_o \triangleq \alpha \left(\frac{2\pi}{D_2} \right) \quad \text{radians/wavelength} \tag{4.44}$$

where D_2 is the total length of the expanded aperture in wavelengths and α is a scale factor.

When using the Fourier series model for hologram expansion, it was found that the prediction accuracy depends on the choice of the fundamental frequency ω_o. In the case of a 1-point object, and for a given hologram configuration, there exists an optimum value for ω_o which minimizes the prediction error. This optimum value can be considered as that at which the pattern of spatial variations represented by the Fourier series matches that of the hologram function of the point object.

Fig. 4.7a shows a plot of the error in the first predicted point, E(M+1), versus the scale factor α in eqn. (4.44) for a total available aperture of 8 points (M=4). The prediction error with the optimum value of the fundamental frequency is $6.6\times10^{-5}\%$ while for the same configuration the error is $2.7\times10^{-2}\%$ when the power series model is used. This shows that the Fourier series expansion for the space-variant predictive model provides greater prediction accuracy at optimum conditions. However, the optimum value of ω_o would be difficult to determine in practice where the true hologram is not known over the expanded aperture. This would require a search for the optimum image from the expanded predicted apertures which implies the knowledge of some a priori information about the object.

It is interesting to note that the optimum value of ω_o is not altered when the point object is laterally offset from the centre while keeping at the same range. This is demonstrated in Fig. 4.7b where the point object is shifted 4 wavelengths from the centre. This agrees with the concept of matching between the hologram pattern and the Fourier model function at optimum frequency. When the point object is shifted laterally, the hologram pattern stays basically the same and is merely shifted sideways with the object, see section 7.5.

The relationship between the object range and the spatial frequency over the hologram aperture can be obtained using eqn. (2.23) and Fig. 2.4 which assume the general case when a reference wave is used. When the complex hologram is measured directly, this is equivalent to using a normally incident plane reference wave, i.e. $\theta_2 = 0$ and $r_2 = \infty$. Moreover, for a point object on the hologram axis $\theta_1 = 0$ and $r_1 = z$. Eqn. (2.23) for the spatial frequency f_x at point x on the hologram reduces to:

$$f_x = \frac{x}{\lambda z} \qquad (4.45)$$

where z is the object range and λ is the wavelength. The radial spatial frequency ω_x is therefore:

$$\omega_x = K \frac{x}{z} \qquad (4.46)$$

where K is the wavenumber. Eqn. (4.46) shows that for a given point on the hologram the spatial frequency is inversely proportional to the range.

According to the concept of matching between the Fourier expansion and the hologram function, the optimum value of ω_0 should therefore vary in a similar fashion with range to that of eqn. (4.46) for a point object. Fig. 4.8a shows the optimum value of ω_0 which minimizes the error in the first predicted point at every value of the range versus the object range, z. Shown also in the same figure is a plot of the function 1/z which shares the same value at the first point in the curve with the ω_0 plot. The figure shows that the two curves are in good agreement which indicates that the optimum frequency is inversely proportional to the range as expected.

Fig. 4.8b shows the minimum error in the first point versus range. The slight irregularity in the curve may be parially due to the limited accuracy in determining the optimum frequency at each value of the range.

4.5 Error Analysis

As illustrated in section 4.4, prediction errors result when the model fitted to the available hologram data is used to predict the signals outside the given aperture. These errors arise because the model does not fit the hologram function exactly due to the limited number of terms in the model and to the discrepancy between the model function and the actual hologram function. In this section, the fitting error in the case of the power series polynomial model describdd in section 4.4.1 is considered.

Referring to the imaging configuration in Fig. 4.1, consider the case of a 1-point object at (x_o,z) in the Fresnel region of the hologram. The hologram signals at points k-1 and k within the available aperture are:

$$H(k-1) = \frac{1}{z} \exp j\, Kz[1+ \frac{1}{2} (\frac{d_{k-1}-x_o}{z})^2]$$

$$H(k) = \frac{1}{z} \exp j\, Kz[1+ \frac{1}{2} (\frac{d_k-x_o}{z})^2] \tag{4.47}$$

where K is the wavenumber, $K = 2\pi/\lambda$ and d_{k-1} and d_k are the distances between each of the two points and the centre of the hologram. The power series model described by eqns. (4.37) and (4.38) can be formulated such that the modelled function is the ratio $h_o(k)$ between H(k) and H(k-1),

$$h_o(k) = \frac{H(k)}{H(k-1)} \tag{4.48}$$

Substituting for H(k) and H(k-1) from eqn. (4.47) gives:

$$h_o(k) = \exp j \frac{K}{2z} [(d_k-x_o)^2-(d_{k-1}-x_o)^2] \tag{4.49}$$

This leads to:

$$h_o(k) = \exp j \frac{K}{2z} [(d_k-d_{k-1})(d_k+d_{k-1}-2x_o)] \tag{4.50}$$

Eqn. (4.50) indicates that $h_o(k)$ is a function of the location of both points k and k-1 on the hologram, the location of the object point in the object plane, and the range z. In the polynomial model, eqn. (4.38), $h_o(k)$ is assumed to be a function of d_k only and therefore the model does not represent $h_o(k)$ exactly even if a power series expansion of infinite length is used. It is expected that the polynomial model would represent a better fit to the hologram function, and therefore produce more accurate prediction, if d_{k-1}, x_o and z are included in the expansion. This, however, would require the knowledge of the object location which might restrict the application of the technique in situations where such a priori information is not available.

Replacing (d_k-d_{k-1}) by Δ , the sample spacing, and using the approximation:

$$d_k+d_{k-1} \approx 2d_k \quad ; \quad k >> 1 \tag{4.51}$$

Eqn. (4.50) can then be written approximately as:

$$h_o(k) \approx \exp[j \frac{K\Delta}{z} (d_k-x_o)] \tag{4.52}$$

Therefore, for a given object range and hologram configuration, the hologram function $h_o(k)$ can be approximated to a complex exponential function of the distance (d_k-x_o). This explains the improvement in the prediction accuracy when the polynomial model is replaced by an optimum Fourier series model since the latter is based on expansion in exponential functions and therefore fits $h_o(k)$ more accurately.

Although a complex exponential function can be represented by a power series, the accuracy of this representation depends on the number of terms in the expansion, which should ideally extend to infinity. Using a polynomial model of order L, the ratio function $h_o(k)$ is approximated as:

$$h_o(k) \triangleq f(d_k) \triangleq a_1+a_2 d_k + a_3 d_k^2 + \ldots + a_L d_k^{L-1} \tag{4.53}$$

Expanding $f(d_k)$ in a Taylor-series about the origin $(d_k=0)$ yields:

$$f(d_k) = f(o) + f'(o) d_k + \frac{1}{2!} f''(o) d_k^2 + \ldots$$
$$+ \frac{1}{(L-1)!} f^{(L-1)}(o) (d_k)^{L-1} + R_L (d_k) \tag{4.54}$$

where $R_L(d_k)$ is the error due to the truncation of the Taylor-series to the first L terms only. The upper bound on this error is obtained using the Lagrange's remainder formula [136] as:

$$|R_L|_{max} = \frac{1}{L!} | f^{(L)}(d) |_{max} (d_k)^L$$

$$; 0 < d < d_k \tag{4.55}$$

Comparison between eqns. (4.53) and (4.54) shows that fitting a polynomial model is equivalent to expanding the hologram ratio function in a Taylor-series expansion about the hologram origin where the model coefficients are given by:

$$a_i = \frac{1}{(i-1)!} f^{(i-1)}(o) \quad ; \quad i=1,2,\ldots,L \tag{4.56}$$

Therefore, the upper bound on the modulus of the fitting error when using a polynomial model of order L is given by $|R_L|_{max}$ in eqn. (4.55). From eqn. (4.52), the L th derivative of the hologram ratio function, $h_o(k)$ at distance d is given by:

$$f^{(L)}(d) = \left(\frac{jK\Delta}{z} \right)^L \exp\left[\frac{jK\Delta}{z} (d-x_o)\right] \tag{4.57}$$

Since the modulus of $f^{(L)}(d)$ does not depend on the distance d, it can be substituted for $|f^{(L)}(d)|_{max}$ in eqn. (4.55). This yields:

$$|R_L|_{max} = \frac{1}{L!} \left(\frac{K\Delta d_k}{z}\right)^L \tag{4.58}$$

Substituting $K=2\pi/\lambda$ gives:

$$|R_L|_{max} = \frac{1}{L!} \left(\frac{2\pi\Delta d_k}{z} \right)^L \tag{4.59}$$

where the values of Δ, d_k, and z are in wavelengths.

Subject to the approximations made in deriving eqn. (4.52), eqn. (4.59) gives an estimate of the maximum fitting error due to the use of a truncated polynomial expansion. It does not, however, account for the errors arising from representing the modelled signal as a function of d_k only and ignoring the rest of the parameters in eqn. (4.52).

Therefore, $|R_L|_{max}$ can be used only as a guide which serves to illustrate the way in which the error is influenced by the various parameters of the imaging configuration such as the model order, L, the range z, the sample spacing Δ, and the location of the sample on the hologram aperture as represented by d_k.

Eqn. (4.59) shows that for all values of L, the polynomial truncation error increases with the increase in the sample spacing Δ and, as expected from the results obtained in section 4.4, increases with the increase in d_k and the decrease in the range z. Since Δ must be small enough to prevent aliasing effects in the sampled hologram and $d_k \ll z$ for most practical applications, the ratio ($\frac{2\pi\Delta d_k}{z}$) is normally less than one and the error decreases with the increase in L. The truncation error is plotted in dbs relative to unity versus the distance d_k in Fig. 4.9 for various values of L,Δ , and z. Figs. 4.10 and 4.11 show similar plots versus the object range z and the model order L respectively. The figures indicate the greater influence of variations in the model order on the error compared to the other parameters.

4.6 The Linear, Space-Invariant, Predictive Model

When the polynomial, space-variant, model, eqn. (4.38), was used to expand the hologram of a 2-point object the prediction errors were large and although the two points can be resolved using the predicted hologram, the image contains spurious details of large magnitude as shown in Fig. 4.12. This indicates that the space-variant model does not adequately represent the hologram signal for the case of a multiple-point object. In fact, this type of model is best suited for describing a single pattern of spatial variations centred at the projection of the 1-point object on the hologram plane. For an object consisting of a number of points, the hologram contains the same number of such patterns which interfere coherently at every point on the hologram aperture.

One other possibility for representing the signal at any point in the available aperture is to express it as a linear combination of the signals at all the preceding points in the aperture half under consideration. Therefore, the signal at point k can be written as:

$$H(k) = \sum_{j=1}^{L} c_j \, H(j)$$

with

$$H(j) = 0 \text{ for } j \geqslant k \tag{4.60}$$

where c_j; $j=1,2,\ldots,L$ are the model coefficients. When eqn. (4.60) is written for the last (M-1) points on the aperture half; i.e. $k=2,3,\ldots,M$, the following complex matrix is obtained for an exactly determined system in which the model size L is equal to M-1:

$$\begin{bmatrix} H(2) \\ H(3) \\ \cdot \\ H(M) \end{bmatrix} = \begin{bmatrix} H(1) & 0 & 0 & \cdot & 0 \\ H(1) & H(2) & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ H(1) & H(2) & H(3) & \cdot & H(M-1) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \cdot \\ c_L \end{bmatrix} \qquad \ldots(4.61)$$

It is noted in this case that the square matrix of dimension (LxL) which represents the system of equations takes a triangular form where the upper triangle is filled with zeros. This is due to the windowing effect in eqn. (4.60) since only the data corresponding to one half of the available aperture is used to construct the predictive model for use on that side of the aperture. This matrix form has certain advantages in increasing the stability of the prediction accuracy against variations in the object parameters as discussed in section 4.10 and in reducing the effects of disturbing noise, see sections 5.3 and 5.7.

Once the model coefficients $c_1, c_2, .., c_L$ are obtained by solving the above matrix equation, the signal signal at the points in the aperture extension can be obtained as follows:

$$G(M+1) = \sum_{j=1}^{L} c_j \; G(j+1)$$

$$G(M+2) \quad \sum_{j=1}^{L} c_j \; G(j+2)$$

$$\cdot \qquad \cdot \quad \cdot \quad \cdot$$

$$G(M+\ell) = \sum_{j=1}^{L} c_j \; G(j+\ell)$$

$$\cdot \qquad \cdot \quad \cdot \quad \cdot$$

$$G(N) = \sum_{j=1}^{L} c_j \; G(j+N-M) \tag{4.62}$$

The linear predictive model was used to increase the size of hologram aperture four times for both a 1-point and a 2-point object. The errors for the case of the 1-point object are shown in Fig. 4.13a for both the linear and the polynomial models. It is obvious from the figure that the polynomial model is superior to the linear model for the case of the 1-point object for small expansion ratios. The results for the 2-point object are shown in Fig. 4.13b from which it is clear that the linear model gives better overall results. It is generally noted that the polynomial model gives better performance when predicting the first few points near the available aperture while the linear model gives greater prediction accuracy towards the edge of the expanded aperture.

4.7 The Hybrid, Polynomial / Linear, Predictive Model

The results in the previous section have indicated the complementary nature of the polynomial (space-variant) and the linear (space-invariant) prediction models. The polynomial model is best suited for 1-point objects and, in general, gives better accuracy in the vicinity of the available aperture. On the other hand, the linear model is more suited for 2-point objects and gives improved performance away from the available aperture. This suggests that a combination of both types of models should be advantageous.

In the hybrid, polynomial/linear predictive model, the signal at any point on the available aperture is related to the signal at the immediately preceding point through a space variant function, as in the polynomial model, and is also linearly related to the signals at a number of the preceding points, as in the linear model. Assuming a power series expansion for the space variant function, the hologram function over the available aperture is written as:

$$H(k) = \sum_{i=1}^{p} a_i (d_k)^{i-1} + \begin{cases} \sum_{j=1}^{q} c_j H(j) & k=2,\ldots,q+1 \\ ;\ H(j)=0 \text{ for } j \geqslant k & \\ \sum_{j=1}^{q} c_j H(k-q+j-1) & k=q+2,\ldots,M \end{cases} \quad (4.63)$$

where p is the number of polynomial (space-variant) terms and q is the number of linear (space-invariant) terms in the hybrid model. $a_i (i=1,\ldots,p)$ are the polynomial coefficients and $c_j (j=1,\ldots,q)$ the linear coefficients. The composition of the hybrid model will be represented throughout in the format p/q indicating the number of polynomial/linear terms respectively.

Since both the polynomial and the linear parts of the hybrid model must share the number of degrees of freedom in the available aperture, and assuming an exactly determined system of equations, then

$$p+q = L = M-1 \quad (4.64)$$

Writing eqn. (4.63) for all values of k over the available aperture produces L linear equations in the L unknown model coefficients. For a power series polynomial, these equations take the following matrix form:

$$\begin{bmatrix} H(2) \\ H(3) \\ \cdot \\ H(q+1) \\ H(q+2) \\ \cdot \\ H(M) \end{bmatrix} = \begin{bmatrix} H(1) & H(1)d_1 & \cdot & H(1)d_1^{p-1} & H(1) & 0 & \cdot & 0 \\ H(2) & H(2)d_2 & \cdot & H(2)d_2^{p-1} & H(1) & H(2) & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ H(q) & H(q)d_q & \cdot & H(q)d_q^{p-1} & H(1) & H(2) & \cdot & H(q) \\ H(q+1) & H(q+1)d_q & \cdot & H(q+1)d_{q+1}^{p-1} & H(2) & H(3) & \cdot & H(q+1) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ H(M-1) & H(M-1)d_{M-1} & \cdot & H(M-1)d_{M-1}^{p-1} & H(M-q) & H(M-q+1) & \cdot & H(M-1) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ a_p \\ c_1 \\ \cdot \\ c_q \end{bmatrix}$$

...(4.65)

The model coefficients $a_1, a_2, \ldots, a_p$, $c_1, c_2, \ldots, c_q$ are obtained by solving eqn. (4.65) . The model is then used to predict hologram points outside the given aperture. G(M+1) is calculated from G(M) together with the other q-1 preceding points. Similarly G(M+2) is obtained from G(M+1) and the other q-1 preceding points, and so on. In general, the signal at point (M+ℓ) is given by:

$$G(M+\ell) = [\sum_{i=1}^{p} a_i (d_{M+\ell})^{i-1}]G(M+\ell-1) + \sum_{j=1}^{q} c_j G(M+\ell-q+j-1)$$

...(4.66)

By choosing the value of p and q, subject to the condition in eqn. (4.63), the structure of the hybrid model can be selected to suit the particular application at hand. For p=L, q=0 all the terms correspond to a polynomial expansion which represents the optimum model for 1-point objects as described in section 4.4. For p=0, q=L the hybrid model is reduced to the linear model discussed in section 4.6.

For a total available aperture of 16 points, i.e. M=8 and the model order L=7, the optimum model for the case of a 2-point object was found to contain 5 polynomial terms and 2 linear terms. This hybrid model was used to double the size of a 16 point aperture spaced at 1λ to image a 2-point object symmetrically positioned in the object plane and separated by a distance of 6λ at a range of 105λ from the hologram. The prediction errors are shown in Fig. 4.14a. The images reconstructed from the small, predicted, and corrected apertures are shown in Fig. 4.14b.

Referring to Fig. 4.14b, the small available aperture is not large enough to resolve the two points since the Rayleigh resolution limit predicts a resolution distance of 7λ for the small aperture. Moreover, the actual resolution limit would be worse for the case of coherent radiation since the two simulated points are assumed to radiate in phase [33]. The two points are clearly resolved with both the true and predicted expanded holograms. The predicted hologram gives a reconstructed image which is in good agreement with that obtained from the true hologram. This shows that the image reconstruction process is tolerant to prediction errors as high as 20% at the edge of the expanded aperture. It should be noted, however, that the reconstruction results are also affected by the distribution of the errors over the hologram aperture.

To demonstrate the advantage of this optimum hybrid model over the polynomial and the linear models, the later two models were used to expand the same hologram aperture in the above example. The results are shown in Fig. 4.15a for the polynomial model; p=7 and q=0 and in Fig. 4.15b for the linear model; p=0, q=7 . In the case of the polynomial model, Fig. 4.15a, the images are shown normalized to the same peak since the intensity of the image obtained from the predicted aperture far exceeds those of the other two images due to the large prediction errors. Although the image from the predicted aperture shows two intensity peaks at the correct positions corresponding to the two points in the object, it suffers from poor contrast and is cluttered with additional spurious peaks. Fig. 4.15b reveals that the linear model, though inferior to the optimum hybrid model, gives much better results than the polynomial model for the case of the 2-point object considered. The two points can be clearly resolved with this model and the central dip in the image exceeds the 19% value which corresponds to the Rayleigh resolution criterion. However, the image is obviously worse than that obtained with the optimum hybrid model.

Fig. 4.16 shows the prediction errors and the reconstructed images for a 2-point object in which the amplitude of the signal from the R.H. point is 0.8 of that from the other point. The optimum 5/2 hybrid model was used for the prediction. Although the relative signal strength is not recovered exactly even in the case of the true aperture due to the limited aperture size, the image from the predicted aperture is in good agreement with that from the true aperture.

The optimum hybrid model was used to expand a given 16-point aperture four times to image an object consisting of two points separated by a distance of $3{\cdot}4\lambda$ at a range of 80λ. Fig. 4.17a shows the reconstructed images from the small, predicted, and true apertures. From this figure, the two points can be clearly resolved in the image from the predicted aperture.

Throughout the above examples using the polynomial or the hybrid models, the polynomial part in the model is based on the power series expansion in eqn. (4.38) . In the hybrid model, it is also possible to use other types of expansions for the polynomial part. When a Fourier series expansion, eqn. (4.43), is used for the 5 polynomial terms in the optimum hybrid model for the case 2-point objects; it is possible to find an optimum value for the fundamental spatial frequency of the Fourier series which minimizes the prediction errors. Comparison between Figs. 4.17a and 4.17b indicates the advantages of the Fourier polynomial over the power series polynomial in reducing the prediction errors and therefore improving the reconstructed images for the case of 2-point objects as has been demonstrated in section 4.4.3 for the case of 1-point objects.

4.8 The Corrective Model

Throughout the prediction procedures described so far in this chapter, only one half of the available aperture is used at a time to predict the hologram signal at the aperture extension adjacent to it. For example, referring to Fig.4.18, the hologram signal H_1 at the R.H. half of the available aperture is used to predict the signal G_1' at the R.H. aperture extension.

The possibility of improving the prediction accuracy by subsequently correcting the predicted values of the signal using a predictive model has been investigated. Such a model is derived by relating the true signal on one half of the available aperture to the data predicted at the same half. This is achieved by making use of the fact that the hologram signals are known at both halves of the available aperture. Therefore, for each half of the available aperture, both the true data and the data predicted using a corrective model which is based on the other half are available. The corrective model for use with predictions in the R.H. aperture extension is constructed as follows; see Figs. 4.18 and 4.19a:

(i) Construct a predictive model, model 1, based on the L.H. half of the available aperture H_2.

(ii) Predict the hologram signals at the R.H. half of the available aperture using the model constructed in (i). Denote the predicted signals H_1'.

(iii) Construct a corrective model based on relating the known signals H_1 to the predicted signals H_1' over the R.H. half of the available aperture.

Having constructed the corrective model, it can now be used to correct predicted data in the R.H. aperture extension, G_1', to obtain more accurate results, G_1, as follows; see Figs. 4.18 and 4.19b:

(i) Construct a predictive model, model 2, based on the R.H. half of the available aperture, H_1.

(ii) Predict the hologram signals at the R.H. aperture extension. Denote the predicted signals G_1'.

(iii) Apply the corrective model to correct the predicted signals G_1' to obtain more accurate data G_1.

The same process is repeated for correcting the L.H. half of the predicted aperture.

4.8.1 The Polynomial, Space-Variant, Corrective Model

Let the true hologram signal at point k in the R.H. half of the available aperture be $H_1(k)$ and the predicted signal at the same point be $H_1'(k)$. The polynomial, space-variant corrective model of the maximum size is constructed by expressing the ratio $H_1(k)/H_1'(k)$ as a power series in the distance from the point to the aperture centre, for all points on the R.H. half of the available aperture,

$$H_1(k)/H_1'(k) = \sum_{i=1}^{M} u_i \, (d_k)^{i-1} \quad ; \ k=1,2,..,M \qquad (4.67)$$

where M is the number of points in one half of the available aperture, $u_i (i=1,2,...,M)$ are complex coefficients and d_k is the distance from the centre.

The set of equations represented by eqn. (4.67) can be solved for the corrective model coefficients; $u_1, u_2, ..., u_M$. If $G_1'(k)$ is the predicted signal at point k in the R.H. aperture extension; k=M+1, M+2,...,N, then the corrected signal $G_1(k)$ at the same point can be obtained as:

$$G_1(k) \ = \ G_1'(k) \sum_{i=1}^{M} u_i (d_k')^{i-1} \quad ; \ k=M+1,M+2,..,N \qquad (4.68)$$

where $u_1, u_2, ..., u_M$ are the corrective model coefficients obtained from eqn. (4.67) and d_k' is the distance from a reference point located midway between the Mth and the (M+1)th sample points, Fig. 4.18.

The polynomial corrective model was used to correct hologram data obtained using a polynomial predictive model for the case of a 1-point object. Fig. 4.20 shows the improvement in prediction accuracy over the aperture extension obtained over the aperture extension with a corrective model having the maximum size M. A reduction of approximately 55% in the prediction error at the edge of the expanded aperture is achieved through correction with this model. At a shorter object range, Fig. 4.20b, the improvement in the prediction accuracy is reduced. This is expected, since the corrective model is based on the predictive model. At short ranges the effectiveness of the predictive model is reduced due to the higher frequency of the spatial variations over the hologram. This in turn reduces the effectiveness of the corrective model.

The effect of varying the size of the corrective polynomial model is shown in Fig. 4.21 where the maximum error over the aperture after correction is plotted against the size of the model for the case of data predicted with a polynomial predictive model and the hologram of a 1-point object. For corrective models of less than the maximum size,M, eqn. (4.67) is written only for a limited number of points L' on the available aperture where L'<M. The isolated point to the left of the curve in Fig. 4.21 represents the maximum error over the aperture when no correction is applied. The figure shows that in this case the correction accuracy is generally improved with the increase in the size of the corrective model.

4.8.2 The Hybrid,Polynomial / Linear,Corrective Model

When the polynomial corrective model was used to correct hologram data obtained using a hybrid polynomial/linear predictive model, there was no improvement in the overall prediction accuracy and the supposedly corrective process actually produced worse results than the original predicted data. In its general form, the corrective model could also take a hybrid form in which the ratio between the true and predicted signals, $H_1(k)$ and $H_1'(k)$ respectively, at one half of the available aperture is represented as a spatial function plus a linear combination of the same ratio at a number of preceding points.

$$\frac{H_1(k)}{H_1'(k)} = \sum_{i=1}^{r} u_i (d_k)^{i-1} + \begin{cases} \sum_{j=1}^{s} v_j \frac{H_1(j)}{H_1'(j)} & k=1,\ldots,s+1 \\ ; \frac{H_1(j)}{H_1'(j)} = 1 \text{ for } j \geqslant k & \\ \sum_{j=1}^{s} v_j \frac{H_1(k-s+j-1)}{H_1'(k-s+j-1)} & k=s+2,\ldots,M \end{cases} \tag{4.69}$$

where r is the number of polynomial terms, s is the number of linear terms, $u_1,u_2,\ldots,u_r$ are the polynomial coefficients, $v_1,v_2,\ldots,v_s$ are the linear coefficients, d_k is the distance from the centre. For an exactly determined system of equations:

$$r + s = M \tag{4.70}$$

With the corrective model coefficients $u_1,u_2,\ldots,u_r$, $v_1,v_2,\ldots,v_s$ determined by solving eqn. (4.69), the predicted signal $G_1'(k)$ at the aperture extension can be corrected to give:

$$G_1(k) = G_1'(k)\left[\sum_{i=1}^{r} u_i (d_k')^{i-1} + \sum_{j=1}^{s} v_j \frac{G_1(k-s+j-1)}{G_1'(k-s+j-1)}\right]$$

$$; k=M+1,M+2,\ldots,N$$

with

$$\frac{G_1(k-s+j-1)}{G_1'(k-s+j-1)} = 1 \text{ for } j \leqslant s-k+1 \tag{4.71}$$

where d_k' is as defined for eqn. (4.68), see Fig. 4.18.

