

Generalised Phase Diversity Wavefront Sensing

Heather Isla Campbell, M.Phys (Hons)

Submitted for the degree of Doctor of Philosophy

Heriot-Watt University

School of Engineering and Physical Sciences

June 2006

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that the copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the prior written consent of the author or of the University (as may be appropriate).

Abstract

Phase Diversity (PD) is a phase retrieval algorithm which traditionally uses a defocus kernel aberration to create the phase diverse data images. This thesis concerns the generalisation of this method to allow the use of other diversity kernels. The necessary and sufficient conditions such a kernel must satisfy for use in a Generalised Phase Diversity (GPD) null wavefront sensor are derived. GPD is extended to full wavefront sensing through the development of a new phase retrieval algorithm. The motivation for this work is to create a wavefront sensor capable of dealing with scintillated and discontinuous wavefronts thus overcoming the limitations of the current Intensity Transport Equation (ITE) based phase retrieval methods used with defocus-only PD sensors. In this thesis a new deconvolution based algorithm, which uses the small angle approximation to reach an analytic solution for the wavefront phase, is described. The Small Angle Expansion (SAE) algorithm is shown to work well with continuous, discontinuous and non-uniformly illuminated wavefronts. The GPD wavefront sensor, coupled with the SAE algorithm, is tested experimentally and the initial results are very encouraging. It will be shown that GPD provides a versatile, compact and robust method of wavefront sensing that may be useful in a wide variety of applications.

Acknowledgments

First and foremost I would like to thank my supervisor, Prof. Alan Greenaway, and my colleague Dr Sijiong Zhang for their support, guidance, and endless patience throughout the course of my PhD. I would also like to thank my CASE sponsors, the UK Astronomy and Technology Centre for supporting my project.

I would like to thank my family for their love and support (both financial and emotional!) over the years. I would never have gotten this far if it hadn't been for years of encouragement, kind words, and hugs. Thanks Dad, for telling me that the world is my oyster!

I would like to thank Paul for being my rock, for loving me and keeping me sane through this degree as well as the last one.

For all the lunchtime loafers (in no particular order) – Lorraine, Jess, Clare, Stuart, Karl, Euan, Matthias, Stefan, Arran, Martin and Tim. Thanks for keeping me laughing the whole way through! I would also like to thank The Hoff for the music.

My sincere thanks go to my research group, the Heriot-Watt Waves and Fields Group, both past members and present. It has been a joy to work with you. I would also like to thank Mr Forbes, my teacher at Lornshill Academy, who first showed me that physics is cool and inspired me to become a physicist.

ACADEMIC REGISTRY

Research Thesis Submission



Name:	Heather Isla Campbell		
School/PGI:	School of Engineering and Physical Sciences		
Version: <i>(i.e. First, Resubmission, Final)</i>	Final	Degree Sought:	PhD

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

- 1) the thesis embodies the results of my own work and has been composed by myself
- 2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
- 3) the thesis is the correct version of the thesis for submission*.
- 4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying, subject to such conditions as the Librarian may require
- 5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate:		Date:	
-------------------------	--	-------	--

Submission

Submitted By <i>(name in capitals)</i> :	
Signature of Individual Submitting:	
Date Submitted:	

For Completion in Academic Registry

Received in the Academic Registry by <i>(name in capitals)</i> :			
Method of Submission <i>(Handed in to Academic Registry; posted through internal/external mail):</i>			
Signature:		Date:	

Contents

Abstract	ii
Acknowledgments	iii
Thesis Submission Form	iv
Author's Publications	x
Chapter 1: Introduction	1
1.1 Introduction	1
1.2 Adaptive Optics and the Earth's Atmosphere ~ Historical Background	1
1.3 The Basic Principles of Adaptive Optics	3
1.4 Wavefront Sensing	5
1.4.1 The Shack-Hartmann Wavefront Sensor	6
1.4.2 Phase Diversity and Curvature Wavefront Sensing	9
1.4.2.1 Phase Diversity and Curvature Sensors – Advantages & disadvantages ..	11
1.4.3 The Paterson-Dainty Hybrid Sensor	12
1.4.4 Wavefront Shearing Interferometry	13
1.4.5 The Smartt-Point Diffraction Interferometer	14
1.4.6 Schlieren Imaging	17
1.4.7 Scanning Knife-Edge	18
1.4.8 Pyramid or Scanning Prism Knife Edge	19
1.4.9 The Modal Wavefront Sensor	21
1.4.10 Image Sharpening and Indirect Methods	22
1.4.11 Summary of Wavefront Sensing Technologies	23
1.5 References	24
Chapter 2: Phase Diversity Wavefront Sensing	29
2.1 Introduction	29
2.2 The Defocus Only Phase Diversity Wavefront Sensor	30
2.2.1 The Intensity Transport Equation	32
2.3 The Quadratically Distorted Diffraction Grating	33
2.3.2 Examples of Data	38
2.4 Data Reduction	41
2.4.1 The Green's Function Solution	41