The hybrid corrective model can lead to improvement in the prediction accuracy when used with either the polynomial or the hybrid predictive model. The performance of the corrective model, however, depends upon the composition of both the predictive and the corrective models. To investigate the optimum composition of the corrective model, the size of the model was kept fixed at the maximum possible value, i.e. r+s = M, and the number of linear terms,s , increased from 0 to M. The maximum error over the expanded aperture after correction was calculated in each case.

Fig. 4.22a shows the results when the model is used with a polynomial predictive model (p=L,q=0) for the case of a 1-point object. The isolated point corresponds to the maximum prediction error over the aperture without any correction. The figure shows that the optimum corrective model for use with the polynomial predictive model is a linear model (r=0,s=M) which reduces the prediction error to an almost negligible quantity. Fig . 4.22b is the corresponding plot for the case of a 5/2 hybrid model and a 2-point object. It shows that it is possible for certain compositions of the corrective model to increase the errors in the predicted signals rather than reduce them and a search for the optimum corrective model would therefore be required. The optimum model for the case considered is a 4/4 polynomial/linear which reduces the error by a factor of approximately 85%. Fig. 4.22c for the linear predictive model and a 2-point object shows that the prospects for a major improvement in the prediction accuracy through correction are much less in this case.

The plots in Fig. 4.23 demonstrate the resolution improvement that can be achieved using the corrective technique on the predicted data. A small aperture containing 8 points (M=4) is expanded four times in order to image a 2-point object; optimum models were used for both prediction and correction. Fig. 4.23a shows the reduction in the prediction error and Fig. 4.23b the reconstructed images from the small, predicted, corrected, and true apertures. Because of its limited size, the small aperture cannot resolve the two points in the object. Due to the large prediction errors, the expanded predicted aperture cannot resolve them either and it is only after the correction process that the two points are clearly resolved.

4.9 The Effect of Aperture Expansion on Range Resolution

In the previous sections the effect of expanding the hologram aperture on the lateral resolution has been considered. When the range resolution is aperture-limited, its value is given, according to one criterion, by eqn. (2.71). This equation defines the range resolution as the distance in range between the plane of focus, at which the image intensity is at its peak, and the plane at which the intensity drops by a factor of 20% of its peak value.

To verify that range resolution is improved when the aperture is expanded, the hologram of a 1-point object at the centre at a distance of 50λ from a 15λ aperture (spacing = 1λ) was simulated. The size of this aperture was doubled using the optimum predictive model, a polynomial model. The images corresponding to the small, predicted, and true apertures were reconstructed at a number of reconstruction distances from 5λ to 85λ from the hologram in steps of 5λ. For every image reconstruction, the value of the image intensity at the centre was recorded. In this way, an image in-range of the point object is created for each of the three apertures. Fig. 4.24 shows these reconstructed images versus range, normalized to the same peak value.

This figure indicates that both the predicted and the true holograms produce a peak at a distance of 50λ which corresponds to the correct object range. The image obtained from the small aperture is much wider than those corresponding to the holograms of the expanded aperture and its peak occurs at a slightly shorter range. The two images from the predicted and true holograms are in good agreement and both indicate an average resolution distance of 5.7λ while eqn. (2.71) predicts a distance of 5.2λ for the imaging geometry considered.

The improvement in range resolution through aperture expansion allows the correct imaging of two objects close in range which would not be possible using the small available aperture. The hologram of 2 points located on the hologram axis at ranges 20λ and 50λ was simulated and the resulting aperture expanded by a factor of 2 using an optimum 4/4 hybrid predictive model. The images in-range which were obtained are shown in Fig. 4.25 for the small, predicted, and true apertures. Due to its shorter range, the point nearer to the hologram is imaged at the correct range with all three apertures, although the definition is better in the case of the two expanded apertures. The small aperture fails to resolve the farer point at the correct range while both the predicted and the true apertures image the point with an error of ±10% in its location in range. Part of this error may be due to the large steps in range (5λ) at which the images were reconstructed.

4.10 The Effect of Increased Object Complexity

It has been shown in section 5.7 that for more than one point in the object to be imaged, it is necessary to introduce a number of linear terms in the predictive model in order to achieve adequate prediction accuracies. These linear terms relate the signal at every point in the available aperture to the signal at the points in its near vicinity. A 5/2 polynomial/linear hybrid model gives optimum results for the 2-point objects considered. Although the linear model (0/7) gives less accurate prediction, the prediction accuracy can be adequate in many cases as demonstrated in Fig. 4.15b. The linear model has the advantage of providing greater stability with noisy data which will be discussed in Chapter 5. In this section the effects of increasing the complexity of the multi-point object on the performance of both models are considered. This includes increasing the spacing between the points in a 2-point object and the simulation of quasi-continuous objects which are represented by a large number of points within a given finite object width.

4.10.1 Effect of Increasing the Point Spacing in a 2-Point Object

Fig. 4.26 shows the maximum error over the predicted aperture for an expansion ratio of 2 versus the spacing between the points in a 2-point object symmetrically positioned relative to the hologram axis for the case of the 5/2 hybrid model and the 0/7 linear model. The prediction errors tend to increase with increasing the spacing between the two points due to the increased spatial frequency content of the hologram. This trend is manifested more clearly in the case of the hybrid model than in the case of the linear model. The figure also shows that while the hybrid model gives better performance for small spacings compared with the linear model, the latter gives more accurate prediction for large spacings.

4.10.2 Effect of Increasing the Number of Points Representing an Object of a Given Width

As shown in Fig. 4.26, the extent of the object in the direction parallel to the hologram plane has a marked effect on the prediction errors for both predictive models under consideration. In order to observe the effects of increasing the point density on the object, the object width must be fixed. The objective of this investigation is to determine the effectiveness of the prediction technique for common continuous objects which represent the limit for the discrete point objects when the number of points taken to represent the object tends to infinity. The object of a given width is represented by a number of uniformly spaced points along its width in the direction parallel to the hologram line and the density of these points can be increased to approach the case of a continuous object.

The maximum prediction error encountered when doubling the size of a 16-point aperture using the 5/2 hybrid model is plotted versus the number of points taken on the object in Fig. 4.27a for 3 values of the object width, at the same range. The corresponding results for the linear model are shown in Fig. 4.27b. In both cases, as the number of points on the object increases, for the same object width, the error generally decreases towards a constant steady-state value. This indicates that the prediction error stabilizes as the case of a continuous object is approached which implies that this prediction technique would be useful for improving the resolution when imaging ordinary spatially bounded objects. It is also noted that the steady-state value of the error is considerably lower than the error in the case of a 2-point object where the two points are situated at the extreme ends of the object length. Therefore, the expansion of the hologram of a continuous object with a finite width would be even more accurate than when the object is represented by two points marking its ends.

A comparison between Figs. 4.27a and b shows that the steady-state value is reached more rapidly in the case of the linear model which indicates that this model is relatively insensitive to changes in the point density. The steady-state error increases with the increase in the object length for the hybrid model in Fig. 4.27a while it decreases in the case of the linear model in Fig. 4.27b. The variations in the steady-state error with the object width is only about 15 dbs in the case of the linear model compared to approximately 35 dbs for the hybrid model. This shows that the linear model is also relatively insensitive to variations in the object width.

Fig. 4.28 shows the same effect of increasing the point density but for the same object width at 3 values of the object range. While the steady-state error in the case of the hybrid model varies by approximately 50 dbs over the range values used, Fig. 4.28a, the corresponding variations in the case of the linear model does not exceed 5 dbs; Fig. 4.28b. Therefore, the linear model is obviously less sensitive to variations in the distance of a quasi-continuous finite object from the hologram.

The general tendency for the prediction error to decrease as the point density on the object increases and the fact that it approaches a steady-state value can be explained with reference to Fig. 4.29. For simplicity, assume that the object range is large enough for the Fraunhofer approximations to apply. In this case, as discussed in section 3.4, the hologram is approximately the Fourier transform of the object. Due to its limited aperture size, the hologram acts as a low-pass filter with a cut-off frequency f_c which is proportional to the numerical aperture of the hologram.

In Fig. 4.29 the spectrum of the discrete finite object is derived in steps, starting with a continuous object of a finite width which has a sinc(x) spectrum as shown in Fig. 4.29a. Representing such an object using a number of uniformly spaced discrete points is equivalent to sampling the continuous object with a train of Delta functions spaced at distance τ in the object (space) domain. This train of Delta functions has the spectrum shown in Fig. 4.29b which also consists of another train of Delta functions spaced at $(1/\tau)$ in the frequency (hologram) domain. The multiplication involved in the sampling process in the space domain is equivalent to covolution in the frequency domain and therefore the spectrum of the discrete finite object consists of an infinite number of sinc(x) spectra with their centres spaced at a spatial frequency interval equal to $(1/\tau)$, Fig. 4.29c.

Superimposed on the discrete object spectrum in Fig. 4.29c is the transfer function of the hologram with a cut-off frequency f_c. As the number of points used to represent an object of a given width increases, the spacing, τ, between the points decreases and therefore the spacing between the sinc(x) spectra increases. This reduces the number of these spectra within the passband of the hologram aperture. This reduction in the spatial frequency content of the hologram reduces the prediction accuracy as noted in section 4.4.1 for the case of 1-point objects. When the density of the discrete points on the object is increased further, a stage will be reached when all the sinc(x) spectra will be removed outside the passband with the exception of the central spectrum which corresponds to the case of the continuous object. When this stage is reached, any further increase in the point density does not lead to any change in the frequency content of the hologram and therefore the prediction accuracy reaches its constant steady-state value which would be obtained if the discrete object were a continuous one.

In practice, the simplified argument presented above is further complicated by the fact that in the Fresnel region the hologram is the spectrum of the product of the object function with a quadratic phase function, cf. eq. (2.48). Moreover, a more rigorous analysis should also take into account the nature of the relationship between the prediction accuracy and the spatial frequencies on the hologram.

It should be noted that the Fraunhofer integral which describes the hologram in the far field can also be expressed in terms of the angular spectrum of the object distribution. In this case the hologram signal at a point which subtends an angle θ at the object plane is given by [128]

$$H(\sin\theta) = \int_{-X_o}^{X_o} a\left(\frac{x_o}{\lambda}\right) \exp\left\{-j2\pi\left(\frac{x_o}{\lambda}\right)\sin\theta\right\} d\left(\frac{x_o}{\lambda}\right)$$

where $\pm X_o$ represent the object extent and $a\left(\frac{x_o}{\lambda}\right)$ is the object distribution. Comparison with the standard form of the Fourier transform shows that the frequency is now represented by $\sin(\theta)$. This form also describes the radiation pattern of the object. Fig. 4.29d shows the angular spectrum of a 2-point object where the points are separated by a distance of $\lambda/2$. In a practical situation the data can be collected only over a limited portion of the angular space. Fig. 4.29e illustrates the case when θ is restricted to $\pm \pi/2$ and the two points cannot be resolved.

The relative insensitivity of the linear model to the point density, object width, and object range can be explained by referring to the matrix equation which describes the model; eqn. (4.61). Because of the triangular form of the matrix, the model coefficients are less influenced by the hologram signals at the various points on the aperture compared to the case of the square matrix. This obviously has the disadvantage of reducing the prediction accuracy since the model does not represent the signal over the hologram aperture as accurately. On the other hand, this makes the results obtained with the triangular matrix model more stable against factors affecting the hologram signals such as the object width, object range, and the number of points used to represent a given object as demonstrated in Figs. 4.26 - 4.28. Such factors also include noise and measurement errors in realistic hologram signals as will be discussed in Chapter 5.

4.10.3 Imaging of Quasi-Continuous Objects

The results described above in section 4.10.2 are encouraging since they indicate that prediction errors approach a stable value as the case of the continuous object is approached. In this section the results obtained when imaging two quasi-continuous segments instead of 2-point objects are discussed. Each segment is of a finite width w and is represented by a number of uniformly spaced points as described in section 4.10.2. The point density is defined as ρ points/wavelength. The two segments are positioned in the object plane symmetrically about the hologram axis and separated by a distance of 5.6λ, centre to centre. The images are obtained from holograms predicted using a 5/2 hybrid model and a linear model. To provide some form of a reference for comparison, Figs. 4.30 and 4.31 show the prediction errors and the reconstructed images for two points located 5.6λ apart using the hybrid and the linear models, respectively, for prediction.

Figs. 4.32 - 4.35 show the corresponding results for a segment width of 4λ. In this case the gap between the two segments is 1.6λ and is clearly resolved by both the true and predicted holograms. In Figs. 4.32 and 4.33 the point density is 25 points/λ while in Figs. 4.34 and 4.35 the density is 50 points/λ. Comparing these figures with the corresponding figure for the case of a 2-point object indicates that the prediction errors are generally lower. The two fold increase in the point density has negligible effect on the image quality both in the case of the hybrid and linear models.

4.11 Conclusion

A technique for increasing both lateral and aperture-limited range resolution in acoustic holography by aperture expansion has been described. The expansion is achieved by fitting a model to the hologram signal at the available aperture and using the model so constructed to predict the signal at new points outside the aperture. This approach has the advantage that it puts no restrictions on the object range or width which are imposed by the aperture estimation techniques suggested by Sato [128]. However, as in the majority of object restoration techniques, the object is assumed to be of finite width. Moreover, this prediction method operates directly on the measured complex hologram signal without requirement for any modifications which in some cases require the accurate knowledge of the object range for the signal processing to be performed [128]. Restrictions on the imaging configurations are reduced since the Fresnel hologram of a spatially bounded object has been shown to be an analytic function and therefore can be extended beyond an accurately known portion of its analyticity. Other estimation techniques have relied on the fact that the object spectrum is an analytic function and therefore restrictions and modifications had to be introduced on the hologram signal in order to arrive at a signal proportional to the object spectrum.

Because the technique enjoys a certain degree of freedom in the choice of the model to be fitted for the hologram signal, it offers some flexibility in combating the effects of measurement errors and disturbing noise since the noise effects are greatly influenced by the form of the system matrix. In the technique by Sato [128], the noise effect is reduced by modifying the procedure for data acquisition which can be costly, time consuming or unsuitable for use with conventional image reconstruction algorithms.

A number of models have been described which suit a variety of object configurations. A polynomial, space-variant model gives optimum performance for the case of a 1-point object. In this model the signal at every point is related to that at the immediately preceding point through a spatial function. Both power series and orthogonal polynomial expansions have been investigated for this model. For multiple-point objects it was found that in order to obtain adequate prediction accuracies the model should contain a number of linear, space-invariant terms which relate the signal to that at a number of preceding points. For 7 terms in the predictive model, 5 polynomial and 2 linear terms provide optimum results for most configurations considered for multiple-point objects with noise-free data. Resolution improvement by a factor of 4 has been demonstrated using this model for the case of a 2-point object. Although the performance of the linear model is less than optimum regarding prediction accuracy, this model has the advantage that no decision about the model composition is required since it contains no polynomial terms and therefore it represents a more universal model for prediction in the case of multiple-point objects. Moreover,this model has important advantages regarding stability with changes affecting the hologram data in general, cf. sec. 4.10.2 and 3, and robustness against noise as will be discussed in Chapter 5.

A method has been suggested for improving the prediction accuracy obtained with the predictive model. This is achieved by correcting the predicted data using a corrective model based on comparing predicted data with corresponding data which is known to be true. The corrective model is constructed by relating the known data over one half of the available aperture to the predicted data at the same half which is obtained using a predictive model based on the other half.
The performance of the corrective model depends on the structure of both the predictive and corrective models employed and, in general, a search for the optimum corrective model may prove necessary.

The application of the prediction techniques for more complex objects has been investigated. The results obtained with simulated quasi-continuous objects have indicated that as the number of points representing a given object is increased, the prediction error approaches a constant steady-state value. The linear predictive model, which contains no polynomial terms, exhibits greater stability with increased object complexity due to the triangular form of its matrix.

The imaging of quasi-continuous objects made up of a large number of closely spaced points has been demonstrated. This indicates that the prediction technique described in this chapter should be capable of achieving resolution improvement in the case of continuous objects.

CHAPTER 5

APERTURE EXPANSION IN THE PRESENCE OF DISTURBING NOISE

5.1 Introduction

Noise and other practical considerations set the limit for the resolution obtained from a given imaging aperture [116]. Attempts which aim at improving the resolution beyond the diffraction limit or the wavelength limit are faced with the same problem, cf. chapter 3. The limiting effect of noise on the majority of resolution improvement techniques becomes immediately apparent from the principle of analytic continuation which such techniques employ, since this principle requires the exact knowledge of the analytic function within a finite region. This is not possible in practice due to noise, measurement and quantisation errors, and the limited accuracy of scanning the hologram and of performing the required numerical calculations. Since each of these factors can be considered as some form of noise, this term will be used when referring to these factors in general.

All object restoration techniques based on continuing the object spectrum involve the solution of a set of linear equations. In the techniques which utilize the sampling theorem, for example those due to Harris [118] and Sato [128], the equations are solved for new values of the spectrum outside the spatial frequency passband of the imaging system. In general, these equations take the following matrix form:

$$A \underline{x} = \underline{b} \tag{5.1}$$

where $\underline{x}$ is a vector containing the unknown values of the independent samples in the extended spectrum, $\underline{b}$ is the data vector containing the measured samples within the passband, and A is a complex matrix, usually square for an exactly determined system. The elements of A are defined by the positions of the sampling points within the passband. For example, in the technique by Sato [128], the (i,j)th element of this matrix is given by: cf. eqn. (3.25):

$$A_{ij} = \text{sinc}\left[\frac{1}{\Delta f_o}\ (i\Delta f - j\Delta f_o)\ \right] \tag{5.2}$$

where Δf_o is the critical sampling interval for the given object spectrum and Δf is the actual sampling interval, i is the index indicating the position of the measured sample within the passband, j is the index of the independent sample in the extended spectrum, and sinc(x) denotes $\sin(\pi x)/(\pi x)$. The unknown samples are given by:

$$\underline{x} = A^{-1} \underline{b} \tag{5.3}$$

Equation (5.3) indicates that each value of the unknown samples is a linear combination of all the measured values. Since each of these measured values contribute some noise, the result is a greater error in the calculated samples, $\underline{x}$. The error magnification is a function of the magnitude of the elements of the inverse matrix A^{-1}. The larger these elements the greater the error magnification that results and therefore the greater the sensitivity to noise. A useful parameter of the magnitude of the matrix elements is the Euclidean norm of the matrix [136]. For a matrix C, the norm is defined as:

$$||C|| \triangleq [\sum_i \sum_j |C_{ij}|^2]^{\frac{1}{2}} \tag{5.4}$$

In practice, the hologram is sampled at small intervals which correspond to small values of Δf. Moreover, the hologram expansion ratio in the estimation technique by Sato increases with the decrease in Δf. On the other hand, for small values of Δf, the row vectors of the matrix A become almost identical especially for matrices of large sizes. As a result, the norm $||A^{-1}||$ becomes very large and small amounts of errors in the measured data cause large errors in the resulting image. Increasing the sample spacing improves the situation but reduces the margin of resolution improvement with this technique since a stage will be reached when all the samples within the hologram will be independent samples. The dependency between the rows of the system matrix can be reduced through random sampling. This, however, is not convenient since conventional image reconstruction techniques assume uniform sampling of the hologram. Other methods which attempt to solve this problem by modifying the data acquisition phase of the imaging process may also be undesirable due to the increased time or cost of the operation [128].

The prediction technique for hologram aperture expansion described in Chapter 4 is also affected by disturbing noise in a similar way. In the stage of model construction for either prediction or correction, the system is described in the matrix form given in eqn. (5.1). In this case, $\underline{x}$ is the vector containing the unknown model coefficients and $\underline{b}$ contains the measured complex hologram signals in the available aperture. The structure of the matrix A depends on the type of model used and in general its elements are a function of both the measured signals and the positions of the sampling points. In the general case of the hybrid model, the matrix includes zero elements, cf. eqn. (4.65). The effect of noise on the model coefficients and hence on the resulting errors in the predicted signal is therefore a function of the model order as well as its type and composition.

This indicates that this prediction technique enjoys a certain degree of flexibility in combating the effects of noise since the type of model can in principle be chosen to reduce the effect of noise. However, this is unlikely to be without compromising the prediction performance since the model with good stability to noise is not necessarily that which gives best prediction accuracies and hence a compromise has to be made. Nevertheless, the advantage of this approach is that adequate performance with noise can be achieved through signal processing means only with uniform sampling, therefore avoiding any modifications to the data acquisition phase or the requirement for non-conventional algorithms for image reconstruction.

In this chapter, a statistical analysis of the effects of noise on the hologram signal and on the performance of the polynomial model is presented. The effects of the order, type, and composition of the models described in Chapter 4 on the performance with noise are then considered together with possible modifications which would increase the stability of the matrix solution. The strategy for achieving such improvements is to introduce some sparisty into the model matrix or utilize existing sparsity in the matrix. Due to the reduced dependence of the model coefficients on the measured data with such matrices, the model becomes more robust against variations in such signals. When the square matrix, polynomial predictive model is modified to take a triangular form, it provides adequate means of predicting holograms of 1-point objects in the presence of high levels of noise; although it must be followed by a corrective model of the same type. The triangular form comes naturally in the linear predictive model, eqn. (4.61) , and provides high stability and adequate prediction accuracy for the case of multiple-point objects. This chapter also deals with the effect of computation accuracy in general and on the choice of the optimum order of the prediction model in particular. Simulation results showing the effects of noise on range resolution and on the imaging of objects with increased complexity are also presented.

Throughout this chapter, the noise is simulated by adding random components with zero mean to the calculated real and imaginary components of the hologram data as mentioned in section 4.3. The noise levels are expressed as the relative amplitude of the peak of the noise component with respect to the hologram signal.

5.2 Effect of Noise on the Polynomial Predictive Model

In the simulation results discussed in Chapter 4, it has been assumed that the hologram data is known to an accuracy which is limited only by the word length of the computer used and the accuracy of performing the required computations. In a practical situation, however the process of constructing the predictive model for aperture expansion will be subject to the effect of noise contributed by the following sources of error:

(1) Measurement errors in the hologram signal due to the presence of noise and the limited accuracy of the measuring circuits.

(2) Quantization errors introduced when the hologram data is converted into a digital form for processing.

(3) Errors introduced in the positions of the sampling points due to the limited precision of the scanning arrangement and the jitter in the movement of the hydrophone detector when mechanical scanning is used. When arrays are used to sample the hologram, such errors will be caused by the limited manufacturing tolerance on the locations of the array elements . The errors in the positions of the sampling points will be referred to as distance errors throughout this chapter.

(4) The round-off errors due to the limited computation accuracy when solving the set of linear equations which describe the model.

(5) Interference due to extraneous coherent returns from nearby objects as a result of multipath effects in a practical imaging system. This can be reduced by using locallized insonification and range gating techniques. However, this problem is beyond the scope of this thesis.

Errors in (1) and (2) above can be considered as additive noise in the measured hologram signal. This section is concerned with the effects of this additive noise together with the distance errors on the hologram signal and the resulting errors in the model coefficients which, in turn, influence the prediction accuracy. The effect of the computation accuracy is discussed in section 5.8.

5.2.1 Effect of Noise and Distance Errors on the Measured Hologram Signal

Referring to eqn. (4.37) and Fig. 4.1, assume that the modelled hologram function is the ratio between the complex hologram signals H(k) and H(k-1) at points k and k-1. Let this ratio in the absence of any noise or distance errors be $h_o(k)$, therefore:

$$h_o(k) = \frac{H(k)}{H(k-1)}$$

$$= \psi(k) \qquad (5.5)$$

where $\psi(k)$ is the spatial function describing the model. Additive noise in the measured quantities H(k) and H(k-1) introduce a random component n_k into $h_o(k)$. Assuming that no distance errors exist in the positions of the sampling points, the ratio function h(k) in the presence of additive noise is given by:

$$h(k) = h_o(k) + n_k \qquad (5.6)$$

To investigate the effect of the distance errors in the positions of the samples in the hologram plane on the construction of the polynomial predictive model of order L, the ratio function is expanded as a truncated power series of the form:

$$h_o(k) = \sum_{i=1}^{L} a_i\, d_k^{\,i-1} \qquad (5.7)$$

assuming no additive noise or distance errors. In eqn. (5.7) d_k is the distance from point k to the centre of the aperture and a_i ; i=1, 2,...,L are the complex model coefficients. In practice, d_k will be in error by ϵ_k , where ϵ_k is a random error in the position of sample k. The resulting hologram signal is therefore:

$$h(k) = \sum_{i=1}^{L} a_i\,(d_k + \epsilon_k)^{i-1}$$

$$= \sum_{i=1}^{L} a_i[d_k^{i-1}+(i-1)\epsilon_k d_k^{i-2}+ \frac{(i-1)(i-2)}{2!}\,\epsilon_k^{\,2} d_k^{i-3}+\ldots] \qquad \ldots(5.8)$$

For small values of the distance error, high order terms of ϵ_k can be ignored and eqn. (5.8) can be approximated to:

$$h(k) \approx \sum_{i=1}^{L} a_i d_k^{i-1} + \varepsilon_k [\sum_{i=1}^{L} a_i (i-1)\, d_k^{i-2}] \tag{5.9}$$

Using eqn. (5.7):

$$h(k) \approx h_o(k) + \varepsilon_k [\frac{\partial h_o(k)}{\partial d_k}] \tag{5.10}$$

For a 1-point object in the Fresnel zone, $h_o(k)$ is given by eqn. (4.52), from which:

$$\frac{\partial h_o(k)}{\partial d_k} = j \frac{K\Delta}{z} h_o(k) \tag{5.11}$$

Substituting in eqn. (5.10) gives:

$$h(k) = [1 + j \frac{K\Delta}{z} \varepsilon_k] h_o(k) \tag{5.12}$$

This shows that the distance errors appear in the hologram ratio function as multiplicative noise. A model showing the effects of both types of noise on $h_o(k)$ is shown in Fig. 5.1. Assume that the random error ε_k has zero mean and a uniform probability distribution over the interval $-\varepsilon_o \leqslant \varepsilon_k \leqslant \varepsilon_o$ where ε_o is the maximum distance error, see Fig. 5.2. In a mechanical scanning system, ε_o is determined by the precision of the scanning apparatus and the maximum jitter in the movement of the sampling detector. From the above assumptions,

$$E(\varepsilon_k) = 0 \tag{5.13}$$

where E denotes the expected value. The variance of ε_k is:

$$E[\varepsilon_k^2] = \int_{-\varepsilon_o}^{\varepsilon_o} p(\varepsilon_k)\, \varepsilon_k^2 \, d\varepsilon_k$$

$$= \frac{\varepsilon_o^2}{3} \tag{5.14}$$

From eqn. (5.12), the mean value of h(k) is:

$$E[h(k)] = h_o(k) \qquad (5.15)$$

and the variance, Var [h(k)] is:

$$E[|\, h(k) - E[h(k)]|^2] = \frac{K^2\Delta^2}{z^2} \frac{\varepsilon_o^{\,2}}{3} \,|\, h_o(k)|^{\,2} \qquad (5.16)$$

The normalized mean square error in the hologram ratio h(k) due to the distance errors is therefore:

$$\frac{\mathrm{Var}[h(k)]}{|E[h(k)]|^2} = \frac{K^2\Delta^2\varepsilon_o^2}{3\,z^2} \qquad (5.17)$$

This shows that the relative error in h(k) increases at shorter ranges and with greater sample spacing. Both of these factors increase the sensitivity of the hologram signal to variations in the position of the sampling point over the hologram aperture.

5.2.2 Effect of Noise and Distance Errors on the Model Coefficients

In the noise-free case, the model coefficients are derived by solving the L linear equations obtained by writing eqn. (5.7) for k=2,3,...,M, where M is the number of hologram samples in each half of the aperture, cf. sec. 4.4. In vector representation, this takes the form:

$$\underline{h}_o = A\,\underline{a} \qquad (5.18)$$

where A is the real matrix:

$$A = \begin{bmatrix} 1 & d_2 & d_2^{\,2} & \cdot & d_2^{\,L-1} \\ 1 & d_3 & d_3^{\,2} & \cdot & d_3^{\,L-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & d_L & d_L^{\,2} & \cdot & d_L^{\,L-1} \\ 1 & d_{L+1} & d_{L+1}^{2} & \cdot & d_{L+1}^{L-1} \end{bmatrix} \qquad (5.19)$$

and $\underline{h}_o$ and $\underline{a}$ are the vectors containing the hologram signal ratios at every two neighbouring points k and k-1 and the model coefficients respectively. In the presence of distance errors, the model is described by eqn. (5.12) which takes the vector form:

$$\underline{h} = \{ I + j \frac{K\Delta}{z} \begin{bmatrix} \varepsilon_1 & & 0 \\ & \varepsilon_2 & \\ 0 & & \cdot \\ & & & \varepsilon_L \end{bmatrix} \} A \underline{a} \quad (5.20)$$

where $\underline{h}_o$ has been substituted for from eqn. (5.18), I is the unit matrix, and $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_L$ are the distance errors in the positions of the samples at k=2,3,...,M. Let the diagonal matrix representing the distance errors be denoted, D,

$$D = \begin{bmatrix} \varepsilon_1 & & 0 \\ & \varepsilon_2 & \\ 0 & & \cdot \\ & & & \varepsilon_L \end{bmatrix} \quad (5.21)$$

In practice, $\underline{h}$ will also contain the effect of additive noise in the measured hologram signals. Therefore, the combined effect of both types of noise is described by:

$$\underline{h}_o + \underline{n} = [I + j \frac{K\Delta}{z} D] \; A(\underline{a}+\delta\underline{a}) \quad (5.22)$$

where $\delta\underline{a}$ is the net error in the model coefficients and $\underline{n}$ is the vector representing the random components due to additive noise in the hologram signals,

$$\underline{n} = \begin{bmatrix} n_1 \\ n_2 \\ \cdot \\ n_L \end{bmatrix} \quad (5.23)$$

Subtracting eqn. (5.18) from (5.22) gives:

$$\underline{n} = j\frac{K\Delta}{z} D A \underline{a} + A \delta\underline{a} + j\frac{K\Delta}{z} D A \delta\underline{a} \qquad (5.24)$$

The third term at the R.H.S. of eqn. (5.24) contains the products of rather small quantities representing the errors in the model coefficients and in the positions of the sampling points and therefore can be considered negligible, hence:

$$\delta\underline{a} \approx A^{-1} \underline{n} - j\frac{K\Delta}{z} A^{-1} D \underline{h}_o \qquad (5.25)$$

With ϵ_k and n_k as uncorrelated random variables with zero mean, the mean error in the solution vector $\underline{a}$ is:

$$E[\delta\underline{a}] = A^{-1} E[\underline{n}] - j\frac{K\Delta}{z} A^{-1} E[D] \underline{h}_o$$

$$= 0 \qquad (5.26)$$

The covariance of the error is:

$$E[\delta\underline{a}\ \delta\underline{a}^{*T}] = E[\{A^{-1}\underline{n} - j\frac{K\Delta}{z} A^{-1}D\underline{h}_o\}\ \{A^{-1}\underline{n} + j\frac{K\Delta}{z} A^{-1}Dh_o\}^{*T}]$$

$$= A^{-1} E[\underline{n}\ \underline{n}^{*T}](A^{-1})^T + \frac{K^2\Delta^2}{z^2} A^{-1}E[D\underline{h}_o \underline{h}_o^{*T}D^T](A^{-1})^T \qquad (5.27)$$

since the cross-products average to zero. '*' denotes the complex conjugate and 'T' indicates vector or matrix transpose. The contribution of the additive noise is represented by:

$$E[\underline{n}\ \underline{n}^{*T}] = E \begin{bmatrix} |n_1|^2 & n_1 n_2^* & \cdot & n_1 n_2^* \\ n_2 n_1^* & |n_2|^2 & \cdot & n_2 n_L^* \\ \cdot & \cdot & \cdot & \cdot \\ n_L n_1^* & n_L n_2^* & \cdot & |n_L|^2 \end{bmatrix} \qquad (5.28)$$

The components of the noise vector $\underline{n}$ are assumed uncorrelated, with identical noise distributions at the various sampling points. In this case, the variance of n_k is constant and equal to σ_h^2 for all values of k, therefore:

$$E[n_i n_j^*] = \begin{cases} 0 & i \neq j \\ \sigma_h^2 & i = j \end{cases} \tag{5.29}$$

Substituting from eqn. (5.29) into eqn. (5.28):

$$E[\underline{n}\,\underline{n}^{*T}] = \sigma_h^2\, I \tag{5.30}$$

Consider $E[D\,\underline{h}_o\,\underline{h}_o^{*T} D^T]$ which represents the contribution of the distance errors:

$$D\,\underline{h}_o\,\underline{h}_o^{*T} D^T = \begin{bmatrix} \varepsilon_1 & & & \\ & \varepsilon_2 & & 0 \\ & & \cdot & \\ 0 & & & \varepsilon_L \end{bmatrix} \begin{bmatrix} h_{o1} \\ h_{o2} \\ \cdot \\ h_{oL} \end{bmatrix} \begin{bmatrix} h_{o1}^* & h_{o2}^* & \cdot & h_{oL}^* \end{bmatrix} \begin{bmatrix} \varepsilon_1 & & & 0 \\ & \varepsilon_2 & & \\ & & \cdot & \\ 0 & & & \varepsilon_L \end{bmatrix}$$

$$= \begin{bmatrix} \varepsilon_1^2 |h_{o1}|^2 & \varepsilon_1\varepsilon_2\, h_{o1}h_{o2}^* & \cdot & \varepsilon_1\varepsilon_L\, h_{o1}h_{oL}^* \\ \varepsilon_2\varepsilon_1\, h_{o2}h_{o1}^* & \varepsilon_2^2\, |h_{o2}|^2 & \cdot & \varepsilon_2\varepsilon_L\, h_{o2}h_{oL}^* \\ \cdot & \cdot & \cdot & \cdot \\ \varepsilon_L\varepsilon_1\, h_{oL}h_{o1}^* & \varepsilon_L\varepsilon_2\, h_{oL}h_{o2}^* & \cdot & \varepsilon_L^2\, |h_{oL}|^2 \end{bmatrix} \quad \ldots(5.31)$$

If the distance errors are assumed to be uncorrelated at the hologram sampling points and to have the same probability distribution with variance σ_d^2 then:

$$E[\epsilon_i \epsilon_j] = \begin{cases} 0 & i \neq j \\ \sigma_d^2 & i=j \end{cases} \tag{5.32}$$

Using eqns.(5.31) and (5.32) produces:

$$E[D\underline{h}_o \underline{h}_o^{*T} D^T] = \sigma_d^2 \begin{bmatrix} |h_{o1}|^2 & & & 0 \\ & |h_{o2}|^2 & & \\ & & \cdot & \\ 0 & & & |h_{oL}|^2 \end{bmatrix}$$

...(5.33)

Substituting from eqns. (5.30) and (5.33) into (5.27):

$$E[\delta\underline{a} \;\; \delta\underline{a}^{*T}] = A^{-1}\{\sigma_h^2 I + \sigma_d^2 \frac{K^2\Delta^2}{z^2} \begin{bmatrix} |h_{o1}|^2 & & & \\ & |h_{o2}|^2 & 0 & \\ & & \cdot & \\ 0 & & & |h_{oL}|^2 \end{bmatrix}\}(A^{-1})^T$$

$$\triangleq A^{-1} \; Q \; (A^{-1})^T \tag{5.34}$$

where Q is the triangular matrix:

$$Q = \begin{bmatrix} \sigma_h^2 + \alpha^2|h_{o1}|^2 & & & \\ & \sigma_h^2 + \alpha^2|h_{o2}|^2 & & 0 \\ & & \cdot & \\ 0 & & & \sigma_h^2 + \alpha^2|h_{oL}|^2 \end{bmatrix}$$

...(5.35)

and

$$\alpha \triangleq \frac{K\Delta}{z} \sigma_d \tag{5.36}$$

Let M_a be the covariance matrix of the error $\delta\underline{a}$:

$$M_a = A^{-1} Q (A^{-1})^T \tag{5.37}$$

The Taxicab norm [136] of M_a, which is defined as the sum of the moduli of all the matrix elements, satisfies the inequality:

$$||M_a||_1 \leqslant ||A^{-1}||_1 ||Q||_1 ||(A^{-1})^T||_1$$

$$\leqslant ||A^{-1}||_1^2 ||Q||_1 \qquad (5.38)$$

Substituting for $||Q||_1$ from eqn. (5.35) and using eqn. (5.36) yields:

$$||M_a||_1 \leqslant ||A^{-1}||_1^2 [L \sigma_h^2 + \frac{K^2 \Delta^2}{z^2} \sigma_d^2 \sum_{i=1}^{L} |h_{oi}|^2] \qquad ...(5.39)$$

The Euclidean norm [136] for the vector $\underline{h}_o$ is defined as:

$$||\underline{h}_o||_2 \triangleq (\sum_{i=1}^{L} |h_{oi}|^2)^{\frac{1}{2}} \qquad (5.40)$$

For the matrix A this norm is:

$$||A||_2 = (\sum_{i=1}^{L} \sum_{j=1}^{L} |A_{ij}|^2)^{\frac{1}{2}} \qquad (5.41)$$

The Euclidean norms for the components of eqn. (5.18) satisfy the inequality:

$$||\underline{h}_o||_2 \leqslant ||A||_2 \; ||\underline{a}||_2 \qquad (5.42)$$

therefore,

$$\frac{1}{\sum_{i=1}^{L} |a_i|^2} \leqslant \frac{||A||_2^2}{\sum_{i=1}^{L} |h_{oi}|^2} \qquad (5.43)$$

From eqns. (5.39) and (5.43):

$$\frac{||M_a||_1}{\sum_{i=1}^{L} |a_i|^2} \leqslant ||A^{-1}||_1^2 \; ||A||_2^2 \; [\frac{L\sigma_h^2}{\sum_{i=1}^{L} |h_{oi}|^2} + \frac{K^2 \Delta^2}{z^2} \sigma_d^2] \qquad ...(5.44)$$

The quantity at the L.H.S. of eqn. (5.44) represents the noise-to-signal power ratio at the output of the model construction process while the first term inside the square brackets at the R.H.S. represents the noise-to-signal ratio in the input data vector $\underline{h}_o$ and the second term, $\frac{K^2\Delta^2}{z^2}\sigma_d^2$, represents the contribution due to distance errors. Therefore, this equation shows that the noise-to-signal ratio in the model coefficients is determined by the sum of the additive noise-to-signal ratio in the data vector due to measurement and quantization errors and the contribution caused by errors in the positions of the hologram samples. The distance errors are weighted by the factor $\frac{K^2\Delta^2}{z^2}$. Since this factor is normally <<1 for most imaging configurations, the effect of these errors would be much smaller compared to that of the additive noise in the vector $\underline{h}_o$. Moreover, the equation indicates that the effect of both sources of error is magnified by the product $||A^{-1}||_1^2 \; ||A||_2^2$. When the system matrix is ill-conditioned, this product becomes large, and small errors in the data vector $\underline{h}_o$ or the sample positions appear greatly magnified in the model coefficients $\underline{a}$. The norm product of the system matrix and its inverse is a useful parameter for determining the sensitivity of the predictive and corrective models to noise.

5.3 Sensitivity to Noise versus Model Type and Composition

When noisy hologram data was simulated for the case of 1 or 2-point objects, the prediction errors were so large with both the polynomial and the hybrid models that only a few predicted points could be obtained with adequate accuracy even for noise levels as low as 1%. This led to investigating the effect of noise as a function of the model composition in general. For the case of the hybrid model discussed in section 4.7, the optimum composition was obtained on the sole basis of maximizing the prediction accuracy when the hologram data is noiseless. However, from section 5.1 and the analysis given in section 5.2, the norm of the model matrix and its inverse, and hence the model type and composition, affect the stability of the matrix solution with noise and therefore the overall prediction accuracy. This factor must be taken into account when selecting an optimum model to be used with noisy data.

The plots in Fig. 5.3 show the effect of increasing the number of linear terms, q, in a hybrid prediction model of order L=9 on both the maximum prediction error over the aperture without noise, Fig. 5.3a, and with 1% noise in the hologram signal, Fig. 5.3b for the case of a 1-point object when doubling the size of the available aperture. The prediction error is plotted in dbs relative to 1%.

These plots indicate that while the prediction performance of the model deteriorates with the increase in the number of linear terms, the sensitivity of the prediction accuracy to noise generally decreases. It is noted that the reduction in the sensitivity to noise far exceeds the increase in the prediction error without noise. Therefore, it is possible that the use of a model which is not considered optimum in the case of noiseless data would lead to results which are better than those obtained with an optimum but noise-sensitive model. For example, a 7/2 hybrid model produces an error of -26.5 dbs without noise but this increases to 347 dbs with 1% noise while for a 0/9 model these figures are 45.7 dbs and 45.4 dbs respectively.

Fig. 5.3c shows the variations in the Euclidean norm of the inverse matrix $||A^{-1}||$ with increasing the number of linear terms in the case of noiseless data. As expected, the decrease in the sensitivity to noise shown in Fig. 5.3b is accompanied by a decrease in the norm of the inverse matrix which is responsible for the increased magnification of the errors in the noisy hologram data.

Fig. 5.3 also indicates the influence of the triangular form of the model matrix on both the sensitivity to noise and the prediction accuracy. When the number of linear terms in a hybrid model is increased, the size of the triangular portion at the top R.H. corner of the matrix, eqn (4.65), increases. This reduces the dependence between the rows of the model matrix and improves the stability of the solution with noise. However, the resulting increase in the number of zeros in the matrix makes the model less representative of the hologram signal over the available aperture and therefore the prediction accuracy of the model is reduced. This shows that by selecting a model which is represented by a suitable matrix it is possible in principle to improve the stability of the performance with noise at the expense of increasing the basic prediction errors. Since the improvement in the stability with noise can be chosen to outweigh the reduction in the basic prediction accuracy, it is possible to achieve an overall improvement in the prediction performance in the presence of noise.

5.4 Stability of the Square and the Triangular Polynomial Models

Consider a polynomial prediction model of order L of the type described in section 4.4. In order to isolate the effect of disturbing noise in the hologram data so that it does not influence the model matrix, the model is described using eqn. (5.18) where the model matrix is real and is given by eqn. (5.19). This matrix is only a function of the positions of the hologram sampling points and is not affected by noise in the hologram signal. This form of the matrix is useful in investigating the effect of the structure of the model on the performance with noise. For hologram samples spaced uniformly at distance Δ wavelengths, the system matrix, A, is given by:

$$A = \begin{bmatrix} 1 & 1.5\Delta & (1.5\Delta)^2 & \cdot & (1.5\Delta)^{L-1} \\ 1 & 2.5\Delta & (2.5\Delta)^2 & \cdot & (2.5\Delta)^{L-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & (L-0.5)\Delta & [(L-0.5)\Delta]^2 & \cdot & [(L-0.5)\Delta]^{L-1} \\ 1 & (L+0.5)\Delta & [(L+0.5)\Delta]^2 & \cdot & [(L+0.5)\Delta]^{L-1} \end{bmatrix} \qquad (5.45)$$

Due to the uniform sampling, the lower rows of the matrix will tend to become identical for large values of the matrix size L. This leads to an increase in the norm $||A^{-1}||$ of the inverse matrix and therefore an increase in the sensitivity of the solution vector, $\underline{a}$, representing the model coefficients,to small perturbations in the data vector, $\underline{h_o}$, cf. eqn. (5.18). It is interesting to note that the uniform sampling has also been responsible for the high noise sensitivity in the estimation technique proposed by Sato since it affects the system matrix in basically the same way [128].

As a measure of the dependency between the neighbouring high order rows of the matrix in eqn. (5.45), consider the relative difference between the last two rows. If A_{L-1} and A_L are the last two row vectors, then the relative norm of the difference between the two rows is given by:

$$\delta_L = \frac{||A_L - A_{L-1}||}{||A_L||} \qquad (5.46)$$

where $||\ \ ||$ denotes the Euclidean norm of the vector.

In the matrix of eqn. (5.45), let:

$$u = (L+0.5)\Delta$$
$$v = (L-0.5)\Delta \tag{5.47}$$

For this matrix:

$$||A_L-A_{L-1}||^2 = (u-v)^2+(u^2-v^2)^2+\ldots+(u^{L-1}-v^{L-1})^2$$
$$= u^2+u^4+u^6+\ldots+u^{2(L-1)}$$
$$+v^2+v^4+v^6+\ldots+v^{2(L-1)}$$
$$-2[uv+(uv)^2+(uv)^4+\ldots+(uv)^{L-1}] \tag{5.48}$$

Substituting with the sum of the L terms in each geometric series yields:

$$||A_L-A_{L-1}||^2 = u^2\,\frac{(u^{2L}-1)}{u^2-1} + v^2\,\frac{(v^{2L}-1)}{v^2-1} - \frac{2uv((uv)^L-1)}{uv-1} \qquad \ldots(5.49)$$

and

$$||A_L||^2 = 1 + u^2 + u^4 + u^6 +\ldots+ u^{2(L-1)}$$
$$= \frac{u^{2L}-1}{u^2-1} \tag{5.50}$$

Substituting from eqns. (5.49) and (5.50) into (5.46) gives:

$$\delta_L^{\;2} = u^2+v^2\left(\frac{u^2-1}{v^2-1}\right)\left(\frac{v^{2L}-1}{u^{2L}-1}\right) -2uv\left(\frac{u^2-1}{uv-1}\right)\left(\frac{(uv)^L-1}{u^{2L}-1}\right) \tag{5.51}$$

For large matrix sizes:

$$L \gg 0.5 \tag{5.52}$$

and, from eqn. (5.47), u and v become approximately equal. In most practical cases Δ, the sample spacing in wavelengths, is not much smaller, if not greater, than unity and therefore:

$$u \approx v \approx L\,\Delta >> 1 \tag{5.53}$$

The square of the relative norm of the difference between the last two rows in the matrox is therefore:

$$\begin{aligned}\delta_L{}^2 &\approx u^2 + v^2 - 2v^2 \\ &\approx u^2 - v^2 \\ &\approx 0 \qquad \text{for large matrices}\end{aligned} \tag{5.54}$$

This shows that for large model sizes, the high order rows of the matrix in eqn. (5.45) will tend to become equal, therefore making the matrix nearly rank deficient. Attempts to solve the matrix under this condition yields large magnification of errors in the data vector due to the large norm of the inverse matrix A^{-1} [138].

Eqn. (5.53) suggests that for a given model size, L, reducing the sample spacing Δ increases δ_L and therefore reduces the amount of error magnification, In practice, however, it is often required that a given number of samples should be as widely spaced as possible for efficient sampling of the required hologram aperture. Moreover, this approach has the disadvantage that it affects the data acquisition phase of the imaging process in such a manner in which parameters such as the sample spacing are compromised in order to satisfy the requirements of the hologram expansion technique.

A more convenient approach is to modify the model matrix in order to reduce the row dependence for large sizes. Consider the triangular matrix which consists of the top left triangle of the matrix in eqn. (5.45) including the diagonal elements. This takes the form:

$$A = \begin{bmatrix} 1 & 1.5\Delta & (1.5\Delta)^2 & . & (1.5\Delta)^{L-1} \\ 1 & 2.5\Delta & (2.5\Delta)^2 & . & 0 \\ . & . & . & . & . \\ 1 & (L-0.5)\Delta & 0 & . & 0 \\ 1 & 0 & 0 & . & 0 \end{bmatrix} \tag{5.55}$$

The dependence between the rows of this matrix is reduced since the difference between any two adjacent rows A_i and A_{i+1} will be greater than the matrix element $A_{i,L+1-i}$ which is the last non zero element in row A_i. With this matrix it is less straightforward to predict the rows with the minimum relative difference. If the criterion in eqn. (5.46) is used as a guide it gives:

$$\delta_L = (L-0.5)\Delta \tag{5.56}$$

Contrary to the case of the square matrix, eqn. (5.56) indicates that for a given L, reducing the sample spacing Δ will have the opposite effect of increasing the ill-conditioning of the triangular matrix. It should be noted, however, that this result is based on the assumption that the last two rows are the cause of the ill-conditioning.

In order to compare the stability of the solution using both the square and the triangular matrices consider the bounds on the perturbations in the solution vector which are caused by a given perturbation in the data vector. Ignoring any distance errors in the positions of the hologram sampling points , assume an error $\delta\underline{h}_o$ in the data vector which causes an error $\delta\underline{a}$ in the solution vector. From eqn. (5.18), the model is described by:

$$\underline{h}_o + \delta\underline{h}_o = A\,(\underline{a} + \delta\underline{a}) \tag{5.57}$$

Subtracting eqn. (5.18) from eqn. (5.57) yields:

$$\delta\underline{a} = A^{-1}\,\delta\underline{h}_o \tag{5.58}$$

The Euclidean norms of the components of eqn. (5.58) satisfy the inequality:

$$||\delta\underline{a}|| \leqslant ||A^{-1}||\ \ ||\delta\underline{h}_o|| \tag{5.59}$$

Similarly, from eqn. (5.18):

$$||\underline{h}_o|| \leqslant ||A||\ \ ||\underline{a}|| \tag{5.60}$$

From eqns. (5.59) and (5.60):

$$\frac{||\delta \underline{a}||}{||a||} \leqslant ||A|| \; ||A^{-1}|| \; \frac{||\delta \underline{h}_o||}{||\underline{h}_o||} \tag{5.61}$$

Similar to the results obtained in eqn. (5.44) for the covariance matrix of the error in the solution vector $\underline{a}$, eqn. (5.61) indicates that the norm product $||A|| \; ||A^{-1}||$ can be considered as a magnification parameter for the relative error in $\underline{h}_o$ since the larger the value of this product the greater the influence of a relatively small disturbance in the data vector on the solution.

Another quantity which is used in the literature as an error magnifying parameter [138] is defined as:

$$\beta \triangleq \frac{S_{max}}{S_{min}} \tag{5.62}$$

where S_{max} and S_{min} are the maximum and minimum non-zero values of the moduli of the singular values of the system matrix A [138].

The norm product $||A|| \; ||A^{-1}||$ and the ratio β can be used to investigate the effects of the matrix size and the sample spacing and the advantages of using a triangular matrix. The singular values of the matrix A which are required for calculating the ratio β in eqn. (5.62) are obtained as the square roots of the corresponding eignevalues of the matrix AA^T where A^T is the transpose of A.

In Fig. 5.4 the ratio β is plotted versus the matrix size, L, for both the square and the triangular models. The curves in Fig. 5.4a for the square model show that the error magnification increases rapidly with the increase in the matrix size. As expected from eqn. (5.53), a reduction in the sample spacing by a factor of 4 achieves some reduction in the value of β. A much more significant improvement is obtained by using the triangular matrix model as shown in Fig. 5.4b. With the same sample spacing of 1λ in both cases, the triangular matrix reduces the error magnification factor by approximately 120 dbs compared to the case of the square matrix for L=7. The effect of the sample spacing in the case of the triangular matrix is shown in Fig. 5.4c. For the values used for the model size and the sample spacing, the error magnification increases with the reduction in the sample spacing as predicted by eqn. (5.56). The curves corresponding to Figs. 5.4a,b,c are shown in Fig. 5.5 for the norm product $||A|| \, ||A^{-1}||$ and are very similar to those in Fig. 5.4.

The plots in Figs. 5.4 and 5.5 show the potential advantage of the triangular matrix over the square matrix regarding the magnification of errors. This is confirmed by the results shown in Fig.5.6 where the relative norm of the error in the solution vector $||\delta \underline{a}||/||\underline{a}||$ is plotted versus the matrix size for both the square and the triangular matrices. The error in the data vector is simulated by inserting random components with a relative amplitude of 10% into the hologram signals of a 1-point object located on the hologram axis at a distance of 100λ from the hologram. While the relative error is large and increases steadily with the increase in the model size in the case of the square matrix, it is much smaller and is less sensitive to variations in the model size in the case of the triangular matrix. This type of matrix, therefore, provides an adequate means of obtaining better stability for the solution without imposing restrictions on the sampling of the hologram.

5.5 The Triangular, Polynomial, Predictive Model

Consider the simple case of an available aperture with M=4 shown in Fig. 5.7. The triangular, polynomial, predictive model is constructed by writing the following set of equations for the signal at points 2,3, and 4.

$$h_{o1} = \frac{H(2)}{H(1)} = a_1 + a_2\, d_2 + a_3\, d_2^{\,2}$$

$$h_{o2} = \frac{H(3)}{H(2)} = a_1 + a_2\, d_3$$

$$h_{o3} = \frac{H(4)}{H(3)} = a_1 \qquad (5.63)$$

The coefficients a_1, a_2, and a_3 are obtained by solving:

$$\begin{bmatrix} 1 & d_2 & d_2^{\,2} \\ 1 & d_3 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} h_{o1} \\ h_{o2} \\ h_{o3} \end{bmatrix} \qquad (5.64)$$

where the matrix takes the triangular form proposed in section 5.4. While this form provides better solution stability as discussed in the previous section, eqn. (5.63) suggests that this would be at the expense of reduced prediction accuracy. This is because the signals at the available aperture do not contribute fully in determining the model coefficients. For example, coefficient a_1 is determined by h_{o3} only while in the case of the square matrix a_1 would be a function of h_{o1} , h_{o2} and h_{o3} and the model coefficients would be a more accurate representation of all the available data on the hologram aperture.

Having solved eqn. (5.64) for the model coefficients, the signal at points 5,6, and 7 outside the aperture extension can be predicted point by point. The origin point is shifted from O to O', see Fig. 5.7, so that the configuration of points, 4,5,6 and 7 relative to the new origin O' is a replica of that of points 1,2,3 and 4 relative to the origin point O. With the signal at point 4 known , the prediction proceeds as follows:

$$H(5) = H(4)\,[a_1 + a_2\, d_5' + a_3\, d_5'^2]$$

$$H(6) = H(5)\,[a_1 + a_2\, d_6']$$

$$H(7) = H(6)\,[a_1] \qquad (5.65)$$

where the distances d_5' and d_6' are measured relative to the new origin O'. Further points can be predicted with reduced accuracy, e.g.:

$$H(8) = H(7)\,[a_1]$$

The inferior performance of the triangular matrix model compared to the square model in the case of noiseless data is shown in Fig. 5.8a for a 1-point object. The maximum prediction error in dbs relative to 1% when doubling the aperture size is plotted versus the matrix size for both the triangular and the square models. The error in the case of the square matrix is much smaller and decreases with the increase in the matrix size. On the other hand, the error is larger for the case of the triangular model and is less sensitive to the increase in the matrix size.

Fig 5.8b shows the improvement in the performance with noise when using the triangular matrix. The maximum prediction error is plotted versus the model size with 10% noise in the hologram signal. While the error is larger and increases rapidly in the case of the square matrix it is much smaller and is almost constant for the triangular matrix. Comparison between Figs. 5.8a and 5.8b shows that for large matrix sizes the improvement in the stability with noise obtained by using the triangular model outweighs the associated increase in the prediction errors.

The effect of increasing the amount of noise on the performance of both types of models is shown in Fig. 5.9 for a model size L=7. The norm of the error in the solution vector, $||\delta \underline{a}||$, due to noise is plotted in Fig. 5.9a for noise levels from 5% to 30% and Fig. 5.9b shows the corresponding results for the maximum prediction error. As expected, Fig. 5.9b indicates that the prediction error is fairly insensitive to the level of noise in the case of the triangular model.

5.6 The Triangular, Polynomial, Corrective Model

Although the triangular form of the polynomial model is relatively stable with noise, the levels of prediction errors using this model are too high for any significant improvement in resolution to be attained. To overcome this limitation, the predicted signals obtained with this model are corrected using a corrective model based on the principle discussed in section 4.8 . However, for the corrective model to have the required stability with noise, it is assumed to be of the triangular form.

Consider the simple case of an 8-point available aperture in Fig. 5.10 which is to be extended to 16 points. As shown in section 4.8, the correction of the four points to the right of the available aperture is achieved by using a corrective model which is based on relating the true signals and the predicted signals at the R.H. side of the available aperture. First, a predictive triangular model with size L=3 is derived from the data H_2 at the L.H. half of the available aperture with the origin point at O_1. This model is then used to predict signals at the four points in the R.H. half of this aperture with the origin point at O_2. Denote the predicted signals H_1'.

A triangular corrective model of size L=4 is constructed by writing the following four equations which relate the known signals H_1 to the predicted signals H_1', the distances are measured from the origin point at O_1.

$$h'_{o1} = \frac{H_1(1)}{H_1'(1)} = u_1+u_2\, d_1 + u_3\, d_1^2 + u_4\, d_1^3$$

$$h'_{o2} = \frac{H_1(2)}{H_1'(2)} = u_1+u_2\, d_2 + u_3\, d_2^2$$

$$h'_{o3} = \frac{H_1(3)}{H_1'(3)} = u_1+u_2\, d_3$$

$$h'_{o4} = \frac{H_1(4)}{H_1'(4)} = u_1 \qquad (5.66)$$

The model coefficients $u_1,\dots,u_4$ are obtained by solving the matrix equation:

$$\begin{bmatrix} 1 & d_1 & d_1^2 & d_1^3 \\ 1 & d_2 & d_2^2 & 0 \\ 1 & d_3 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} h'_{o1} \\ h'_{o2} \\ h'_{o3} \\ h'_{o4} \end{bmatrix} \qquad (5.67)$$

Having determined the coefficients of the predictive model, it can be used to correct the predicted signals in the R.H. aperture extension. The prediction and correction steps are summarized as follows:

(i) Construct a triangular predictive model, L=3, using the true signals H_1 at the R.H. half of the available aperture with the origin at O_3.

(ii) Use the predictive model construction in (i) to predict 4 points in the R.H. aperture extension with origin at O_4. Denote the predicted signals G_1'.

(iii) Obtain the corrected signals, G_1, from the predicted signals G_1' at points 5,6,7,8 with origin at O_3 using the following equations:

$$\begin{aligned} G_1(5) &= G_1'(5)[u_1+u_2\, d_5+u_3\, d_5^2 +u_4\, d_5^3] \\ G_1(6) &= G_1'(6)[u_1+u_2\, d_6 +u_3\, d_6^2] \\ G_1(7) &= G_1'(7)[u_1+u_2\, d_7] \\ G_1(8) &= G_1'(8)[u_1] \end{aligned} \qquad (5.68)$$

The triangular predictive and corrective models have been used to double the size of a 16-point aperture to image a 1-point object in the presence of noise. For the sake of comparison, Fig.5.11 shows the results using the square model with and without noise. In Fig. 5.11a the reconstructed images from the small, predicted, and true apertures in the case of noiseless hologram data. Since the prediction errors with the square model are very small in the absence of noise, the image obtained from the predicted aperture is almost identical with that from the true aperture and no further correction is necessary. Fig. 5.11b shows the corresponding images using the same model with a random component of 0.1% relative amplitude added to the simulated hologram signal. The figure indicates that even with such low levels of noise, the prediction errors using the square model are quite large and the quality of the resulting image is unacceptable.

Fig. 5.12 shows the results obtained using the triangular predictive and corrective models with noiseless data. The prediction errors both before and after correction are shown in Fig. 5.12a which illustrates clearly the effectiveness of the correction technique in reducing the prediction errors. The images reconstructed from the small, predicted, corrected, and true apertures are plotted in Fig. 5.12b. Due to the large errors in the predicted hologram before correction, the corresponding image is of poor quality and no gain in resolution is achieved. However, with the great reduction in the prediction errors achieved through the use of the corrective model, the image from the corrected data is in good agreement with the image from the true aperture.

The results with 10% noise level are shown in Fig. 5.13 . Comparing Figs. 5.13a and 5.12a, the prediction errors exhibit only a small increase due to noise because of the high stability of the triangular model. The image reconstructed from the corrected data is comparable with that from the true hologram. Images obtained with 20% and 30% noise levels are shown in Fig. 5.14. The effect of using different random values for the noise is shown in Fig. 5.15. Because of the stability achieved by the triangular form of the predictive and corrective models, changes in the noise levels or the noise values have little effect on the reconstructed images.

5.7 Imaging of Multiple points in the Presence of Noise

In section 4.7, a hybrid polynomial/linear model has been identified as the optimum predictive model for the holograms of multiple-point objects. Moreover, it was also found that the linear model which gives adequate , if not optimum, prediction performance exhibits a number of desirable stability characteristics. For example, it has been shown in section 4.10 that prediction errors obtained using this model are fairly insensitive to variations in the object range, object width and in the number of points taken to represent a discrete-point object of a given width. In this section, it is shown that this stability extends also to the effects of noise in the hologram signal and therefore the linear model provides adequate means for predicting the holograms of multiple-point objects in the presence of disturbing noise.

Referring to eqn. (4.65) for the general hybrid model and considering the case of L=7 (M=8), for the 5/2 optimum hybrid model there are only two linear terms and the model matrix is dominated by the polynomial expansion terms in the L.H. portion of the matrix. It has been shown in section 5.4 that the polynomial expansions, when carried to their full length, tend to make the matrix nearly rank defficient and therefore the 5/2 hybrid model greatly magnifies the errors due to noise in the hologram signals. On the other hand, the linear model described by eqn. (4.61) takes the form of a triangular matrix where the upper triangle consists of zero elements. It has been shown in section 5.4 that the triangular matrix exhibits better stability with noise in the case of the polynomial model. With this model, the triangular form helps reduce the tendency of the higher order rows of the matrix to become identical and therefore cause instability. It can be shown that the triangular form performs a similar role in the case of the linear model, with the difference that this form appears naturally as a result of the windowing condition in eqn. (4.60). If the data contributing to the construction of the linear model is not restricted to one half of the available aperture at a time, eqn. (4.60) would be modified such that the hologram signal at each point on the aperture half under consideration is expressed as a linear combination of L preceding points, a number of which would be situated on the other half. The matrix representing the model in this case would be:

$$A = \begin{bmatrix} H(1) & H(-1) & H(-2) & . & H(-L+1) \\ H(2) & H(1) & H(-1) & . & H(-L+2) \\ . & . & . & . & . \\ H(L-1) & H(L-2) & H(L-3) & . & H(-1) \\ H(L) & H(L-1) & H(L-2) & . & H(1) \end{bmatrix} \tag{5.69}$$

where H(-1), H(-2),...,H(-L+1) are hologram signals at sampling points on the other half of the available aperture.

Eqn. (5.69) indicates that neighbouring rows of the matrix contain hologram signals which correspond to a shift of only one sample along the hologram. Since the samples are taken close enough to prevent undersampling, neighbouring rows would be almost identical particularly when the spatial frequency content of the hologram is low; for example when the object range is large. This would cause instability in the same manner described in section 5.4 although it is expected that this instability would exist even for small matrix sizes. With the triangular form of the matrix for the linear model, cf. eqn.(4.61), neighbouring rows differ by at least the value of one hologram signal and therefore the matrix solution is more stable with variations in the data vector. This explains the relative insensitivity of the linear model to variations in the hologram and object parameters and suggests that this should also be the case regarding noise in the hologram data.

The maximum prediction error when doubling the size of a 16-point available aperture is plotted in Fig. 5.16 versus the level of noise from 0 to 30% for both the 5/2 hybrid model and the 0/7 linear model. The object consists of two points separated by 5.2λ at a distance of 105λ from the 15λ available aperture. With zero noise, the hybrid model gives better prediction accuracy but the prediction error increases steadily with noise while for the linear model the error is fairly constant with the increase in the noise level. This curve is similar to that in Fig. 5.9b for the polynomial model.

Fig. 5.17 shows the prediction errors and reconstructed images obtained when the linear model is used to double the size of a 16-point aperture to image 2 points separated by a distance of 5.2λ, range 105λ and positioned symmetrical about the hologram axis assuming noiseless data. The corresponding results with 10% noise are shown in Fig. 5.18. The increase in the prediction errors is very small compared to the case of noiseless data and the two points can still be clearly resolved. The images reconstructed with 20% and 30% noise in the simulated hologram signals are shown in Fig.5.19. The effect of using different random values for the noise is shown in Fig. 5.20. The last three figures indicate that the linear predictive model provides the required stability for imaging multiple-point objects in noisy environments using the expanded apertures.

In the above mentioned figures for images reconstructed from noisy hologram data, it is noted that in some cases the image reconstructed from a predicted hologram may appear to be of better quality than the image obtained from the 'true' hologram. It should be noted,however, that in the above examples the data in the 'true' hologram is itself noisy and therefore the image reconstructed from this hologram is in error when compared to the case of an ideal noise-free hologram. Due to the random nature of noise, it is possible that the errors in the predicted holograms are such that they partially offset the errors caused by noise in the 'true' hologram, leading to a more accurate representation of the exact hologram and therefore a better image.

5.8 Optimum Size of Available Aperture and the Effects of Computation Accuracy

The number of points in the available aperture affects the accuracy of the predicted holograms in two ways. First, it determines the size of the predictive and corrective models for an exactly determined system and, for a given expansion ratio, it also determines the number of predicted points in the aperture extension. Consider the simple case of a 1-point object and ignore the effects of disturbing noise in the hologram signal. The available aperture contains a total number of 2M points and it is required to predict a number of points equal to L, the size of the predictive model,on both sides of the available aperture; where L=M-1. Assume that the optimum predictive model, i.e. the square polynomial type, is used and that the object range is large enough to prevent undersampling.

Since the hologram in this case contains a single pattern of spatial variations which corresponds to the point object, it can be assumed that as the model size increases, with the increase in the number of points in the available aperture, the model becomes a more accurate representation of the hologram function. This assumption is justified by the error analysis in section 4.5 which shows that prediction errors using the polynomial model should decrease steadily with the increase in L for:

$$\frac{2\pi\Delta\ d_k}{z} \leqslant 1 \qquad (5.70)$$

where Δ is the sample spacing, d_k is the distance from the predicted point to the centre of the aperture and z is the range. Therefore, under the assumption in eqn. (5.70), the increase in L leads to more accurate prediction of the first point in the aperture extension, k=M+1, and the prediction error E(M+1) decreases steadily with the increase in L. Due to the accumulation of the prediction errors as more new points are predicted, the error at any point in the aperture extension will depend both on the accuracy with which the first point (M+1) is predicted as well as the location of the point under consideration relative to the available aperture.

The maximum error at the edge of the expanded aperture, E(M+L), when predicting L points is therefore subject to two contradicting effects as L increases with the increase in the size of the available aperture, namely

(i) Improved prediction accuracy in the first predicted point E(M+1) which results in the reduction of the error E(M+L) as long as eqn. (5.70) is satisfied.

(ii) As L increases, the last point in the predicted aperture, when predicting L points, becomes further removed from the first predicted point and therefore more prediction errors accumulate.

Accordingly, the maximum error over the aperture, E(M+L), exhibits a minimum at the value of L which causes the two factors to offset one another. This value of L corresponds to the optimum size of the available aperture for the given expansion ratio required.

The above argument, however, is somewhat idealized since it neglects the effects of the practical limitations on the accuracy of performing the computations necessary for solving the matrix equation and performing the prediction. The weakness of this argument in a practical situation can be seen from the assumption that the error in the first predicted point E(M+1) decreases steadily with the increase in L. Obviously, even with eqn.(5.70) satisfied, the value of this error cannot decrease indefinitely since, sooner or later, it will be limited by the basic accuracy of the machine used to do the calculations. When E(M+1) starts to rise, it would also cause E(M+L) to rise.

The computation accuracy is a function of the word length of the computer and the method used for solving the matrix equation. As the computation accuracy decreases, the accuracy with which the coefficients of the prediction model which represents the signal over the available aperture decreases. From eqn. (5.18), assuming that both the matrix A and the data vector $\underline{h}_0$ are free from noise, the solution for the model coefficients $\underline{a}$ is given by:

$$\underline{a} = A^{-1}\,\underline{h}_o \qquad (5.71)$$

In practice, due to the limited wordlength which causes round-off errors, the calculated inverse matrix will be error. Let this matrix be A_e^{-1} ,

$$A_e^{-1} = A^{-1} + E \qquad (5.72)$$

where E is an error matrix which represents the perturbations in A_e^{-1} due to computation errors. Substituting this matrix in place of A^{-1} in eqn. (5.71) leads to an erroneous solution vector $\underline{a} + \delta\underline{a}$.

$$\underline{a} + \delta\underline{a} = (A^{-1}+E)\,\underline{h}_o \qquad (5.73)$$

From eqns. (5.18), (5.71) and (5.73), an upper bound on the error $\delta\underline{a}$ is given by:

$$\frac{||\delta a||}{||a||} \leqslant ||A||\;||E|| \qquad (5.74)$$

where $||E||$ is the norm of the perturbation matrix E.

If the solution vector $(\underline{a} + \delta\underline{a})$ in eqn. (5.73) is substituted in the original equation, $\underline{h}_o = A\underline{a}$, a new value of the original data vector is obtained as:

$$\underline{h}'_o = A\,A_e^{-1}\;\underline{h}_o \qquad (5.75)$$

since the matrix product $A\,A_e^{-1}$ does not exactly produce the unit matrix, the data calculated from substituting the solution in the original equations will be different from the original data. Therefore, a measure of the effects of the computation accuracy can be obtained by calculating the percentage error E(M) at the last point in the available aperture. Referring to eqn. (4.35) this error is given by:

$$E(M) = 100\,\{[g_1(M)-p_1(M)]^2+[g_2(M)-p_2(M)]^2\}^{\frac{1}{2}}/P_{max} \qquad (5.76)$$

where $P_1(M) + j\,P_2(M)$ is the true complex hologram signal at point M and $g_1(M) + j\,g_2(M)$ is the corresponding signal obtained through substituting the erroneous solution into the original equations and P_{max} is as defined in eqn. (4.34).

Assuming noiseless hologram data, the error in the predicted points outside the available aperture can be considered as the resultant of two components:

(i) Errors which are strictly due to the prediction process.

(ii) Perturbations contributed by the limited computation accuracy which cause errors in the model coefficients.

At short ranges, the errors of the first category are large due to the high spatial frequency content of the hologram and therefore these tend to mask the effects of the second type of errors. However, as the range increases, prediction errors decrease and the effects of the limited computation accuracy become more tangible.

The round-off errors due to the limited wordlength and the limited accuracy of performing the calculations represent uncertainties in the numerical values and therefore can be considered as some form of noise. The effects of the computation accuracy on the matrix solution and the prediction errors are therefore similar to those of noise in general. Accordingly, these effects and their influence on the optimum size of the model and of the available aperture would be different for the square and the triangular models, cf. sec. 5.4.

5.8.1 Effects of Computation Accuracy on the Performance of the Square Polynomial Predictive Model

Eqn. (5.74) indicates that the influence of the limited accuracy in performing the matrix inversion is proportional to the norm of the system matrix A. From eqn. (5.19) for the case of the square polynomial model, as the model order L increases, both the distances d_k and the powers to which these distances are raised increase. This increases the norm $||A||$ and hence the sensitivity to computation errors. Fig. 5.21. shows the errors in the last available point E(M) (due to the limited computation accuracy), in the first predicted point E(M+1), and in the last, Lth, predicted point E(M+L) versus the matrix size L;

L=1 to 15 for 3 different values of the range ; z=40λ, 80λ, 160λ . The object consists of one point located on the hologram axis. The computations were made on the tektronix 4051 minicomputer using the Gauss-Jordan elimination method for solving the matrix equation.

From the figure, the error E(M) increases steadily with the increase in the model size, L. This indicates that the limited computation accuracy using this method for solving the matrix on the Tektronix machine severely limits the performance of the square model for large values of L. The effect of the deterioration in the computation accuracy with the increase in the model size on the overall prediction accuracy depends on the object range. At short ranges, Fig. 5.21a, the relatively large prediction errors dominate and therefore the error in the first point E(M+1) decreases steadily up to L=11. Within this range of the values of L (L=1 to 11) over which the error in the first predicted point E(M+1) continuously decreases, the error in the last predicted point E(M+L) exhibits a shallow trough with a minimum at L=7. Since the performance in this region is governed mainly by prediction errors, this minimum is of the type described at the beginning of this section where the improvement in the prediction accuracy of the first point is offset by the increase in the size of the aperture extension.

From L=11 upwards, the errors due to the limited computation accuracy dominate, therefore 'saturating' the system. In this region, the curve E(M+1) approximately coincides with that of E(M), with the error E(M+L) following closely. Since E(M) is determined by the effects of the computation accuracy only, the performance of the square model for large model sizes will be limited by this accuracy rather than by prediction considerations.

At larger ranges, Figs. 5.21b,c, errors attributed to prediction are lower and therefore the saturation due to the effects of the limited computation accuracy occurs at lower values of L with the increase in range. Although the error E(M+L) still exhibits a minimum, this minimum is not dictated by prediction alone, as in the case of Fig. 5.21a. Since this minimum coincides with the point at which the curve E(M+1) intersects that of E(M), the value of the model size at which this minimum occurs is a function of the range.

The results were repeated for the same imaging geometries using the QR decomposition method [138] for solving the matrix equation on a Nova 2 computer with double precision arithmatics. The corresponding curves are shown in Fig.5.22. Due to the larger wordlength and the use of the QR decomposition, the effects of the computation accuracy are much less pronounced compared to the case of Fig. 5.21. For example, the error E(M) is much smaller in the case of Fig. 5.22 and the rate at which it increases with the increase in the model size is significantly lower. Another sign of improvement is that at the shortest range, Fig. 5.22a the error in the first predicted point E(M+1) shows no sign of bottoming over the range of values L=1 to 15. This is expected when prediction is the predominant factor as discussed at the beginning of this section. The error E(M+L) has a minimum at L=7 which agrees with the value obtained from Fig. 5.21a.

The results in Figs. 5.21 and 5.22 show that for large sizes of the square model, the overall accuracy of the predicted signals is greatly influenced by the accuracy of the machine used to perform the data processing and the method employed for solving the set of equations which describe the model. This influence is increased at large ranges when the small errors due to prediction alone are swamped by errors due to the limited computation accuracy. In this case, the optimum size of the available aperture for a given expansion ratio will therefore depend on the range, the computation accuracy , and the method used for solving the matrix equation.

5.8.2 Effects of Computation Accuracy on the Performance of the Triangular Polynomial Predictive Model

Referring to eqn. (5.55), it is obvious that the triangular form of the polynomial model reduces the norm $||A||$ of the matrix compared to the case of the square model. Therefore, this model should be less sensitive to the effects of the limited computation accuracy. Fig.

5.23 shows the results corresponding to those in Fig. 5.21 when the triangular polynomial model is used for prediction. A comparison between the two figures shows that the size of the triangular model has a negligible effect on the error E(M) which represents the sensitivity to computation accuracy. This indicates that the performance of this model is more robust against the effects of limited computation accuracy regardless of the model size and that it can tolerate a larger amount of round-off errors compared to the case of the square model.

The variations in the prediction error at the edge of the predicted aperture, E(M+L) , with the model size do not exceed 35 dbs for any value of the object range in Fig. 5.23, while such variations can be as high as 195 dbs in the case of the square model. This shows that in the case of the triangular model, operation at an optimum value for the size of the available aperture is not as critical as in the case of the square model.

The results obtained using the Nova 2 with double precision are shown in Fig. 5.24. In spite of the 60 dbs improvement in the computation accuracy as represented by E(M), compared to the results in Fig. 5.23; the prediction errors are almost identical to those obtained using the Tektronix 4051 computer. This indicates that the triangular model is fairly insensitive to the word length used in the calculations and the method employed for obtaining the model coefficients.

5.9 Range Resolution of Apertues Expanded in the Presence of Noise

In section 4.9, the effect of aperture expansion on the resolution in range was discussed for the case of noiseless data where optimum models were used in order to maximize the prediction accuracy. When noisy data are considered the stability of the model with noise must be taken into account and therefore models which are less than optimum from the point of view of prediction may provide better performance in the presence of noise. Fig. 5.25 shows the results corresponding to those in Fig. 4.24 for imaging a 1-point object in range with relative noise level of 10%. Triangular polynomial models were used for both prediction and correction in order to achieve the required stability with noise. The figure indicates that correction is essential for obtaining the resolution improvement corresponding to the expansion of the hologram aperture.

The results corresponding to Fig. 4.25 for resolving two points in range are shown in Fig. 5.26 with 10% noise. The point in the object which is further from the hologram is imaged with 30% error in the value of the range using the predicted hologram and -5% error using the corrected hologram. It should be noted also that the image from the true hologram has a peak at a range value which is+5% in error. The definition of the peak representing this object point in range using the corrected hologram is inferior to that in the case of the noiseless data, Fig. 4.25. This is mainly due to the increased prediction errors with the use of the triangular models which do not fit the hologram function as accurately as the 4/3 hybrid predictive model used in section 4.9.

5.10 Imaging of Objects with Increased Complexity in the Presence of Noise

The effects of increased object complexity and the imaging of quasi-continuous objects were considered in section 4.10. An example for imaging such objects with 10% noise is shown in Fig. 5.27. In this case the object consists of two segments of equal width spaced at 5.6λ , centre to centre, at a distance of 95λ from the hologram. The segment width is 4λ and the linear predictive model is used for prediction. The prediction errors and the reconstructed images are similar to the results obtained for the case of noiseless data in Fig. 4.35 for the same point density of 50 points per wavelength used in both cases.

Another aspect of complexity in the case of 2-point objects was considered in section 4.7 where two points of different values for the radiation strength were imaged using a 5/2 hybrid model to double the aperture size, Fig. 4.16. The corresponding results with 10% noise in the hologram signals are shown in Fig. 5.28 where the linear model was used for prediction. Although the prediction errors are larger compared to the case of the optimum model, the two points are clearly resolved from the predicted hologram. As in Fig. 4.16 for the case of noiseless data, the correct relative radiation strength of 0.8 is not adequately reproduced in the true hologram due to the limited aperture size and the added effect of noise in the signals of that hologram in the case of Fig. 5.28. The apparent improvement in the relative radiation strength in the image reconstructed from the predicted hologram may be due to the offsetting of some of the effects of noise in the true hologram by the prediction errors in the predicted hologram, cf. sec. 5.7.

5.11 Conclusion

Disturbing noise poses a limit on the effectiveness of the prediction technique for hologram expansion described in chapter 4. As in the majority of similar object restoration and hologram estimation techniques,the effects of noise in the measured data are accentuated by the requirement for uniform sampling of the hologram over a limited aperture. Contrary to such techniques, however, this prediction method has the flexibility that the prediction model employed can be tailored to achieve a compromise between high prediction accuracy and adequate stability with noise. In this way, adequate performance in the presence of noise can be achieved through data processing means only; without affecting the data acquisition or image reconstruction phases of the imaging process or the need to compromise important system parameters such as the sample spacing in order to achieve the same objective.

The effects of noise are primarily due to the fact that small errors in the measured data are magnified in the model coefficients when solving the matrix equation describing the model. Therefore, these effects are a function of the norms of both the model matrix and its inverse. Due to uniform sampling, the higher order rows of the square polynomial matrix tend to be identical as the matrix size increases and therefore , this model exhibits large sensitivity to noise. The row dependence can be reduced by modifying the matrix to an upper triangular form which achieves a considerable improvement in stability with noise. However, this is obtained only at the expense of a reduction in the prediction accuracy and it was found that correction using a triangular corrective model is required in order to achieve adequate accuracy when imaging1-point objects.

The optimum hybrid predictive model in the case of noiseless data is dominated by polynomial terms and therefore suffers from the same drawback of noise sensitivity as the square polynomial model. However, the hybrid model assumes the form of a lower triangular matrix when the number of polynomial terms is equal to zero, i.e.

when the model consists of linear terms only. This matrix has similar stability characteristics to the triangular polynomial matrix and leads to adequate prediction accuracies without the requirement for correction when imaging multiple-point objects. Polynomial and linear triangular models have been used to image 1-point and multiple-point objects with relative noise levels as high as 30% when doubling the size of the available aperture. The limit on the amount of noise that can be tolerated appears to be set by the deterioration in the measured hologram rather than the error magnification in the model coefficients.

Another aspect of the practical limitations which affect the prediction accuracies obtainable with this technique is that caused by the limited accuracy of performing the computations during the solution of the matrix equation. The effects of this limitation become more significant at large ranges when the small errors they contribute become comparable with the small prediction errors at such ranges. The influence of the resulting round-off errors on the square and the triangular models is similar to the effect of noise on the performance of the model. Therefore, the triangular model has the added advantage that it is more tolerant to this type of errors. This means that the prediction technique using a model of this form can be implemented using relatively shorter wordlengths and simpler methods for the solution of the matrix equation . This simplifies the memory and hardware requirements and reduces the execution times required and therefore makes it feasible for the resolution enhancement to be achieved using small inexpensive microprocessor systems.

CHAPTER 6

EXPERIMENTAL HOLOGRAPHIC IMAGING SYSTEM

6.1 Introduction

An experimental holographic imaging system has been designed and built for measuring acoustic holograms in water at a frequency of 1 MHz and reconstructing them digitally. This system allows the principle of holographic aperture expansion described in Chapters 4 and 5 to be verified experimentally and helps demonstrate a number of concepts in holographic imaging in general as will be shown in Chapter 7. Additionally, the system demonstrates the capabilities and advantages of using microprocessor technology in achieving the desirable integration between the various aspects of the holographic imaging process which include scanning, data acquisition, hologram reconstruction, and display. It also illustrates the merits of reconstructing the complex hologram digitally in simplifying and speeding up both the acquisition and reconstruction phases of the process.

The system is based on the TMS9900 16-bit microprocessor and employs mechanical scanning of a small hydrophone detector for sampling the hologram aperture. The microprocessor controls the mechanical scanning, the transmission and reception of signals, hologram data acquisition, image reconstruction and the display of both hologram and image information. Moreover, the microprocessor supervises the interface between the system and a number of peripheral devices. The system incorporates the following features, see Fig. 6.1:

(1) Measuring/simulating two-dimensional complex valued holograms and displaying the amplitude and phase information in real-time on a colour TV monitor.

(2) Reconstructing the measured/simulated holograms digitally using the frequency domain backward propagation method which employs an FFT algorithm, and displaying the reconstructed images on the TV monitor.

(3) Obtaining a hardcopy of the displayed holograms and images on an electrostatic picture recorder.

(4) Transferring the measured holographic data to the Tektronix 4051 minicomputer for further processing, including aperture

expansion, and for storing on magnetic tape. Plots may be obtained on an interactive digital plotter.

(5) Recording the measured holographic data on paper tape for further processing of larger holograms on mainframe computers if required.

An overall view of the system components is shown in Fig. 6.2. For the purpose of describing the system, it can be conveniently divided into the following main components:

(i) The microcomputer system
(ii) The hologram acquisition system
(iii) The display system
(iv) The software for hologram simulation and image reconstruction.

These will be covered in the following sections. The electronic circuits of the above subsystems are housed in a 19" rack which is shown in Fig. 6.3.

6.2 The Microcomputer System

The TMS 9900 microprocessor system represents the central processing and control component of the holographic imaging system. It controls the aperture scanning, signal transmission and reception, data acquisition, display, and performs the image reconstruction. Additionally it controls the interface with the Tektronix 4051 mini-computer, the picture recorder, the paper tape punch, and the keyboard, see Fig. 6.1.

The microprocessor system is represented by the functional diagram shown in Fig. 6.4. It uses the TMS 9900 microprocessor as the central processing unit (CPU) [139]. 4k words of Erasable Programmable Read Only Memory (EPROM) are used to store the program codes and the fixed data. 12k words of Random Access Memory (RAM) are used to store the hologram and image data and to provide the working space required during hologram simulation and image reconstruction. A simple monitor program, the TIBUG Monitor [139], resides in a small Read Only Memory (ROM) and allows the inspection/modification of the system memory and registers via the keyboard. It can therefore be used to set up the starting points for executing programs in addition to performing a limited range of debugging operations. An alpha-numeric memory and display interface module allows the TV monitor to be used for displaying the messages exchanged with the monitor through the keyboard.

The microprocessor performs the main external control functions through input and output lines which constitute the Communication Register Unit (CRU) [139]. The signals at these input/output lines are multiplexed/demultiplexed to/from the microprocessor through the CRU bus by decoding the address of the particular line to be enabled. The CRU output lines can be set individually to the required logic levels. Alternatively, a group of such lines can be used to transfer data sequentially from the microprocessor to external devices. Similarly, individual CRU input lines may be tested for low or high logic levels or may be used to transfer data from external devices to the microprocessor. In the experimental holographic imaging system the CRU lines provide the following facilities:

a) RS 232 interface with the system keyboard or a teleprinter/ paper tape punch.

b) Control of the mechanical scanning

c) Interfacing with the hologram acquisition circuits

d) Data and control lines for the display system

e) Data and control lines for the parallel interface with the Tektronix 4051 minicomputer through its General Purpose Interface Bus (GPIB)

f) Data and control lines for the interface with the electrostatic picture recorder.

The microprocessor operates at a clock frequency of 3 MHz and has an extensive instruction set which includes both Multiply and Divide instructions. The 16-bit data bus is particularly useful in implementing the display system since it allows any point in the 256x 256 matrix of the display memory to be accessed by the microprocessor with only one instruction.

6.3 The Hologram Acquisition System

This data acquisition system measures the complex hologram signal over a two-dimensional aperture in the hologram plane. The complex signal is obtained by measuring the amplitude and phase of the signal received by a small hydrophone which is mechanically scanned in a raster pattern over the hologram area. The system employs a novel technique for measuring the phase of the received signal relative to the transmitted signal [140]. A copy of this reference is enclosed in Appendix D. This technique measures the phase digitally by counting the number of pulses of a high frequency clock and has the advantage that the width of the acoustic pulse can be smaller compared to other conventional techniques for measuring the complex hologram, cf. sec. 2.4.2. All of the system functions, including mechanical scanning, data acquisition and storage, and interfacing with other equipment, are performed under the control of the microprocessor system described in section 6.2.

Fig. 6.5 shows a block diagram of the basic components of the hologram acquisition system. The point object to be imaged is simulated by an active source in the form of a focused acoustic transducer which is placed at the bottom of a water tank. The signal received over the horizontal hologram plane is sampled at a number of discrete points in both directions by mechanically scanning a small hydrophone under microprocessor control.

The hydrophone element, see Fig. 6.6, has a diameter of 0.8 mm. At the acoustic frequency of 1 MHz, this corresponds approximately to $\lambda/2$ in water, where λ is the wavelength. Substituting this value as the width of the sampling detector in eqn. (2.32) shows that the maximum spatial frequency over the hologram aperture can be approximately as high as $2/\lambda$ before the transfer function due to the finite sampling aperture starts to influence the spectrum of the sampled hologram, cf. sec. 2.4.1. Assume a total width of 31λ for the hologram aperture. For a point object on the hologram axis and a normally incident reference, the maximum spatial frequency over the aperture is given by eqn. (4.45) as:

$$f_x = \frac{15.5}{z} \tag{6.1}$$

where z is the object range in wavelengths. Since z in the experiments performed is of the order of 100λ or greater, the size of the hydrophone element is small enough for sampling such holograms without significantly distorting the high frequency components.

Referring to Fig. 6.5, a 1 MHz TTL square wave is converted to a bipolar signal and applied to a gated amplifier circuit. At every position of the scanning hydrophone, this circuit receives a gating signal of a width which is equal to the required width of the transmitted pulse. The gated RF signal is connected to an RF power amplifier which drives the acoustic transducer. The transmitted acoustic pulse is received by the hydrophone and is fed to a 2-stage low-noise pre-amplifier which can provide up to 60 dbs of gain, Fig. 6.7. After further amplification in wide-band operational amplifiers, the signal is applied to the amplitude and phase measuring circuits. The phase measuring circuit derives its reference from the same 1 MHz TTL signal used to generate the transmitted pulse. At every position of the scanning hydrophoen, the amplitude and the phase of the received signal relative to the transmitted signal are measured and their values are fed in a digital form into the micro-processor system where they are stored. The amplitude and phase are displayed in real-time on a colour TV monitor as the scan proceeds. At the end of the scan a hardcopy of the two-dimensional hologram can be obtained on the picture recorder. The data can also be fed to the 4051 minicomputer through the GPIB interface.

6.3.1 Mechanical Scanning Arrangement

The hologram acquisition system uses the water tank facility in the departments' offshore technology laboratory. This tank is 2.1 m long, 1.2 m wide,and 0.7 m deep and is fitted with arrangements for mechanically scanning a carriage in the x and y directions using two stepping motors which move the carriage in the two perpendicular directions, Fig. 6.8. The four phases of each of the stepping motors are driven from a motor drive module. This module allows for TTL signals corresponding to the required phasing of each motor to be applied externally, therefore controlling the motion of the motor.

Before a scan is initiated for the first time, the position of the carriage to which the hydrophone is attached is adjusted manually so that the hydrophone coincides with the first point in the hologram area to be scanned. The sample spacing and the number of samples are entered along with other hologram parameters into the microprocessor system memory before the program is executed. After transmission and reception of the acoustic signal is performed at the first sampling point, the x-motor is turned on for a specified number of turns which causes the carriage to move the required increment Δx in the x directions. This is achieved by outputting a sequence of data values stored in the microprocessor system EPROM, which corresponds to the motor phasings, on the CRU output lines which control the motor drive module. The operation is repeated after transmission and reception in the new position until a line scan is completed. At the end of the line scan, the y-motor is rotated to move the carriage in the y direction by an increment Δy. The motion in the x direction is then reversed in direction by outputting the phasing values in the opposite sequence. The process is repeated until the raster scan is completed. At the end of the scan the hydrophone is moved automatically to the first point in the raster in preparation for a new scan. Fig. 6.9 shows a simplified flow chart for the software of the scanning operation.

6.3.2 Transmission and Range Gating Circuits

The transmission and range gating circuits allow the number of cycles in the 1 MHz transmitted acoustic pulse, together with the range delay and the width of the receiver gate,to be set by the microprocessor. A block diagram for these circuits is shown in Fig. 6.10. Before each transmission, the microprocessor loads the required number of cycles in the transmitted pulse into the parallel load inputs of a synchronous counter, C1 in Fig. 6.10. This allows the number of the 1 MHz cycles in the transmitter pulse to have any desired value from 1 to 15. The counter is driven by the 1 MHz TTL clock but is normally inhibited until the arrival of a transmit command from the microprocessor which initiates a sequence for transmission, reception, and data acquisition. Since the microprocessor signal is asynchronous to the 1 MHz clock, this signal is first synchronized to this clock before being used to enable the pulse width counter C1. Therefore, the gating period generated by this counter starts in phase with the 1 MHz clock and lasts for the duration of the required number of cycles. This signal is fed to the gate input of a gating amplifier which passes only the set number of cycles of the bipolar 1 MHz signal to drive the power amplifier. The greatly amplified signal is used to drive the acoustic transmitter.

The synchronised transmission trigger signal is also used to generate a receiver gate which is delayed by a given range delay from the transmitted pulse. The required values for both the range delay and the width of the receiver gate are loaded into the corresponding counters before each transmission. The carry signal from the range delay counter, C2, enables the receiver gate counter C3. To allow for accurate setting of both the range delay and width of the receiver gate, counters C2 and C3 which generate these signals are driven from a 10 MHz clock from which the 1 MHz signal is derived. The circuit is wired-up such that both intervals can be selected to within 0.2 μsec. At 1 MHz this corresponds to a distance of 0.2 λ of one way travel of the acoustic signal. The receiver gate is used to enable the receiver circuits so that they receive only the direct first-arrival signal that reaches the hydrophone from the acoustic source and therefore ignore any other multipath arrivals due to reflections at the water surface or the walls of the tank.

6.3.3 Phase and Amplitude Acquisition Circuits

A block diagram of the circuits used for the acquisition of both the amplitude and phase of the hologram signal is shown in Fig. 6.11. The received signal from the hydrophone is amplified and connected to the phase acquisition circuits. At the front end of these circuits, the signal is connected to an ECL threshold detector which is enabled by the receiver gate described in section 6.3.2. In order to limit the processing of the received signal the first received pulse only and ignore any other nearby arrivals that follow it, an additional processing gate is used. This gate starts at the point in time when the received signal exceeds the threshold level for the first time and is as wide as the gate used to generate the transmitted pulse. This gate is generated only once for every transmission and therefore guarantees that any further signals exceeding the threshold level will be ignored in all subsequent processing even if these occur within the receiver gate. The processing gate is shown in Fig. 6.12 in relationship to other gating signals in the transmission, reception, and data acquisition and display sequence. The use of this gate has the following advantages :

(i) Reducing the tolerance on the width of the receiver gate. The receiver gate can now be of a fixed width which is chosen to be large enough to accommodate variations in the arrival time of the received signals over the hologram aperture. In this case there is no need for calculating a value for the range delay at every point in the scan.

(ii) This method is useful in ignoring the effects of any ringing in the transmitted acoustic pulse. This ringing is substantial when the transducer is air-backed, which is the case in the focused transducer used for obtaining the experimental results.

The processing gate enables an ECL zero crossing detector which converts the received signal into a square wave lasting for the duration of the gate. This square wave signal is fed to a phase measuring circuit together with the reference 1 MHz signal which is used to generate the transmitted pulse. This circuit measures and stores the value of the phase between the received signal and the 1 MHz reference signal at every cycle of the received acoustic pulse by counting the number of pulses of a high frequency clock for the duration of the phase difference between the two signals. To be able to make one measurement every cycle, the cycle is viewed as one that varies from plus to minus 180^{o}. By using only one half of the cycle for the actual phase measurement, the circuit is free during the other half to store the results and to clear the phase counter for the next measurement in the new cycle. The basic phase measuring circuit is described in more detail in Appendix B and in [140], a copy of which is enclosed in Appendix D.

Referring to Fig. 6.11 for the phase acquisition circuits in the hologram acquisition system, the values of the phase for every cycle of the received signal are stored in a small fast RAM memory. The RAM address is derived from a multiplexer which selects the address either from a cycle counter when the phase values are written into the RAM during signal reception or from the microprocessor during the post-reception data processing. The leading edge of the selected (leading)input; the 1 MHz reference signal or the received signal, is used to increment the cycle counter for every cycle of the received signal. This counter is reset to zero between transmissions. Since the value of the threshold level is always greater than zero in order to prevent the system from being triggered by noise, the first cycle of the received acoustic pulse is ignored. This avoids erroneous phase measurement due to the missing part at the beginning of this cycle.

The number of pulses of a 100 MHz clock are counted for the duration of the phase delay signal by the phase counter. The trailing edge of the selected leading input signal, which always comes after/ at the end of the phase delay signal, generates a narrow pulse from monostable 1 which strobes the data corresponding to the terminal phase count, together with the phase sign bit, into the phase RAM. Once this data has been stored, the phase counter can be cleared in preparation for the new cycle. This is accomplished by another narrow pulse which is generated by monostable 2. The process is repeated for the remaining cycles of the received signal which lie within the processing gate.

With a 100 MHz clock, the relative phase between the two 1 MHz signals can be measured to within $\pm 1.8^{o}$. Higher measurement accuracy can be obtained by using higher frequency clocks. Emitter Coupled Logic (ECL) circuits with a typical gate propagation delay of 2 nsec are used for gating and multiplexing the input signals. This is necessary in order to minimize the phase errors due to the differential propagation delays when the phase delay signal is obtained. The first stages of the phase counter are also of the ECL type to suit the high frequency clock. The 100 MHz clock is obtained from a crystal controlled ECL oscillator which is mounted, together with all the ECL components,on a special wire-wrap board, Fig. 6.13.

The amplitude acquisition circuits are shown at the top right corner of the schematic block diagram of Fig. 6.11. The amplified received signal is applied to a peak detector circuit whose DC output is connected to an analogue to digital (A/D) converter. The peak detector, which is normally inhibited and its output set to zero, is enabled at the beginning of the processing gate. At the end of the analogue to digital conversion, the resulting digital number which corresponds to the detected peak amplitude is latched to the 8-bit outputs of the converter, ready for accessing by the microcomputer. The output pulse from the A/D converter which marks the end of conversion resets the peak detector to zero in preparation for the next transmission sequence. This pulse also indicates the end of the hardware signal acquisition during reception and signals the microprocessor to start the post-reception data acquisition.

6.3.4 Post-Reception Data Processing and Display

Software post-reception data processing commences after the acoustic pulse is received, the phase of each of its cycles stored in the phase RAM, and its peak amplitude detected, converted to a digital number, and latched at the output of the A/D converter. The address of the phase RAM is then switched so that it comes from the microcomputer system. The microprocessor transfers the stored phase data (in magnitude and sign format) through CRU lines to an area in the microcomputer RAM. Each value of the phase clock count, C_i, stored in the phase RAM covers the range:

$$-50 \leqslant C_i < +50 \quad ; \quad i=2,3,\ldots, N_c \tag{6.2}$$

where N_c is the number of cycles in the transmitted pulse. This corresponds to a range of phase values given by:

$$-180^o \leqslant \theta_i < +180^o \quad ; \quad i=2,3,\ldots,N_c \tag{6.3}$$

Each value of the phase count is converted to a positive number C_i' according to the sign of C_i

$$C_i' = \begin{cases} C_i & C_i \geqslant 0 \\ 100-|C_i| & C_i < 0 \end{cases} \tag{6.4}$$

The range of C_i' is therefore given by:

$$0 \leqslant C_i' < 100 \tag{6.5}$$

which corresponds to the phase range

$$0 \leqslant \theta_i' < 360^o \tag{6.6}$$

After converting the values of the 0 to 180^o phase counts into their equivalent 0 to 360^o positive counts, the microprocessor calculates the average of these values as:

$$C'_{av} = \frac{1}{N_c-1} \sum_{i=2}^{N_c} C_i' \tag{6.7}$$

The average phase count is stored together with the 8-bit value of the peak amplitude in the two bytes of the 16-bit word in a block of momory within the microprocessor system RAM which is reserved for the storage of the hologram data. This amplitude and phase data is stored in a word whose location within this memory block corresponds to the position of the scanning hydrophone in space within the hologram aperture. The most significant 3 bits of the peak amplitude and the average phase count are fed into the display memory and displayed in real-time on a colour TV monitor as the scan is performed.

After the amplitude and phase data are written into the display memory the hydrophone is moved to the next point in the scan and a new cycle of transmission, reception, and data acquisition is repeated, see Fig. 6.12 . A flow chart for the software of the post-reception data processing is shown in Fig. 6.14.

6.3.5 Interface with the Electrostatic Picture Recorder

Hardcopies of measured or simulated holograms and reconstructed images can be obtained using a modified Muirhead electrostatic picture recorder type K-580-A. The recorder is originally intended for recording picture information transmitted over a telephone line by a complementary remote transmitter. For the particular application at hand, the most significant 4 bits of the digital information to be recorded are converted to an analog form to provide the 16-gray level video signal of the picture. This analog signal replaces the information which is normally derived from the modulated carrier signal that would be received in usual operation. A TTL enable signal is also required in addition to the video signal.

At the outset of recording a picture, the video signal must contain an initial start-up pulse followed by a number of phasing pulses which are used to synchronize the motion of the belt which carries the recording styli. The microprocessor is programmed to generate the required phasing pulses at the correct timing. To ensure synchronism between the recording mechanism and the retrieval of data from the microcomputer memory, a synchronising pulse which occurs at the beginning of each picture line is derived from the recorder. The microprocessor waits for this pulse to arrive before the data corresponding to a new picture line is sent to the recorder.

A simplified flow chart for the microprocessor software for the recording of hardcopies is shown in Fig. 6.15. The information corresponding to each picture sample remains fixed for a period of time which depends on the number of samples in the picture. To represent each sample with a square picture element, the same information corresponding to one line of the picture data is repeated for a number of recorder lines. In order to retain the appearance of a sampled hologram or image, a spacing is left blank between the picture elements both in the x and y directions by setting the picture data to produce a 'white' video signal. The blank spacing between the sampled picture elements is determined by the spacing between the hologram samples in space. Two examples of the hardcopies obtained in this way are shown in Fig. 6.16 for two simulated holograms of a 1-point object.

6.3.6 GPIB Interface with the Tektronix 4051 Minicomputer

The Tektronix 4051 minicomputer was used to perform aperture expansion and image reconstruction on one-dimensional line holograms obtained with the hologram acquisition system. It also allowed for plots of the holograms and images to be obtained on an interactive digital plotter. The General Purpose Interface Bus (GPIB) facility on the 4051 was utilized to transfer the hologram parameters and data from the microcomputer system to the minicomputer. This is a byte-serial, bit parallel method of interface [141] in which the data is transferred in groups of bytes on an 8-bit data bus. The GPIB has another 8 lines for interface control signals. The microcomputer system uses CRU input/output lines to provide the required data and control lines.

Fig. 6.17 shows a simplified flow chart for the software on both the microcomputer and the 4051 sides of the interface. The microprocessor initiates the data transfer routine by activating a service request from the 4051 . The latter responds by sending a device address which sets the microcomputer system as a talker. The microprocessor then reads the address and, if it agrees with a pre-assigned value, the data is transmitted byte by byte by the microprocessor and read by the minicomputer.

6.4 The Colour Display System

A versatile colour graphics display system has been developed for displaying both simulated and measured two-dimensional holograms and reconstructed images. In this system, the microprocessor is programmed to write the picture information into a dedicated display memory which is organized to provide a 256x256 picture resolution. The display memory is 3-bit wide, which allows for the picture information to be displayed in 8 colours. The microprocessor accesses this memory as a memory mapped device and can write into any of its locations with only a single instruction. Transparant access for writing into the display memory is available during the field blanking intervals in order to avoid any flickering of the display.

The picture is displayed on a modified TV monitor which allows colour information to be supplied in the form of 3 TTL lines; corresponding to the red, green, and blue colour guns. The synchronizing and blanking signals are supplied separately, also in TTL format. When the TV monitor is not used for displaying pictures, the video signals to the monitor can be switched to come from an alpha-numeric display interface module. The TV monitor, together with a keyboard, can then be used as a terminal for communicating with the TIBUG Monitor which is resident in the microcomputer system. This is used for the initial setting of the parameters for the imaging system and the starting points for executing the various programs. This arrangement is also useful for carrying out a number of basic debugging operations.

6.4.1 General Description

A schematic diagram of the colour display system is shown in Fig. 6.18. The system consists of a synchro generator module, a timing and control module, in addition to the display memory cards and two line buffer memories each of which can store one picture line.

The synchro generator module generates the horizontal and vertical synchronizing and blanking pulses, together with a number of horizontal and vertical clock signals. These clocks are used to drive counters whose parallel outputs provide the addresses for the display and line buffer memories when the contents of these memories are displayed on the TV screen.

The timing and control module generates all the signals which control the reading and writing operations of the various memories in the system. The main function performed by this module is the multiplexing of the address bus of the display and line buffer memories. For example, when the microprocessor requires to write picture data into the display memory, a control signal from the microprocessor routes the data bus of the microprocessor to the column and row address lines of the display memory. After the picture information is written into the memory, it can be read by deriving the address from fast counters which are driven in synchronism with the sync and blanking signals from the synchro generator module, see Fig. 6.19. The timing and control module is controlled by CRU lines from the microprocessor.

The display memory consists of 3 boards each containing 16 static RAM devices each having a storage capacity of 4kx1 bits. The 16 devices are wired up to form a matrix array of 256x256 storage locations in which one bit of the picture colour information is stored. The 3 pages of the display memory are always addressed simultaneously. During the display write mode, the address comes from the microprocessor. In this mode, the Write Enable ($\overline{WE}$) of all the devices in the display memory is activated by decading the value of the dummy address assigned to this memory mapped operation. During the read mode, the display memory locations are scanned in real time in synchronism with the TV sync signals and the stored picture is displayed on the screen.

Two line buffer memories are used, each buffer can store 256x3 bits of information corresponding to one picture line. The buffers are two bipolar fast memories which are alternately loaded and unloaded with the colour data corresponding to successive picture lines stored in the display memory. While one buffer is being read into the display, the data corresponding to the next line down the picture is written into the other buffer and the two buffers alternately exchange their roles. A multiplexer selects the signals from the buffer being read and routes it to the colour signal inputs on the TV monitor. The 'Select' signal to this multiplexer is a square wave obtained by dividing the vertical (line) clock by 2. The reason for the use of the line buffers is treated in the following section.

6.4.2 Economic Constraints on the Display Memory

Due to the large size of the display memory, its cost would be high and therefore it would be advantageous to reduce the speed requirements on this memory. This factor was more significant at the time when the display system was designed a few years ago than it is today due to the recent increase in the availability of larger and faster memories at a lower cost. Writing into the display memory is limited by the speed of the microprocessor and can be slowed down even further by incorporating 'wait' cycles. Reading the display memory, however, must be performed fast enough so that the stored picture data is fed to the display at a rate which matches the line and field scan times of the TV raster. If the simple and straight forward approach is adopted, where the picture data is read directly from the memory to the display, then 256 memory locations must be read during each TV line. Assuming an active line period of 52 μsec, the maximum limit on the duration of the memory read cycle is approximately 200 nsec. Although this does not call for particularly fast memory devices, especially by today's standards, the large size of the memory required would make this approach economically unfavourable.

To reduce the speed requirements on the display memory, a technique is employed which has been reported in the literature in connection with digital scan converters [142]. With this technique, each four devices which comprise the rows of the 4x4 device matrix in the display memory are accessed simultaneously to output a whole picture line into the display screen. Therefore, each device can operate at a much slower pace than if they were to be read in tandom as in the case described in the previous paragraph. The picture lines produced in this way are stored in a fast, but small, RAM which has enough capacity to hold only one picture line. This line buffer memory consists of 4 separate devices which are loaded simultaneously and then read out sequentially to the display. However, in order to ensure continuous transfer of picture rows into the display, another identical line-buffer RAM is required. The two line-buffers are wired-up such that while one is being loaded by sections of a picture line , the full length of the other is read sequentially to transfer the preceding picture line to the display screen. The two buffer exchange their roles alternately every picture line.

The arrangement for implementing this approach is shown in Fig. 6.20 for one page of the display memory in which one bit of the picture data is stored. The scheme is identical for the other two pages of memory. The display memory consists of sixteen 4kx1 static RAM devices arranged in 4 rows and 4 columns, each providing a 64x64 matrix of memory locations. Each of the line-buffer memories consist of four 64x1 fast bipolar RAM devices. The data outputs of each four column devices in the display memory are joined together and connected to the data inputs on the corresponding two devices in the two line-buffers. The data input to the display memory is common to all of its 16 devices.

When the microprocessor writes the picture information into the display memory, the data bus of the microprocessor provides the address for the picture pixel to be accessed for writing. The Chip Enable ($\overline{CE}$) inputs on the devices are wired up so that only one of the 16 memory devices can be accessed by the microprocessor at any time. In the display read mode, the full length of the 8-bit row address for the display memory as a whole is derived from counters driven by the vertical (line) clock as shown in Fig. 6.19 to select the required picture line. Similarly, the least significant 6 bits of the column address are obtained from the counters driven by the horizontal clock φ_1. These address the pixel along the selected picture line within each of the four row devices which contain the selected row. The remaining most significant 2 bits of the column address are ignored in such a way that all these row devices are enabled simultaneously. Therefore, 4 symmetrical locations along each row are accessed simultaneously. The four devices in the line-buffer to be loaded are also enabled and their address is fed from the same horizontal clock φ_1.

The data is loaded from each of the four display memory row devices into the corresponding device in the line buffer. Since four data transfers are performed at the same time, only 64 transfers are required during the whole of the active line interval in order to load the 256-element picture line into the buffer memory. Therefore, the period of the horizontal clock can be as long as 0.8 μsec for an active line interval of 51.2 μsec. This figure also represents the maximum limit on the access time of the display memory devices. A reduction in the speed requirement for the display memory by a factor of four can therefore be achieved with this approach.

At the end of the line interval, the line-buffer contains a full picture line. During the next line interval, the four devices in this line-buffer are enabled sequentially and their contents are dumped onto the display screen. Since 256 data transfers are performed during the active line interval in this case, the clock producing the address during the reading of the line-buffer memory must have a period of approximately 0.2 μsec . Therefore two different clock rates are used to provide the addresses of this memory in the read and write modes as shown in Fig. 6.21. The read cycle of line-buffer RAM devices should be 0.2 μsec maximum. However, since the two buffer memories are small, they can be chosen to be fairly fast without any significant economic constraints.

6.4.3 <u>The Display Software</u>

To ensure simple and speedy access for the microprocessor to the display memory when writing the picture information, this memory is considered by the microprocessor as a memory mapped output device. In this mode of operation, parallel data transfer is achieved through the microprocessor data bus which is connected in this application to the address lines of the display memory during the display write mode.

As a memory mapped device, the display memory is given a fixed memory address, e.g. >400 where '>' denotes hexadecimal number representation. Whenever the colour data corresponding to any of the 256x256 picture pixels is to be written into the corresponding location in the display memory, this data is first set on the 3 CRU lines which are connected to the data inputs of this memory. The 8 bits of data corresponding to the x address of the required pixel are loaded into the least significant byte of a microprocessor workspace register [139], e.g. R1. Similarly, the y address is loaded into the most significant byte of the same register. This address information is then transferred to the address bus of the display memory by executing the instruction

```
MOV   R1, @ > 400
```

which transfers the contents of R1 to the dummy memory location whose address is > 400.

The last operation carried out by the microprocessor when executing this instruction is a memory write cycle. During this cycle, the microprocessor address bus has the value >400 while its data bus has the contents of register R1, i.e. the display memory address. To enable the writing of the colour data set at the memory inputs at this particular address, the Write Enable ($\overline{WE}$) input to the display memory is activated at this point in time by decoding the fixed memory mapped address (>400) as shown in Fig. 6.18. Additional memory and data bus control signals from the microprocessor are also decoded in order to prevent inadvertent accessing of the display memory and to ensure that the timing requirements for the write cycle of the display memory are fulfilled.

There are two modes of writing the picture information into the display memory. These can be labelled as the bulk writing and the transparant (or interrupted) writing modes. In the first mode, the time for writing into the memory is set by the microprocessor. During this writing time, the display memory is accessed only by the microprocessor and therefore its contents cannot be displayed until the writing is finished and the microprocessor releases its address lines and switches them to the outputs of the counters of the TV raster. This method is useful in writing large amounts of picture data and in situations where the display of the information in real-time is not required.

In the second mode, writing is performed during the field blanking pulses and is therefore achieved without loss of synchronism of the displayed picture. In this case, the field blanking signal is used as the Read/Write control for the display memory. The same signal is used as a CRU input to be continuously tested by the microprocessor. The microprocessor waits for the blanking pulse to arrive in order to transfer the required display data to the memory. The field blanking lasts for approximately 2 msec and therefore this method can be used to write simple patterns with the execution time not exceeding the duration of the blanking pulse, although in principle transparant writing of more complex patterns may be extended over a number of picture fields. This writing mode is used to display the amplitude and phase data in real-time as the hologram aperture is scanned. cf. sec. 6.3.4. This mode can also be selected by the microprocessor following a bulk writing mode to allow for making small modifications to a picture made up of a large amount of data or for adding variable markings or labels onto it.

The colour display system has been mainly used for displaying the measured and simulated holograms together with the images reconstructed from them; samples of the picturs obtained are given in Chapter 7. This however does not fully utilize the flexibility of the system for displaying colour graphics. In this application the microprocessor writes the graphics information by calculating the x and y coordinates of the points describing the curves to be drawn and accessing the display memory locations corresponding to these coordinates. In a practical holographic imaging system, the graphics facility would be useful in displaying the outline of a test piece for example on which the information on the location and shape of the imaged object can be superimposed.

6.5 Software for Hologram Simulation and Image Reconstruction

A number of microprocessor programs have been written and used to simulate two-dimensional holograms for simple point objects and to reconstruct images from both simulated and measured holograms. The hologram simulation is based on the approach outlined in section 4.3. Image reconstruction uses the backward propagation (frequency domain) method discussed in section 2.5.2. To ensure adequate dynamic range and computation accuracy, all the calculations are performed using floating point format. In this format, the number is represented by two memory words, each 16 bit wide. Performing floating point arithmatics on the TMS9900 microprocessor is simplified by using a number of utility routines supplied by the manufacturer. These routines convert a number from an integer into a floating point format and vice versa and perform addition, subtraction, multiplication, and division in floating point arithmatics. It should be noted that no special effort has been made to optimize the software for speed or economy in utilizing the memory space available. Nevertheless, holograms made up of 32x32 samples are reconstructed and the resulting images displayed in less than 3 minutes.

6.5.1 General Description

Fig. 6.22 shows a network representation of the subroutines used for hologram measurement/simulation and image reconstruction and the display of both holograms and images. The routines are shown in their hierarchal order. Following is a brief description of each of the main routines:

HMES : Measures amplitude and phase of hologram (uses integer representation)and displays the results in real-time as the hologram aperture is scanned mechanically, cf. sec. 6.3.

MOTOR : Outputs the appropriate motor phasing signals under the control of HMES.

CONVERT : Converts the amplitude and phase count information measured by HMES in the integer form into real and imaginary components in floating point format in preparation for image reconstruction.

HSIM : Simulates the hologram of a discrete point object containing up to 3 points using floating point arithmatics. Uses eqns. (4.25) to (4.27) for hologram simulation. Calculates and displays the amplitude and phase of the simulated hologram signal.

HRCON : Reconstructs the complex hologram data to obtain an image at any given value of the reconstruction distance. Uses the backward propagation (frequency domain) method, cf. sec. 2.5.2. This method includes 2 two-dimensional FFT transforms and the use of the FOCUS routine.

FOCUS : Multiplies the hologram data in the frequency domain by a focusing function. This function is determined by the reconstruction distance at which the image is obtained, cf. eqn (2.63).

DFT : Performs a two-dimensional FFT in the forward or inverse directions using the Cooley and Tukey algorithm. Uses the SFT routine on the rows of the original data and then on the columns of the resulting data.

SFT : Performs a one-dimensional FFT in the forward or inverse direction on a row or a column of data.

NDIS : Calculates the intensity of the hologram spectrum or the reconstructed image from the complex amplitude, normalizes the intensity distribution relative to its peak value, and displays the results on the TV screen.

HCOPY : Produces a hardcopy of the hologram, its spectrum; and the image intensity on the picture recoder, cf. sec. 6.3.5.

In addition to the floating point utility routines, the main routines described above use also a number of mathematical functions routines to calculate the functions $\sqrt{X}$, $\sin(\theta)$, $\cos(\theta)$, $\sin^{-1}(\theta)$ and $\cos^{-1}(\theta)$ in floating point . These routines are described in section 6.5.3.

6.5.2 The Two-Dimensional FFT Algorithm (DFT)

Referring to eqn. (2.65), the two dimensional discrete Fourier transform of the complex function $u(p\Delta x, q\Delta y)$ is given by:

$$U(m\Delta f_x,n\Delta f_y) = \sum_{q=0}^{N-1} \sum_{p=0}^{N-1} u(p\Delta x,q\Delta y) \exp\{ -j2\pi[(p\Delta x)(m\Delta f_x) + (q\Delta y)(n\Delta f_y)]\} \quad ; m,n=0,1,\ldots,N-1 \tag{6.8}$$

where $u(p\Delta x,q\Delta y)$ and $U(m\Delta f_x,n\Delta f_y)$ are the discrete functions in the space domain and the spatial frequency domain respectively. N is the number of data samples in each coordinate direction in both domains and is assumed to be a power of 2. p,q are the indices of the data sample in the space domain in the x and y directions respectively while m,n are the indices in the f_x and f_y directions in the frequency domain. This equation can be written in the form

$$U(m\Delta f_x,n\Delta f_y) = \sum_{q=0}^{N-1} \left[\sum_{p=0}^{N-1} u(p\Delta x,q\Delta y) \exp\{-j2\pi[(p\Delta x)(m\Delta f_x)]\} \right]_{m=0,1,\ldots,N-1} \exp\{ -j2\pi[(q\Delta y)(n\Delta f_y)]\}_{n=0,1,\ldots,N-1} \tag{6.9}$$

where the expression between the large square brackets represents one dimensional Fourier transforms on the rows of data while the summation outside these brackets operates on the columns of the resulting rows. This shows that the two-dimensional transform can be performed by applying one-dimensional transforms to the rows of data and then on the columns of the transformed rows. The one-dimensional transforms are performed using the SFT routine.

The SFT routine performs the FFT on N pairs of real and imaginary floating point data which are stored in a buffer area of 4N words within the microprocessor system RAM. At the end of the transform, the resulting data replaces the original data in this area of buffer memory. The microprocessor program uses a number of mathematical function routines for determining the exponential functions in eqn. (6.9). A listing of the SFT program and the associated routines in assembly code is given in Appendix C.

A flow chart for the two-dimensional DFT routine is shown in Fig. 6.23. The algorithm operates on a matrix of NxN pairs of real and imaginary data stored in the system RAM, where N is a power of 2. Rows of data are sequentially loaded into the SFT buffer memory, Fourier transformed, and the resulting data fed back to replace the original data in the system memory. After all the data rows have been transformed, the process is repeated on the columns of the transformed data.

6.5.3 Mathematical Functions Routines in Floating Point Format

The microprocessor programs used for hologram simulation and reconstruction require a number of mathematical functions to be determined with adequate accuracy at a reasonable execution time and memory storage space. This requirement is fulfilled using four routines; SQRT, SINTAB, SINCOS and ANGL. These are shown in the software network in Fig. 6.22 and will be briefly described below.

SQRT:

This routine adopts an iterative method for generating the square root of a floating point positive number using the Newton-Raphson iteration. As a first approximation, the square root R of the number X is determined as the number of odd numbers (starting from 1) whose sum does not exceed X. More accurate values of the root are obtained from the iterative equation:

$$R_{k+1}=0.5[R_k+(X/R_k)] \qquad (6.10)$$

where R_{k+1} is the new root and R_k is the previous root. The routine used employs five iterations.

SINTAB:

This routine contains a look-up table of 91 values of sin(θ) where θ is an integer value in degrees in the range $0 \leqslant \theta \leqslant 90^o$ in steps of 1^o. This range of values is sufficient to determine both sin(θ) and cos(θ) in the range $0 \leqslant x < 360^o$. The routine is used by both SINCOS and ANGL routines.

SINCOS:

This routine is used in conjunction with the SINTAB routine to evaluate sin(θ) or cos(θ) where θ is a real number in degrees in floating point format. If θ' is the nearest integer value which is less than θ and α is the difference (θ-θ') then:

$$\sin(\theta) = \sin(\theta'+\alpha)$$

$$= \sin\theta' \cos\alpha + \sin\alpha \cos\theta' \qquad (6.11)$$

From the definition of θ',

$$0 \leqslant \alpha < 1^{o} \qquad (6.12)$$

Applying the following approximations for small angles

$$\cos\alpha \approx 1$$
$$\sin\alpha \approx \alpha \frac{\pi}{180} \qquad (6.13)$$

reduces eqn. (6.11) to:

$$\sin(\theta) \approx \sin(\theta') + \frac{\alpha\pi}{180} \cos(\theta') \qquad (6.14)$$

Eqn. (6.14) shows that the sine of the real valued angle can be approximately determined with the knowledge of α and the values of the sine and cosine of the nearest lower integer value θ'. These values can be easily obtained from the look-up table in SINTAB. Similarly the cosine function can be evaluated as a sine function using the relationship:

$$\cos(\theta) = \sin(\theta+90^{o}) \qquad (6.15)$$

With the approximations used in this algorithm, the values calculated for the sine and cosine are accurate to the third decimal place. This was considered adequate for the application although the method is capable of achieving better accuracy by using more terms in the series expansion of cosα and sinα beyond the first term in each case as used in eqn. (6.13).

ANGL:

Used in conjunction with the SINTAB routine, this routine determines the nearest integer value θ ; $0 \leqslant \theta < 360^{0}$ to the argument of the complex number $z = x+jy$, given the values of the real and imaginary parts x and y. The sine of the equivalent argument θ_r which is restricted to the first quarter $0 \leqslant \theta_r \leqslant 90^{0}$ is calculated as $|y|/|z|$. Using the look-up table in SINTAB, this sine value is used to determine the nearest integer value to the angle θ_r . The quarter in which the actual argument lies is determined from the signs of both x and y and therefore the integer value θ of the actual argument can be calculated from θ_r.

CHAPTER 7

EXPERIMENTAL RESULTS

7.1 Introduction

Experimental results have been obtained to verify the aperture expansion technique described in Chapters 4 and 5 using the experimental holographic imaging system described in Chapter 6. Both one and two-dimensional holograms of a point object were measured and reconstructed digitally. The line holograms were expanded by a factor of two and the images reconstructed from both small, predicted, corrected, and true apertures (where applicable) using the algorithms described in section 2.5.2. The processing of line holograms was performed on the Tektronix 4051 minicomputer while image reconstruction from two-dimensional holograms was implemented on the microcomputer system using the software routines described in section 6.5. A number of two-dimensional holograms have been simulated and reconstructed for a variety of discrete point objects which help demonstrate a number of criteria of holographic imaging. Pictures of the two-dimensional holograms,their spectra, and the reconstructed images were obtained by photographing the colour display on the TV monitor, cf. sec. 6.3 and 6.5.

7.2 Experimental Set-up and the Problem of Hologram Tilt

The active point-object used in the following experiments was simulated using a focused acoustic transducer which is made up of a spherical shell and has the dimensions shown in Fig. 7.1. The transducer is positioned at the bottom of the water tank and the hologram area on top of the transducer is mechanically scanned in a raster format and sampled with a small hydrophone. At every position of the sampling hydrophone, a burst of 1 MHz signal is transmitted from the transducer and the amplitude and phase of the received signal are measured. The measured holograms are displayed in real-time on a colour TV monitor. The hologram data corresponding to any line scan can be transferred to the Tektronix 4051 for further processing, see Fig. 6.5.

7.2.1 Effect of Tilt on Hologram Recording

The initial tests were performed with the transducer attached to a holder and placed directly on the bottom of the water tank. Two anomalies were observed in the first holograms recorded originally with this arrangement, namely:

(i) Higher spatial frequencies than expected were encountered and signs of under-sampling were evident in spite of the fact that the samples were chosen close enough to prevent any aliasing in the theoretically expected holograms for the given aperture size and object range.

(ii) The holograms were not symmetrical about the central fringe; with fringes cramped on one side of the hologram and sparsely spaced on the other.

These two effects are clearly demonstrated in the photograph in Fig. 7.2. In a line hologram, this may lead to introducing an additional phase fringe at the centre of the hologram where the phase function should not normally exhibit much variations in this region opposite to the point object. This is shown in the phase plot in Fig. 7.3 for a measured hologram.

These effects can be attributed to the presence of a relative tilt angle between the hologram plane and the bottom of the water tank. Referring to Fig. 7.4, Fig. 7.4a shows the case of proper alignment in which both planes are parallel while in Fig. 7.4b a tilt is introduced in the hologram plane. Consider a line hologram in the vertical plane containing the point object O. The rate of variations in the phase of the hologram signal along the line scan is proportional to the rate of variations in the length of the vector OH which joins the point object to point H on the hologram. Comparing Figs. 7.4a and 7.4b, it can be seen that these variations, and accordingly the fringe spacing and the spatial frequency on the hologram, will not be symmetrical on both sides of the tilted hologram. For a positive tilt angle as shown in the figure, higher spatial frequencies exist on the R.H. side of the hologram. Those frequencies would also be higher than the frequencies in the case of correct alignment, hence the undue symptoms of undersampling. While the phase is almost constant in the central region of the hologram in Fig. 7.4a., the phase variations at the centre of the tilted hologram give rise to the additional phase fringe as shown in the plot of Fig. 7.3.

The effect of this positioning error was numerically simulated by introducing a tilt angle in the hologram plane and the results are shown in Fig. 7.5 for the case of 0°, $+5^{\circ}$, and -2.5°. It should be noted that the effect of the tilt in the hologram is different from that caused by a shift in the position of the point object. Fig. 7.6a shows the hologram phase with a 4λ shift in the x direction parallel to the hologram line and no tilt in the hologram plane. In this case the phase pattern merely shifts with the object. Fig 7.6b simulates a shift of 8λ in the object position in the y direction normal to the hologram line. Although it is possible that a new central fringe could be generated in this case for large shifts in the y direction, the hologram is different from that in the case of a tilted hologram since symmetry is always maintained in both halves of the hologram.

7.2.2 Effect of Tilt on Image Reconstruction

Referring to Fig. 7.7, assume a hologram tilted with an angle θ.

For a point object, the signal at a point on the hologram at distance x from the centre is:

$$H(x) = \frac{1}{r'} \exp(-jKr') \qquad (7.1)$$

where K is the wavenumber, $K = 2\pi/\lambda$, and r' is the distance between the object point and the hologram point. For small values of θ and large values of r':

$$r' \approx \begin{cases} r & \text{in the denominator of eqn. (7.1)} \\ r + x\sin\theta & \text{in the exponent of eqn. (7.1)} \end{cases} \qquad (7.2)$$

where r is the distance between the point object and the corresponding hologram point assuming no tilt in the hologram plane. Substituting eqn. (7.2) into eqn. (7.1) yields:

$$H(x) = [\frac{1}{r} \exp -jKr] \exp(-jKx\sin\theta) \qquad (7.3)$$

The terms between the square brackets represent the hologram signal with zero tilt while the exponent outside these brackets represents a plane wave incident at an angle θ with the untilted hologram. Therefore, eqn. (7.3) shows that the effect of a tilt in the hologram plane is equivalent to that of introducing an offset reference wave into the hologram signal that would be received with no tilt in the hologram. When the complex hologram is measured on a hologram tilted with angle θ the resulting signal corresponds to that of a hologram with zero tilt mixed with an off-axis reference wave which makes an angle θ with the normal to the untilted hologram. In fact, tilting the hologram plane has been reported in the literature as a means of providing an off-axis reference [143],[54]. Since the reconstruction algorithm employed, cf. sec. 2.5.2., assumes that the hologram data corresponds to the complex hologram, which is equivalent to the use of an on-axis reference, the tilt in the hologram plane introduces errors in the reconstructed image.

A simplified approach to investigate the type of errors in the reconstructed image due to a hologram tilt is described by referring to Fig. 7.7. The point O to be imaged lies on the vertical line passing through the hologram centre at distance z from the hologram centre. The reconstruction algorithm assumes that the object plane is parallel to the hologram plane, cf. sec. 2.5.2. Therefore, when this algorithm is used to reconstruct the tilted hologram with z as the reconstruction distance, the image is obtained for points on a plane parallel to the tilted hologram and at distance z from it; this plane is marked P_1 in Fig. 7.7. Since point O does not lie on this plane, it will be out of focus in the reconstructed image. In order to be able to image point O in focus, the reconstruction distance should be z' ; $z' \leqslant z$. From the geometry of Fig. 7.7:

$$z' = z \cos\theta \tag{7.4}$$

Moreover, in the reconstructed image at plane P_2, point O appears offset by a distance x_o' to the left if θ is positive or to the right if θ is negative, where:

$$x_o' = -z \sin\theta \tag{7.5}$$

Fig. 7.8 shows the reconstructed images of a point object at $x_o=0$, $z=100\lambda$. The holograms are the middle portions of those whose phase functions are plotted in Fig. 7.5. For the case of zero tilt in the hologram, Fig. 7.8a, the image is at the centre of the reconstructed object plane and the optimum image occurs at 100λ. With $+5^o$ tilt, Fig. 7.8b, the optimum image occurs at 97.5λ and is offset to the left by a distance of 8.5λ. The values predicted by eqns. (7.4)and (7.5) for the optimum range and the offset distance are 99.61λ and 8.71λ respectively. for -2.5^o tilt angle, Fig. 7.8c, the point is offset to the right and the observed values are 97.7λ and 4.5λ while the calculated values are 99.9λ and 4.36λ respectively. The optimum reconstructed image was considered throughout as that having the maximum peak value.

7.2.3 Compensating for the Hologram Tilt

In order to minimize the effects of the relative tilt between the scanned hologram plane and the bottom of the water tank, the focused transducer was mounted on a heavy brass plate fitted with three adjustment screws in a tripod arrangement which can be used to tilt the

the transducer in any direction in order to offset any misalignment, Fig. 7.9. The vertical position of the plate is adjusted such that the two-dimensional hologram recorded in the area directly on top of the transducer is symmetrical in both x and y directions and agrees with the simulated hologram for an object at the same location. A line scan on top of the centre of the transducer is then performed and the corresponding hologram data transferred to the 4051 minicomputer for reconstruction and expansion.

It should be noted that the lateral offset in the position of the reconstructed image allows for the effects an unknown angle of tilt in the hologram to be compensated for computationally for the case of a point object at the centre if the range z is known. First, the value of the tilt angle θ is determined by measuring the offset distance x_o' and using eqn. (7.5), x_o' is positive when the imaged point is offset to the right. The phase of the measured hologram signal $\varphi(x)$ is then modified to $\varphi'(x)$ according to the relationship:

$$\varphi'(x) = \varphi(x) - K\, x \sin\theta \qquad (7.6)$$

where x is the position of the hologram sample relative to the centre, x is positive for points at the R.H. half of the hologram. In the Fresnel zone the effect of tilt on the amplitude of the hologram signal can be neglected, cf. eqn. (7.2).

7.3 Image Reconstruction From Measured Holograms

Fig. 7.10 shows photographs of the two-dimensional holograms recorded for a point object on the hologram axis at a distance of 97λ together with the hologram spectrum in the spatial frequency domain and the reconstructed images at focus and at two other ranges on both sides of the true object range. Shown also in the same figure are the corresponding results for a simulated hologram having the same number of samples and the same sample spacing as the measured hologram. The layout of the contents of each photograph in Figs. 7.10, 12 and 13 is shown in Fig. 7.11a. The relationship between the various colours in these figures and the relative data values they represent is given in the table in Fig. 7.11b. Fig. 7.12 shows the results when the hologram is measured at a distance of 144λ from the point object. Both hologram have a number of samples N=32 spaced at a distance $\Delta = 1.11\lambda$ in both x and y directions. A comparison between the two figures indicates the effect of increasing the object range in reducing the spatial bandwidth of the hologram and in reducing the resolution of reconstructed image. Both figures indicate good agreement between simulated and experimental results.

The two holograms in Fig. 7.10 and 7.12 were measured using a transmitted pulse having a number of cycles $N_c=8$. The same pulse width was used for the holograms presented in the remainder of this chapter. However, in order to show the effect of the number of cycles in the acoustic pulse on the accuracy of the hologram measurement and the quality of the reconstructed image, the hologram at z=144λ was measured with pulses 4 cycles and 15 cycles wide. The results are shown in Fig. 7.13. For the case of the narrow pulse, Fig. 7.13a, although only three active cycles are used for the hologram data acquisition, cf. sec. 6.3.5, the accuracy of the measured hologram is not seriously affected. In fact, the quality of the reconstructed image when the narrow pulse is used is somewhat improved due to the improvement in the range resolution. The phase variations at the centre of the hologram in both cases may be due to a slight residual tilt in the hologram plane.

More quantitative evaluation of the results is obtained by reconstructing the line hologram measured at the middle of the raster scan of the two-dimensional hologram. Fig. 7.14 shows the reconstructed images from a line hologram measured at 93λ from the point object. Fig. 7.14a is a plot of the reconstructed image at the measured range of 93λ while the optimum image, i.e. that having a maximum peak value, is shown in Fig. 7.14b and is obtained at a reconstruction distance of 90λ. Fig. 7.15 shows the corresponding results for the hologram measured at 144λ for which the optimum range was found to be 132λ. Fig. 7.16 shows the measured amplitude and phase of the hologram measured at 93λ compared with the calculated results at a distance close to the optimum range. The corresponding results for the hologram at 144 λ are shown in Fig. 7.17.

As indicated above, the optimum image occurs at a value of the range which is in error compared to the measured range. The error is 3.3% for the hologram at 93λ and 8.3% in the case of the hologram at 144λ , the optimum range being lower than the measured range in both cases. This error can be attributed to the following factors:

(i) Random errors in the measured hologram signal due to additive noise and measurement and quantization errors. As shown in section 5.2.1, errors in the hologram signal are also caused by random errors in the positions of the hologram sampling points due to the limited accuracy of the mechanical scanning apparatus and to jitter in the movement of the sampling detector.

(ii) The measured range is taken as the distance between geometric focal point of the focused transducer and the tip of the sampling hydrophone. However, due to the limited aperture of the focused transducer, this focal 'point' will have a finite extent in the depth as well as the lateral directions. For the dimensions D=25.4 mm and z=21.6 mm for the focusing shell used, see Fig. 7.1, the criterion for range resolution in eqn. (2.71) gives $\delta_r \approx 1.4\lambda$ as the distance from the focal plane at which the radiation strength of the focal point drops by 20%. Since the hologram acquisition circuits operate on the signals that arrive first at the hydrophone detector after the acoustic pulse is transmitted, the measured hologram corresponds to an object at a shorter range. The lower the threshold level at the receiver circuits, cf. sec. 6.3 , the greater the difference between the effective range and the measured range since the receiver circuits respond to weaker hologram signals from points which are further removed from the focal plane of the focused transducer.

(iii) Errors in measuring the values of the object range and the sample spacing. Due to the limited measurement accuracy because of the experimental set-up and the small wavelength at the frequency of operation ($\approx$ 1.5mm), the error in the range measurement can be as high as 1λ. The quality of the reconstructed image was found to be particularly sensitive to the error in measuring the sample spacing as shown in Figs. 7.18 and 7.19 for the holograms at 93λ and 144λ respectively. In each of these two figures a ±5% error in the sample spacing was simulated by inserting in the image reconstruction algorithm a value for the sample spacing which is different from the measured value by the error factor. Fig. 7.18 indicates that up to 22% variations in the peak value of the optimum image are caused by a 5% error in the sample spacing. These variations in the peak value due to the error in the sample spacing influence the choice of the optimum image and therefore cause an error in the value selected as the optimum range.

(iv) Another factor which contributes to the error in the observed object range is the limited range resolution due to the finite pulse width. For a 1MHz transmitted acoustic pulse 8 cycles in length, signals from a sector 8λ thick contribute to the hologram signal and therefore objects within this sector cannot be resolved, cf. sec. 2.6.1.

(v) Residual alignment errors which cause a small relative tilt between the hologram plane and the base of the focused transducer. As shown in section 7.2, these errors also cause the optimum range to be lower than the actual range which is the case observed in practice.

7.4 Expansion of Measured Hologram Apertures

The prediction technique described in Chapters 4 and 5 for expanding a given hologram aperture has been applied for the two measured holograms discussed in section 7.3. Aperture expansion was performed using triangular predictive and corrective models. Three tests were made on each hologram, namely:

(i) The measured 32-point line hologram was split into two halves; an inner half about the hologram axis and an outer half which consists of the exterior portions on each side. The inner half was considered as a 16-point available aperture and expanded to twice its size in much the same way used in the simulation tests described in Chapters 4 and 5. Since the hologram data is known over the outer portions where the small aperture is extended, the prediction errors can be calculated and the image from the true extended aperture can be reconstructed.

(ii) The total measured aperture was expanded to twice its size to produce a 64-point hologram. Since the true hologram signals are not known over the aperture extension in this case, neither the prediction errors nor the image from the true extended hologram can be obtained.

(iii) The effect of smoothing the hologram data in the available aperture to reduce the effect of noise on the prediction accuracy was investigated in both cases mentioned above.

Fig. 7.20 shows the prediction errors and the reconstructed images at the optimum range when the inner half of the hologram at 93λ was expanded to twice its size. The expansion was performed using triangular polynomial predictive and corrective models of the types described in sections 5.5 and 5.6 with model order L=7 and 8 respectively. The corresponding results for the hologram at 144λ are shown in Fig. 7.21. The two figures indicate the effectiveness of the aperture expansion technique in extending the measured hologram data and in achieving the attendant improvement in resolution. In both cases, the images reconstructed from the corrected holograms are similar to those from the true holograms. As in the case of the simulation results described in section 5.6, correction plays a critical part in reducing the prediction errors obtained using the predictive model and therefore improving the reconstructed image. The reconstructed images at the measured ranges for both cases are shown in Fig. 7.22 which indicates that resolution improvement is maintained at ranges other than the optimum range.

Although the use of the corrective model improves the prediction accuracy, the improvement is limited by noise in the small 'available' aperture. Due to the random nature of noise, it causes undue assymmetry between the two halves of the available aperture in the case of the hologram of a point object located at the centre of the object plane. The corrective model is obtained by relating the signals predicted using a model based on one half of the available aperture to the true data in the other half, cf. sec. 4.8. This implies that the smoothness of the hologram function over the two halves of the available aperture is required for the effective application of the corrective model. Since excessive noise in the measured data affects the smoothness of the hologram function, this is bound to influence the amount of improvement achieved using the corrective model. Referring to Figs.7.16a and 7.17a for the measured phase data of the measured hologram and assuming that the simulated hologram in each figure represents the true phase function, it is estimated that noise levels as high as 30% and 45% of the phase data exist within the inner half of the holograms measured at 93λ and 144λ respectively.

To investigate the effect of noise on the effectiveness of the corrective model, the hologram data over the small'available' aperture was smoothed, prior to aperture expansion, using a data fitting utility routine [144] available on the Tektronix 4051 minicomputer. The routine fits cubic splines to the data and the amount of smoothing introduced is determined by the value of a smoothing factor, s. The higher the value of this factor the smoother the resulting function. This, however, is achieved at the expense of losing fine details which may correspond to genuine variations in the function and not to noise. The routine was applied to the hologram data of the 16-point aperture at the inner half of the measured holograms. Since the amplitude data was found to be fairly smooth, it is not significantly affected by the smoothing process. Fig. 7.23a shows the smoothed phase function fitted to the data in the hologram at 93λ with a smoothing factor s=300 together with the values of the original data. Using the coefficients of the spline fit determined by the 4051 routine, the values of the smoothed function were calculated at the hologram sampling points in the inner half of the measured aperture. Fig. 7.23b shows the measured hologram with its inner half smoothed. The corresponding results for the hologram at 144λ are shown in Fig. 7.24 with s=200.

The prediction errors and the reconstructed images using the smoothed hologram function are shown in Fig. 7.25 for the hologram measured at 93λ. Comparing this figure with Fig. 7.20 for the unsmoothed hologram, it can be seen that smoothing the data achieves an improvement in the performance of the corrective model. For example, the mean error over the aperture in the corrected hologram is reduced by approximately 50% and there is a corresponding improvement in the quality of the reconstructed image. The results for the hologram at 144λ are shown in Figs. 7.26. The prediction errors plotted in Fig. 7.26a indicate an improvement in the accuracy of the corrected hologram with smoothing, although this improvement is not as high as in the case of the hologram at 93λ since the mean error is reduced by only 16% compared to the case when row hologram data is used for prediction. This can be attributed to the increased noise in the hologram considered. It should also be noted that the improvement in the performance of the corrective model depends on the choice of the smoothing factor s, therefore it is possible that greater improvement can be achieved by using a more optimum value for s. The effect on the reconstructed image is shown in Fig. 7.26b.

In a practical situation only the measured data over the given aperture will be known, with no knowledge of the hologram signal over the aperture extension. This situation can be simulated by using all the 32 points of the measured data as the available aperture and extending it to twice its size to obtain a 64-point hologram. Prediction was performed using triangular predictive and corrective models of order L=15 and 16 respectively. Since the true hologram data is not known outside the available aperture, no information is available about the prediction errors or the image from the true extended aperture. The reconstructed images from the small, predicted, and corrected holograms are shown normalized to the same peak in Fig. 7.27 for the holograms at 93λ and 144λ. The figure indicates an improvement in resolution in the image from the corrected hologram with the aperture expansion. However, this improvement is much less than expected from doubling the size of the hologram aperture. As in the case of the 16-point 'available' aperture discussed earlier , this can be attributed to the poor performance of the corrective model due to noise in the measured data and therefore can be improved by smoothing the data over the 32-point available aperture. Figs. 7.28 and 7.29 show the smoothed phase function and the smoothed hologram phase with s=2000 for the holograms at 93λ and 144λ respectively. The images obtained from doubling the size of the smoothed holograms in the two cases are shown in Fig. 7.30 for the small and corrected holograms where the images from the predicted holograms were omitted for clarity. Compared to Fig. 7.27, the aperture expansion now leads to better improvement in resolution with the width of the point image at the half-power points reduced by 40% and 50% for the hologram at 93λ and 144λ respectively. This shows that smoothing the hologram data helps increase the overall prediction accuracy and improve the quality of the reconstructed images from expanded holograms.

7.5 Simulation and Reconstruction of Two-Dimensional Holograms

The software described in section 6.5 and the display system of section 6.4 were used for simulating, reconstructing, and displaying a range of two-dimensional holograms to illustrate a number of criteria which are pertinent to holography. Although not directly related to the technique of aperture expansion presented in this thesis and the subject of resolution improvement in general, these results are included here since they serve to demonstrate using colour photographs how resolution is influenced by parameters of the imaging system and by the imaging geometry. Throughout the following figures, the layout of the contents of each photograph is as shown in Fig. 7.11a and the colour code as shown in the Table in Fig. 7.11b. The simulation of the complex holograms for the discrete point objects was performed on the basis outlined in section 4.3 and the images reconstructed using the backward wave propagation (frequency domain) algorithm, cf. sec. 2.5.2.

Fig. 7.31 illustrates the hologram and the reconstructed images for a 1-point object when the object is located on the hologram axis and when offset from the centre. As indicated in section 7.2.1., the hologram pattern is basically the same in both cases but is shifted with the point object.

Fig. 7.32 shows the improvement in resolution and the increase in the spatial frequency bandwidth of the hologram with the decrease in the object range for the case of a 1-point object.

Fig. 7.33 shows the effect of undersampling of the hologram and the resulting multiple images as described in section 2.4.1.

Fig. 7.34 illustrates the effect of varying the reconstruction distance inserted into the reconstruction algorithm when images are obtained from a given hologram. The image in Fig. 7.34b is reconstructed at a distance which is equal to the true range of the object while in Fig. 7.34a and 7.34c the reconstruction distance corresponds to two out-of-focus ranges both smaller and greater than the true object range respectively.

Fig. 7.35 shows two points resolved, with separation in the x direction in Fig. 7.35a and in the y direction in Fig. 7.35b. It is noted that the main variations in the amplitude of the hologram signal occur in the direction of the separation between the two points in the object.

Fig. 7.36 illustrates the effect of increasing the spacing between the two points on resolution. In Fig. 7.36a, the two points, located at a distance of 97λ from the hologram, are separated by a distance of 4λ. The Rayleigh resolution limit for the 31λ aperture at this range is approximately 3λ. However, since the two points are simulated such that they radiate coherently in phase, they are not resolved [33]. It is noted that because of the small separation in this case the phase function over approximately 75% of the area of the hologram aperture is similar to that of a single point object, hence the difficulty in resolving the two points. Increasing the separation between the two points to 6λ allows the two points to be resolved as shown in Fig. 7.36b. In this case the portion of the hologram aperture over which the phase function is similar to that of a single point is less than half of the total aperture.

Fig. 7.37 shows the effect of doubling the linear dimensions of the hologram aperture on improving the resolution . The two points to be imaged are separated by a distance of 6λ. With a 15λ aperture these points cannot be resolved as shown in Fig. 7.37a since the hologram data intercepted by the small aperture carries little information that the object is made up of two points. Increasing the size of the aperture to 31λ by doubling the number of points in the hologram at the same sample spacing allows the two points to be resolved, Fig. 7.37b.

Fig. 7.38 demonstrates the three-dimensional imaging capability of holography, cf. sec. 2.7. The simulated object consists of two points separated laterally and located at different ranges . Reconstructing the hologram at a distance equal to the range of one point images that point in focus while the other point appears out of focus. The separation between the two points in range exceeds the aperture limited depth of focus calculated at the furthest range of the two points, cf. eqn. (2.72). Therefore, when the image of one point is brought into focus, contributions corresponding to the other point are comparatively small and,with the image intensity normalized to only 8 grey levels, these contributions do not appear in the displayed image. Note the difference in resolution in the case of each point which is caused by the differnt values of their ranges.

7.6 Conclusion

In this chapter a number of experimental results have been presented which demonstrate the effectiveness of the prediction approach to hologram aperture expansion discussed in Chapters 4 and 5 when used on realistic holograms. In the experiments described, an active point object is simulated using a focused transmitting transducer and the hologram is measured, reconstructed and displayed using the imaging system described in Chapter 6. The resulting hologram is effected by any relative tilt between the hologram plane and the plane on which the transducer is mounted. It is shown that this is equivalent to introducing an offset reference to the complex hologram signal which corresponds to the case of no tilt. This leads to errors in the images reconstructed using algorithms based on the assumption that the complex hologram is measured relative to an on-axis reference. These are manifested in a lower value for the range of the optimum reconstructed image in addition to an offset in its lateral position. The tilt problem has been overcome by allowing for mechanical alignment of the two planes. Other factors which contribute to the errors in the range of the reconstructed images include the finite depth of the focal region of the transducer, the finite width of the transmitted acoustic pulse, errors in the measurement of the extent of the hologram scan and the distance between the object and the hologram plane, in addition to random errors in the hologram signal and in the positions of the sampling points.

The results discussed have shown the usefulness of the prediction technique as a means of aperture expansion in practical situations. The measured hologram data at 32 sampling points has allowed both a 16-point and a 32-point apertures to be doubled in size. In the case of the 16-point aperture the hologram signal is known over the aperture extensions which allows the prediction errors to be calculated and the images reconstructed from predicted and true holograms over the extended aperture to be compared. In the case of the 32-point available aperture no information is available on the hologram signal outside this aperture which is the case in practical situations of aperture expansion. As in the simulated tests reported in Chapters 4 and 5, the corrective process is essential for the reduction of the prediction errors over the extended aperture in order to achieve the expected improvement in resolution.

Since each corrective model is based on the data in the two halves of the available aperture, it is influenced by the effects of noise and measurement errors in the measured data on the smoothness of the hologram function. The results obtained have demonstrated that the performance of the corrective model can be improved by smoothing the hologram data in the available aperture before expansion in order to reduce noise and measurement errors, with the degree of improvement determined by the amount of smoothing employed. For small amounts of smoothing, the smoothed data follows the noisy data closely and therefore contains a certain amount of noise. Over-smoothing, on the other hand, tends to interfere with fine genuine details in the hologram function. Therefore, it is expected that an optimum condition exists. This optimum condition, however, would be a function of the individual hologram being considered. Another approach to improving the effectiveness of the corrective process is to increase the accuracy of the hologram measurement.

CHAPTER 8
CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK

8.1 Conclusions

From the work reported in this thesis, the following conclusions can be made:

1) A number of the problems facing acoustic holography in its conventional form of optically reconstructing an interference pattern can be largely overcome by measuring the complex hologram and using computer reconstruction. With faster, cheaper, and more efficient computers becoming more readily available, it would be economical to use the microcomputer as an integral part of the imaging system for controlling the various aspects of the imaging process including data acquisition, image reconstruction and processing, and display. With high speed computers and signal processing devices, image reconstruction can be performed fast enough to avoid the problem of time-lag associated with optical reconstruction. As shown in section 2.5.2, the use of the complex hologram reduces the requirement on the spatial resolution of the hologram detectors and therefore helps economize the number of samples required and speed up both the data acquisition and image reconstruction stages of the process. Moreover, the absence of the spatial carrier associated with the use of a reference wave means that there would be no interference with the reconstructed image from other unwanted components. The requirement for a laser is avoided together with all the image distortion caused by the wavelength mismatch and the use of recording and reconstruction optics. The use of the computer together with the linear detection of the hologram signals allows a large amount of signal processing to be cost-effectively performed in order to alleviate a number of the more basic limitations in holography such as the limited resolution.

2) The limited resolution in holography is one facet of the general problem caused by the loss of information on a given function due to the truncation of the function or its spectrum because of the limited window over which the data is collected, cf. sec. 3.4. This limits resolution in imaging systems and restricts the accuracy with which the spectrum of a time-limited function can be determined in spectral analysis. When the object is finite (or the time function is band-limited) the object spectrum (or the time function) is

analytic and it is possible to continue the truncated function uniquely beyond the limited region over which it is known and therefore improve resolution or increase the accuracy of spectral estimation. Noise and measurement errors in the known portion of the spectrum set the limit on the amount of improvement that can be achieved using realistic data. Aperture extension techniques reported in the literature assume imaging in the far-field region so that the data collected correspond to the object spectrum. However, the analogy with the problem in spectral analysis suggests that band-limited holograms correspond to an analytic function in space regardless of the imaging range, cf. sec. 4.2. A formal proof for the case of Fresnel holograms is given in the same section.

3) From the review given in section 3.4, it is noted that the majority of object restoration techniques for resolution improvement can be traced back to the principle of analytic continuation. These techniques, however, can be conveniently divided into two main categories : those which perform this continuation directly by calculating the function at new points beyond the known region using the measured data, and those which perform the continuation indirectly. Examples of the latter type include the error reduction technique by Gerchberg [124] and the statistical approach based on the maximum entropy method by Frieden [125].

Although techniques in this latter category give improved performance with noise compared to conventional direct methods, they have a number of limitations. For example, the technique by Gerchberg requires a large number of iterations, each employing two Fourier transforms. Moreover, the method assumes the knowledge or estimation of the object extent and the effectiveness of the technique depends heavily on how close the estimated value is to the correct value and the algorithm fails completely if the estimated value is less than the correct value. The statistical approach has the basic disadvantage that it operates on intensity distributions and therefore cannot be applied directly on the complex hologram data. Moreover, the method also requires the estimation of a number of parameters which are critical to its performance. Computationally, the method has the disadvantage that it requires the solution of a set of nonlinear equations and that there are situations in which no solution exists [125]. Methods which apply analytic continuation directly include the techniques by Harris [118] in optics and by Sato [128] in acoustic holography. Although these techniques are computationally simpler, they are less robust against noise in their basic form due to the ill-conditioning which results from the use of a sinc(x)/x model, cf. sec. 5.1.

4) The new method proposed in this thesis for resolution improvement by aperture expansion has the advantage of the simplicity of the direct approach to analytic continuation. This method is based on constructing a model to fit the measured hologram data and then using this model to predict new data outside the given aperture. The prediction accuracy increases as the fitting error between the model and the hologram function decreases. This depends on both the model function and the model order for a given hologram, cf. sec. 4.5. From the simulation results discussed in section 4.4., models based on polynomial expansion in space give optimum results for the case of a single point object. For multiple-point objects, linear terms which relate the signal at one point to the signals at the neighbouring points are required, see sections 4.6 and 4.7.

5) Correction of the predicted data is possible using a corrective model which is constructed by relating the true data at one half of the available aperture to the data at the same half which is predicted using a model based on the other half. The effectiveness of the correction process depends on the composition of both the predictive and corrective models, cf. sec. 4.8.

6) As in the case of all analytic continuation techniques, noise poses the limit for the effectiveness of the proposed method. However, since noise sensitivity is caused by the ill-conditioning of the set of the linear equations describing the model, this method is more flexible in overcoming the problem of noise in comparison with similar techniques which employ direct analytic continuation.

In the techniques by Harris [118] and Sato [128] for example, the rigid formulation of the problem in the form of a sinc(x)/x model based on the sampling theorem leaves little room for attempts to overcome the ill-conditioning in the system of equations, apart from manipulating the sample positions by using random sampling for example which is not always desirable since it affects both hologram acquisition and reconstruction. In the proposed method, however, the selection of the model used is more arbitrary, with the only requirement from the prediction viewpoint that it represents the hologram function accurately. This flexibility in the choice of the model type allows the stability with noise to be taken into account in the selection of the predictive or corrective models. Unfortunately, it is often the case that a model which gives good prediction accuracy suffers from high sensitivity to noise. However, as indicated in Chapter 5, a compromize can be reached in which the improvement in the stability with noise far exceeds the deterioration in the prediction accuracy and adequate overall performance is achieved in the presence of noise. This approach has the advantage over Sato's technique [128] that the stabilization with noise is built into the signal processing algorithm and therefore does not affect the way in which the hologram is sampled or reconstructed.

The strategy adopted for stabilizing the performance of the method with noise is to use models which result in a triangular form of the system matrix. As shown in section 5.4, this triangularization counteracts the tendency of neighbouring rows to become equal which is the cause of matrix instability. This is achieved since each pair of neighbouring rows will be different by at least the last non-zero term of the lower order row when the matrix takes an upper triangular form. In the polynomial model, the triangular form of the matrix is obtained by a gradual truncation of the polynomial expansion for points further down the available aperture and the resulting loss in the prediction accuracy necessitates the use of a triangular corrective model, cf. sec. 5.6. In the case of the linear model with no polynomial terms, the triangular form appears naturally as a result of the windowing effect when the model is constructed since data contributing in the formation of the model is restricted to one half only of the available aperture. In the examples given in sections 5.6 and 5.7 for the performance with noise using simulated data, both 1-point and 2-point objects have been imaged correctly from apertures obtained by doubling the size of the available aperture with relative noise levels of up to 30%.

7) The linear predictive model described in sections 4.6 and 5.7 has a number of advantages which enable it to be used as a universal predictive model for the case of multiple-point objects. These advantages can be summarized as follows:

a) Since this model has no polynomial terms, it is of fixed composition and therefore can be used for a variety of imaging configurations without the requirement for the optimum model composition to be determined as in the case of the hybrid model. In practice, determining the optimum model for each imaging configuration is inconvenient since it involves a search for the best image obtained from a set of models and requires a priori information about the shape of the object.

b) When imaging multiple-point objects, the model gives adequate prediction accuracy for it to be used without the requirement for additional correction which simplifies the procedure for aperture expansion.

c) Due to its triangular matrix form, the linear model is stable in the presence of noise in the hologram data and therefore can tolerate errors in the measured data caused by the limited accuracy of scanning and measuring the hologram.

d) The relative insensitivity of the model to variations in the object parameters, cf. sec. 4.10, allows the model to be used for a wide range of imaging and object geometries including continuous objects.

e) The stability of the model with round-off errors arising from the limited wordlength of the computer and the limited accuracy of performing the computations, cf. sec. 5.8, makes it particularly useful for aperture expansion in imaging systems which employ small mini or microcomputer systems.

8) Simulation results using the linear model for expanding the aperture in the case of quasi-continuous objects have indicated that the prediction error does not increase significantly with the increase in the number of points representing an object of a given width. In fact, the error is shown to stabilize at a steady-state value as the case of a continuous object is approached, cf. sec. 4.10. This suggests that the prediction technique described in this thesis should be suitable for use with continuous objects.

9) Experimental results presented in Chapter 7 have verified the basic principles of the aperture expansion techniques including both the prediction and correction algorithms for the case of a 1-point object. A number of practical considerations limit the accuracy of both image reconstruction from measured data and the expansion of the available aperture; these include : noise and measurement errors in the hologram signal, random errors in the positions of the sampling points, and the effect of tilt in the hologram plane. Although correction provides a useful means of improving the prediction accuracy, the results have indicated that the effectiveness of the corrective model is influenced by noise in the measured data which reduce the smoothness of the hologram function in the two havles of the aperture. Smoothing the hologram data before aperture expansion reduces the effect of noise and improves the performance of the correction process. However, since this improvement depends on the amount of smoothing employed, the optimum smoothing factor would have to be determined for each individual hologram which is inconvenient and requires a priori information on the object. Improving the accuracy of the hologram measurement should also reduce the effect of noise on the corrective model. Additionally, more practical applications are concerned with multiple-point objects. In this case the linear predictive model allows adequate performance without the requirement for correction.

10) The experimental holographic imaging system described in Chapter 6 demonstrates the advantages of using microprocessor technology

in integrating the various aspects of the holographic imaging process and reducing the need for special purpose hardware to perform the required functions. The system serves to illustrate the capabilities of small microprocessor systems in performing the tasks of hologram acquisition, image reconstruction and display, in addition to the prespects for improving the image quality through signal processing means.

8.2 Suggestions for Future Work

The work described in the previous chapters leads to a number of areas for improvement and future studies. These will be conveniently divided into those related to the theoretical and experimental aspects of the work.

8.2.1 Suggested Future work Related to Theoretical Aspects

The following theoretical points are believed to be worth investigating in the future:

. Modelling Techniques

The relative flexibility in the choice of the model which is used to represent the hologram function leaves room for investigating possible improvements both in the prediction accuracy and in the stability of the model with noise. In particular, the following points are worth mentioning:

1) For simplicity, all the predictive and corrective models described throughout the thesis are based on an exactly determined system of equations. For the polynomial model, for example, the resulting function represents an interpolating polynomial which satisfies all the equations and therefore the data fit described by the model passes through all the data points. Since such a data fit follows the data closely, it would be more sensitive to noise in the data compared to the case of a least-squares fit obtained when the set of equations is overdetermined.

2) The approach adopted throughout the thesis to the modelling of the hologram signal has been such that the model relies only on the hologram parameters without the requirement for any of the object parameters. However, the analysis described in section 4.5 suggests that a more accurate representation of the hologram signal is obtained when other parameters such as the object range and location are included in the model function. Although this represents a requirement for additional a priori information on the object, this

information may be readily available in certain applications. From eqn.(4.52), it would be expected to obtain better prediction accuracy from the polynomial model if the model equation (4.53) is replaced by:

$$h_o(k) = \sum_{i=1}^{L} a_i \left[\frac{K\Delta}{z}(d_k - x_o) \right]^{i-1} \tag{8.1}$$

where the various parameters are as defined in section 4.5.

3) The results in section 5.4 on the stability of the square and triangular polynomial models with noise have indicated that these two models represent two extremes in the trade-off between prediction accuracy and stability with noise, with the square model giving accurate prediction in the case of noiseless data but is highly sensitive to noise while the triangular model gives larger prediction errors but exhibits remarkable stability with noise. This suggests that some compromize between these two extreme cases might be possible. In particular, it is attractive to think of an adaptive method in which the amount of triangularization in the square matrix is made proportional to the noise level in the hologram signal. The model matrix A_{ad} in this case would be expressed as:

$$A_{ad} = A_{sq} - \alpha A_t \tag{8.2}$$

where A_t is the lower triangular part of the square matrix A_{sq} and $\alpha(0 \leqslant \alpha \leqslant 1)$ is chosen to suit the noise level in the hologram signal. $\alpha=0$ for the noise-free case and the matrix A_{ad} reduces to that corresponding to the square model. For the poorest signal-to-noise ratio anticipated, $\alpha=1$ and the triangular model is obtained. For intermediate noise levels the elements in the lower triangle of A_d will be smaller than those in the case of the square model and greater than the corresponding (zero) elements in the case of the triangular model and therefore a compromize may be achieved between prediction accuracy and stability with noise.

. Investigating the Effects of Errors in the Hologram Data on Image Reconstruction

Errors in the predicted points on the expanded hologram aperture affect the quality of the reconstructed images and therefore set the limit on the amount of aperture expansion that is considered useful. Since the end product of the imaging process is the generation of an image, it is obvious that it is the influence of the errors on the reconstructed image, rather than the errors in themselves, that pose the real limit on the amount of aperture expansion. Fig. 4.5. illustrates this point where prediction errors greater than 200% have almost negligible effects on the reconstructed image of a 1-point object. The effect of errors in the hologram signals caused by amplitude and phase variations in the channel processors is mentioned in the literature [85], [79]. However, the error in this case is assumed to be randomly distributed over the aperture while the error due to prediction would be a function of the coordinates on the aperture since this error increases towards the edge of the expanded aperture.

8.2.2 Suggested Future Work Related to Experimental Aspects

The following points summarize a number of areas for improvement and expansion on the experimental set-up used and the results obtained:

1) In the design and implementation of the experimental holographic imaging system described in Chapter 6, no special effort has been directed towards optimizing the speed and memory utilization throughout the various stages of the imaging process. Since mechanical scanning causes the major portion of the time delay, a significant improvement can be achieved by using arrays to sample the hologram. Moreover, the image reconstruction and data processing can be speeded up by using multi- processors and/or dedicated signal processing devices. For example, an additional microprocessor can be used to compute the FFT transform of the individual rows of data corresponding to the line scans while the scanning of the aperture proceeds under the control of the main processor. Dedicated FFT transformers and signal processing devices, such as the AMI S2814A and the TI TMS320 can be used to perform the FFT operation speedily and achieve considerable savings on the system memory.

2) The accuracy of hologram acquisition can be increased by improving on the mechanical set-up and increasing the accuracy of measurement techniques. For example, the quantization error in the phase measurement may be reduced by increasing the frequency of the 100 MHz clock. Multiple firing may be used to improve the measuring accuracy in general.

3) The technique described for the digital measurement of the phase of the hologram signal using a zero crossing detector should be evaluated more thoroughly. In particular, the advantages of the technique in improving the range resolution and the effect of using short pulses on the accuracy of the phase measurement should be investigated.

4) Experimental results should be extended to the cases of multiple-point objects and continuous objects.

5) The imaging system may be used to perform aperture expansion in two dimensions on both simulated and measured holograms. From the analysis in section 5.8 on the effects of limited computation accuracy, it appears that only triangular models would be suitable for this application due to the limited wordlength of the microprocessor system. The floating point routines employed also allow for double precision arithmatics to be performed.

ACKNOWLEDGEMENTS

The author is indebted to the SERC Marine Technology Directorate who funded the research program from which this thesis developed.

I am grateful to the head of the department, Professor A.M. Rosie, for providing some of the facilities required to carry out the research and write up the thesis.

Special thanks are due to both Professor T.S. Durrani and Dr C.J. Macleod, who jointly supervised this research project, for their invaluable guidance, help, and encouragement throughout the duration of the project.

I would also like to express my thanks to my colleague, Mr. J. Pearson and former colleague, Mr. A. Marshall for their cooperation, to the technicians of project MASS for their assistance in the construction of the hardware, and to Mrs. E. MacKinnon for her sincerity, patience, and efficiency in typing this thesis.

Finally, I wish to record my sincere gratitude to my wife Enaam for her continuous support and encouragement during this period.

References

1. M. Ahmed, K. Wang, and A. Metherell, "Holography and its application to acoustic imaging". Proc. IEEE, Vol 67, 1979, pp. 466-482.

2. B.B. Brenden, "History and present status of liquid surface acoustical holography"., J. Acoust. Soc. Am., Vol. 58, 1975, pp. 951-955.

3. G. Wade, "Historical prespectives"., Acoustical Imaging Vol. 8, Plenum Press, New York, 1978, pp. 21-42.

4. P. Langevin and M.C. Chilowski, "Procedes et appareils pour la production de signaux sous-marins direges et pour la localization a distance d'obstacles sous-marins", French Patent No. 502913, May 29, 1916.

5. S.J. Sokolov, "Ultrasonic oscillations and their applications"., Tech. Phys. U.S.S.R., Vol. 2, 1935, p. 552.

6. S.J. Sokolov, "Ultrasonic microscope"., Akedemia Nauk SSSR, Doklady (Teknicheskaya Fizika), Vol. 64, 1949, pp. 333-335.

7. W.L. Bragg, "The X-ray Microscope", Nature,Vol. 149, 1942, p. 470.

8. D. Gabor, "A new microscope principle", Nature, Vol. 161, 1948, p.777.

9. E.N. Leith and J. Upatneiks, "Reconstructed wavefronts and communication theory"., J. Opt. Soc. Am., Vol. 52, 1962, p. 1123.

10. E.N. Leith and J. Upatneiks, "Wavefront reconstruction with diffuse illumination and three-dimensional objects"., J. Opt. Soc. Am., Vol. 54, 1964, pp. 1295-1301.

11. P. Greguss, "Ultrasonic holograms"., Res. film., Vol. 5. 1965, pp. 330-337.

12. R.K. Mueller and N.K. Sheridon, "Sound holograms with optical reconstruction", Appl. Phys. Lett., Vol. 9, 1966, pp. 328-329.

13. J.L. Ferguson, "Liquid-crystal detectors", Acoustical Holography, Vol. 2, Plenum Press, New York, 1969.

14. J.D. Young, and J.E. Wolfe, "A new recording technique for acoustic holography"., Appl. Phys. Lett., Vol. 11, 1967, p. 294.

15. J.A. Cunningham and C.F. Quate, "Acoustic interference in solids and holographic imaging"., Acoustical Holography, Vol. 4., Plenum Press, New York, 1972.

16. F.L. Thurstone, "Ultrasonic holography and visual reconstruction", Proc. Symp. Biomed. Eng., Vol. 1., 1966, pp. 12-15.

17. R.K.Mueller, E. Marom, and D. Fritzler, "Electronic simulation of a variable inclination reference for acoustic holography via the ultrasonic camera", Appl. Phys. Lett., Vol. 12, 1968, p. 394.

18. A. Korpel and P. Desmares, "Rapid sampling of acoustic holograms by Laser-scanning techniques", J. Acoust. Soc. Am., Vol. 45, 1969, pp. 881-884.

19. N. Booth and B. Saltzer, "An experimental holographic acoustic imaging system"., Acoustical Holography, Vol. 4., Plenum Press, New York, 1972.

20. P. Alias, "Real time acoustical imaging with a 256x256 matrix of electrostatic transducers", Acoustical Holography, Vol. 5, Plenum Press, New York, 1974.

21. A.K. Nigam, "Condenser - microphone arrays in acoustical holography - a review", J. Acoust. Soc. Am., Vol. 55, 1974, pp. 978-985.

22. A.F. Metherell and H.M.A. El-Sum, "Simulated reference in a coarsely sampled acoustical hologram", Appl. Phys. Lett., Vol. 11, 1967, pp. 20-22.

23. B.B. Brenden, "Acoustical holography as a tool for non-destructive testing", Materials evaluation, June 1969, pp. 140-144.

24. D. Fritzler, E. Marom, and R.K. Mueller, "Ultrasonic imaging via the ultrasonic camera"., Acoustical holography, Vol. 1, Plenum Press, New York, 1969.

25. J.W. Cooley and J.W. Tukey, "An algorithm for the machine calculation of complex Fourier series", Math. Comput. Vol. 19, 1965, p. 297.

26. M.M. Sondhi, "Reconstruction of objects from their sound-diffraction patterns", J. Acoust. Soc. Am. Vol. 46, 1969, pp. 1158-1164.

27. A.L. Boyer, J.A. Jordan, Jr., D.L. Van Rooy, P.M. Hirsch, and L.B. Lesem, "Computer reconstruction of images from ultrasonic holograms"., Acoustical Holography, Vol. 2, Plenum Press, New York, 1970.

28. A.L. Boyer, P.M. Hirsch, J.A. Jordan, Jr., L.B. Lesem, and D.L. Van Rooy, "Reconstruction of ultrasonic images by backward propagation", Acoustical Holography, Vol. 3, Plenum Press, New York, 1971.

29. A.S. Gan, "Application of acoustical holography to seismic data processing", Acoustical Imaging and Holography, Vol. 1, 1979, pp. 143-148.

30. D.R. Holbrooke, E.E. McCurry, V. Richards, "Medical uses of acoustical holography"., Acoustical Holography, Vol. 5., Pelnum Press, New York, 1974.

31. M. Wollman, and G. Wade, "Experimental results from an underwater acoustical holographic system"., Acoustical Holography, Vol. 5., Plenum Press, New York, 1974.

32. J.B. Swint, B.G.W. Wee, and N.H. Godbold, "Application of acoustical holography to flaw detection"., Acoustical Holography, Vol. 3., Plenum Press, 1971.

33. J.W. Goodman,"Introduction to Fourier Optics," McGraw-Hill, New York, 1968.

34. T.J. Skinner, "Surface texture effects in coherent Imaging", J. Opt. Soc. Am., Vol. 53, 1963, p. 1350.

35. B.P. Hilderband and B.B. Brenden, "An Introduction to Acoustical Holography", Plenum Press, New York, 1972.

36. B.P. Hilderband, "Scanned acoustic holography", Acoustic Imaging, Ed. G. Wade, Plenum Press, New York, 1976.

37. N. Farhat, Ed., Advances in Holography, Vol. 1, Marcel Dekker, 1975.

38. J.P. Powers, J.Y. De Blois, R.T. O'Bryon and J.W. Patton, "A computer aided ultrasonic imaging system"., Acoustical Imaging, Vol. 8, Plenum Press, New York, 1980.

39. W.R. Fenner and G.E. Stewart, "An ultrasonic holographic imaging system for medical applications"., Acoustical Holography, Vol. 5., Plenum Press, New York, 1974.

40. J.L. Sutton, J.V. Thorn, and J.N. Price, "The effects of circuit parameters on image quality in a holographic acoustic imaging system", Acoustical Holography, Vol.5., Plenum Press, New York, 1974.

41. J.L. Sutton, J.V. Thorn, N.D. Booth, B.A. Saltzer, "Description of a navy holographic underwater imaging system (AIS)", Acoustical Imaging, Vol. 8, Plenum Press, New York, 1980.

42. R.B. MacAnally, "Inclined reference wave in acoustic holography"., Appl. Phys. Lett., Vol. 11, 1967, pp. 266-268.

43. E. Marom, R.K. Mueller, R.F. Koppelmann, and G. Zilinskas, "Design and preliminary test of an underwater viewing system using sound holography", Acoustical holography, Vol. 3, Plenum Press, New York, 1971.

44. E. Marom, D. Fritzler, and R.K. Mueller, "Ultrasonic holography by electronic scanning of a piezoelectric crystal"., Appl. Phys. Lett., Vol. 12, 1968, p.26.

45. P.S. Green, "Acoustic Holography"., LMSC Tech. Report No. 4-77-67-11, Lockhead Missile and Space Company, Palo Alto, Calif., 1966.

46. H.M.A. El-Sum, "The scope of the symposium"., Acoustical Holography, Vol. 1, Plenum Press, New York, 1969.

47. P.N. Keating, R.F. Kopplemann, T. Sawatari, "Holographic aperture synthesis via a transmitter array"., Acoustical Holography, Vol 6., Plenum Press, New York, 1975.

48. K. Nitadori, "Synthetic aperture approach to multi-beam scanning acoustical imaging"., Acoustical Holography, Vol.6., Plenum Press, New York, 1975.

49. P. Alias, "Acoustical imaging by electrostatic transducers"., Acoustical Holography, Vol.4., Plenum Press, New York, 1972.

50. E.E. Aldridge, A.B. Clare, and D.A. Shepherd, "Ultrasonic holography in nondestructive testing", Acoustical Holography, Vol. 3, Plenum Press, New York, 1971.

51. J.L. Pfeifer, "Acoustical reconstruction of holograms and their potential use", Acoustical Holography, Vol.4., Plenum Press, New York, 1972.

52. J.L. Weaver, and G.C. Knollman, "Real-time reconstruction of images from hydroacoustic holograms"., Acoustical Holography, Vol. 4., Plenum Press, New York, 1972.

53. Y. Aoki, "Acoustical holograms and optical reconstruction"., Acoustical Holography, Vol. 1., Plenum Press, New York, 1969.

54. K.R. Erikson, B.J. O'Loughlin, J.J. Flynn, E.J. Pisa, J.E. Wreede, R.E. Greer, B. Stauffer, and A.F. Metherell, "Through-transmission acoustical holography for medical imaging - a status report"., Acoustical Holography, Vol. 6., Plenum Press, New York, 1975.

55. H.D. Collins and B.B. Brenden, "Acoustical holographic transverse wave scanning technique for imaging flaws in thick-walled pressure vessels", Acoustical Holography, Vol. 5., Plenum Press, New York, 1974.

56. G.L. Fitzpatrick, H.R. Nicholls, and R.D. Muson, "An experiment in seismic holography", Bureau of Mines Rep. Inv. 7607, 1972, p. 20.

57. R.K. Mueller, H.R. Farrah, and E. Marom, "Acoustical holography", Bendix Technical Journal, Summer 1969, pp. 15-24.

58. R.K. Mueller, "Acoustical holography", Proc. IEEE, Vol. 59, 1971, pp. 1319-1335.

59. J.L. Kreuzer, P.E. Vogel, "Acoustical holographic techniques for Non-destructive testing", Acoustical Holography, Vol.1., Plenum Press, New York, 1969.

60. J.W. Goodman, "Digital image formation from detected holographic data", Acoustical Holography, Vol. 1., Plenum Press, New York, 1969.

61. R.J. Doyle, and W.G. Glenn, "Lumartron : a high-resolution storage and projection display device", IEEE Trans. Electron. Devices, Vol. ED-18, 1971, p. 739.

62. G.G. Goetz, "Real-time holographic reconstruction by electro-optic modulation",Appl. Phys. Lett., Vol. 17., 1970, p. 63.

63. R.K. Mueller, R.R. Gupta, and P.N. Keating, "A holographic weak-signal enhancement technique", J. Appl. Phys., Vol. 43., 1972, p. 457.

64. P.N. Keating, R.K. Mueller, and R.R. Gupta, "Conventional and weak-signal-enhancement holography in the presence of measurement errors", Acoustical Holography, Vol. 4., Plenum Press, New York, 1972.

65. M. Takagi, N.B. Tse, G.R. Heidbreder, C.H. Lee, and Glen Wade, "Computer enhancement of acoustic images"., Acoustical Holography, Vol. 5., Plenum Press, New York, 1974.

66. C.S. Clark, and A.F. Metherell, "Digital Processing of acoustical holograms", Acoustical Holography, Vol. 5., Plenum Press, New York, 1974.

67. M. Clement, "A new processing technique for scanned ultrasonic holography", Acoustical Holography, Vol. 6, Plenum Press, New York, 1975.

68. R.F. Steinberg, P.N. Keating, and R.F. Kopplemann, "Experimental implementation of advanced processing in acoustic holography", Acoustical Holography, Vol.6. Plenum Press, New York, 1975.

69. C.H. Lee, G. Heidbreder, G. Wade, and A. Coello-Vera, "On-line interactive computer processing of acoustic images", Acoustical Holography, Vol. 7, Plenum Press, New York, 1977.

70. J.C. Stamm and R. Priemer, "Image processing for aberration removal", Acoustical Holography, Vol. 7, Plenum Press, New York, 1977.

71. J.W. Goodman and R.W. Lawrence, "Digital image formation from electronically detected holograms", Appl. Phys. Lett., Vol. 11, 1967, pp. 77-79.

72. J.M. Smith and N.F. Moody, "Application of Fourier transforms in assessing the performance of an ultrasonic holography system", Acoustical Holography, Vol. 1., Plenum Press, New York, 1969.

73. Y. Aoki, "Image reconstruction by computer in acoustical holography", Acoustical Holography, Vol. 5., Plenum Press, New York, 1974.

74. J.P. Powers, " A computerized acoustic imaging technique incorporating automatic object recognition", Acoustical Holography, Vol. 5., Plenum Press, New York, 1974.

75. S. Sepehr, J.M. Reeves, and S.O. Harrold, "A self focusing ultrasonic image reconstruction", Acoustical Imaging, Vol. 12, Plenum Press, New York, 1982.

76. J.P. Powers, "Computer simulation of linear acoustic diffraction", Acoustical Holography, Vol. 7., Plenum Press, New York, 1977.

77. H.R. Farrah, E. Marom, and R.K. Mueller, "An underwater viewing system using sound holography", Acoustical Holography, Vol.2., Plenum Press, New York, 1970.

78. N. Booth and B. Saltzer, "An experimental holographic acoustic imaging system", Acoustical Holography, Vol. 4., Plenum Press., New York, 1972.

79. J.L. Sutton, "A tutorial on underwater acoustic imaging"., Acoustical Holography, Vol. 9., Plenum Press, New York, 1980.

80. G.L. Fitzpatrick, "Seismic imaging by holography", Proc. IEEE, Vol. 67, 1979, pp. 536-553.

81. O.K. Mawardi, "Phase distortions due to nonlinear effects in an acoustic field", Acoustical Holography, Vol. 1., Plenum Press, New York, 1969.

82. E.E. Aldridge, M.I. J. Beale, A.B. Clare, and D.A. Shepherd, "B-scan Fourier transform holography", NDT International, August 1979, pp. 167-174.

83. E. Alasaarela, K. Tervola, J. Ylitalo, and J. Koivukangas, "UHB Imaging"., Acoustical Imaging, Vol. 12., Plenum Press, New York, 1982.

84. W.L. Beaver, D.H. Dameron, and A. Macovski, "Ultrasonic imaging with an acoustic lens", IEEE Trans. Sonics Ultrason. Vol. SU-24, 1977, pp. 235-243.

85. G.L. Sackman, "A comparison of holographic versus lens type acoustic image systems by computer simulation", Acoustical Holography, Vol. 7., Plenum Press, New York, 1977.

86. H.F. Harmuth,"Acoustic Imaging with Electronic Circuits." Academic Press, New York, 1979.

87. J.L. Sutton, "Underwater acoustic imaging"., Proc. IEEE, Vol. 67, 1979, pp. 554-566.

88. P.N. Keating, "Acoustical Holography - a comparison with phased array sonar"., Acoustical Holography, Vol. 5., Plenum Press, New York, 1974.

89. H.K. Wickramasinghe, "High frequency acoustic holography in solids", Acoustical Holography, Vol. 5., Plenum Press, New York, 1974.

90. E.E. Watson, "Detection of acoustic sources using long-wavelength acoustical holography", J. Acoust. Soc. Am., Vol. 54., 1973, pp. 685-691.

91. J.L. Kreuzer, "Ultrasonic holography for nondestructive testing"., Material Evaluation, October 1968, pp. 197-203.

92. A.E. Holt and W.E. Lawrie, "Ultrasonic characterization of defects"., Acoustical Holography, Vol. 7., Plenum Press, New York, 1977.

93. V. Schmitz and M. Wosnitza, "Experiences in using ultrasonic holography in the laboratory and in the field with optical and numerical reconstruction", Acoustical Holography, Vol. 8., Plenum Press, New York, 1980.

94. H.D. Collins, R.P. Gribble, T.E. Hall, W.M. Lechelt, T.J. Luebke, J. Spalek, E.M. Sheen, and A. Stankoff, "Acoustical holography matrix array imaging system for the underwater inspection of offshore oil platform weldments"., Acoustical Holography, Vol. 8, Plenum Press, New York, 1980.

95. B.P. Hildebrand, "Holograms, ultrasound, computers combined to detect flaws"., Industrial Research and Development, November, 1982.

96. H.D. Collins, "Acoustical interferometry using electronically simulated variable reference and multiple path techniques", Acoustical Holography, Vol. 6., Plenum Press, New York, 1975.

97. F. Takahashi, K. Suzuki, and T. Kanamori, "Digital signal processing in acoustical focused image holography, Acoustical Holography, Vol. 8., Plenum Press, New York, 1980.

98. K. Suzuki, and B.P. Hildebrand, "Holographic interferometry with acoustic waves", Acoustical Holography, Vol. 6., Plenum Press, New York, 1975.

99. B.R. Tittmann, "Imaging in NDE", Acoustical Holography, Vol. 9., Plenum Press, New York, 1980.

100. A.Korpel, R. Whitman, and M. Ahmed, "Elimination of spurious details in acoustic images"., Acoustical Holography, Vol. 5., Plenum Press, New York, 1974.

101. J.V. Thorn, N.O. Booth, J.L. Sutton, and B.A. Saltzer, "Test and evaluation of an experimental holographic acoustic imaging system", Naval Undersea Centre Technical Publication 398, Nov. 1974.

102. K. Mano and K. Nitadori, "An experimental underwater viewing system using acoustical holography"., Proceedings of the IEEE Ultrasonic Symposium , 1977 , pp. 272-277.

103. K. Nitadori, K. Mano, and H. Kamata, "An experimental underwater acoustic imaging system using multi-beam scanning", Acoustical Imaging, Vol. 8., Plenum Press, New York, 1980.

104. L. Weiss, and E.D. Holyoke, "Detection of tumors in soft tissues by ultrasonic holography", Surg. Gyn. Obstet. Vol. 128, 1969, p. 953.

105. R.E. Anderson, M.D., "Potential medical applications for ultrasonic holography", Acoustical Holography, Vol. 5., Plenum Press, New York, 1974.

106. R.C. Chivers, "B-scanning and holography in opthalmic diagnosis"., Ultrasonics, September 1974, pp. 209-213.

107. W.E. Leriwell, "Holography at seismic frequency", presented at the European Association of Exploration Geophysicists (EAEG), Vinice, Italy, 1969.

108. G.L. Fitzpatrick and R.K. Mueller, "Seismic holography", presented at the 82nd Annu. Meeting of the Acoust. Soc. of Am., Denver, CO, Oct. 1971.

109. D. Gabor, G.W. Stroke, R. Restrick, A. Funkhouser, and D. Brum, "Optical image synthesis (complex amplitude addition and subtraction), by holographic Fourier transformation", Phys. Lett. Vol. 18., 1965, pp. 116-118.

110. J. Reid, and J. Wild, "Application of echo ranging techniques in determination of structure of biological tissues", Science, Vol. 115, 1952, pp. 226-230.

111. P.N. Keating, T. Sawatari, and G. Zilinskas, "Signal Processing in Acoustic Imaging"., Proc. IEEE Vol. 67., 1979. pp. 496-510.

112. W.E. Kock, "Holography can help radar find new performance horizons", Electronics, October 12, 1970, pp.80-88.

113. T. Sato, M. Ueda, and S. Fukuda, "Synthetic aperture sonar"., J. Acoust. Soc. Am. Vol. 54., 1973, p.799.

114. W.H. Wells, " Acoustical imaging with linear transducer arrays", Acoustical Holography, Vol. 2, Plenum Press, New York, 1970.

115. H.J. Caulfield, "Wavefront sampling in holography"., Proc. IEEE, Vol. 57, 1969, pp. 2082-2083.

116. G. Toraldo di Francia, "Resolving power and information", J. Opt. Soc. Am., Vol. 45., 1955, pp. 497-501.

117. S.A. Schelkunoff, "A mathematical theory of linear arrays"., Bell Sys. Tech. J., Vol. 22, 1943, pp. 80-107.

118. J.L. Harris, "Diffraction and resolving power", J. Opt. Soc. Am., Vol. 54., 1964, pp. 931-936.

119. E.A. Guillemin, "The Mathematics of Circuit Analysis"., John Wiley and Sons, New York, 1949.

120. Y.I. Khurgin and V.P. Yakovlev, "Progress in the Soviet Union on the theory and applications of band-limited functions"., Proc. IEEE, Vol. 65, 1977, pp. 1005-1029.

121. C.W. Barnes, "Object restoration in a diffraction-limited imaging system", J. Opt. Soc. Am., Vol 56, 1966, pp. 575-578.

122. C.K. Rushforth, and R.W. Harris, "Restoration, resolution and noise"., J. Opt. Soc. Am., Vol. 58., 1968, pp. 539-545.

123. Y. Biraud, "A new approach for increasing the resolving power by data processing", Astron. & Astrophys., Vol. 1., 1969, pp. 124-127.

124. R.W. Gerchberg, "Super-resolution through error energy reduction", Opt. Acta. Vol. 21, 1974, pp. 709-720.

125. B.R. Frieden, "Restoring with maximum likelihood and maximum entropy", J. Opt. Soc. Am., Vol. 62, 1972, pp. 511-518.

126. S.F. Burch and J.A. McKnight, "The application of maximum entropy to the processing of ultrasonic images of nuclear reactor components immersed in liquid sodium"., Acoustical Imaging, Vol. 12., Plenum Press, New York, 1982.

127. T. Yokota and T. Sato, "3-D active incoherent ultrasonic imaging", Acoustical Imaging, Vol. 12., Plenum Press, New York, 1982.

128. T. Sato, K. Sasaki, and K. Uemura, "Super-resolution imaging using a sequential estimation of a hologram in an extended area from data obtained in a limited area"., J. Acoust. Soc. Am. Vol. 65., 1979. pp. 976-984.

129. O. Ikeda and T. Sato., "On a method of super-resolution of a limited number of point targets using ultrasonic scanning", J. Acoust. Soc. Am., Vol. 57, 1975, pp. 334-337.

130. E.G. Williams, J.D. Maynard, and E. Skudrzyk, "Sound source reconstruction using a microphone array", J. Accust. Soc. Am., Vol. 68, 1980, pp. 340-344.

131. E.G. Williams, and J.D. Maynard, "Holographic imaging without the wavelength resolution limit", Phys. Rev. Lett., Vol. 45., 1980, pp. 554-557.

132. T. Sato, K. Uemura, and K. Sasaki, "Super-resolution acoustical passive imaging system using algebraic reconstruction", J. Acoust. Soc. Am., Vol. 67., 1980, pp. 1802-1808.

133. A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation", IEEE Trans. on Circuits and Systems, Vol. CAS-22, 1975, pp. 735-742.

134. S.D. Conte and C. DeBoor, "Elementary Numerical Analysis", McGraw-Hill, New York, 1972.

135. R.E.S. Abdel-Aal, C.J. Macleod, and T.S. Durrani, "Expansion of acoustic hologram apertures using ARMA modelling techniques", Acoustical Imaging, Vol. 12, Plenum Press, New York, 1982.

136. G.A. Korn and T.M. Korn, "Mathematical Handbook for Scientists and Engineers"., McGraw-Hill, New York, 1968.

137. M. Born and E. Wolf, "Principles of Optics", Pergamon Press, Oxford, 1980.

138. C.L. Lawson, "Solving Least Square Problems", Prentice-Hall, N.J., 1974.

139. "9900 Family System Design and Data Book", Texas Instruments Inc., 1978.

140. R.E.S. Abdel-Aal, "Digital phase meter updates measurement each cycle", Electronics, September 22, 1981, pp 156-157.

141. "4051 Graphics System Reference Manual", Tektronix Inc., 1976.

142. J. Ophir and N.F. Maklad, "Digital scan converters in diagnostic ultrasonic imaging", Proc. IEEE, Vol. 67, 1979, pp. 654-664.

143. G.A. Massey, "Acoustic holography in air with an electronic reference", Proc. IEEE, Vol 55, 1967, pp. 1115-1117.

144. "Plot 50 : Mathematics, Volume 2", Tektronix Inc., 1975.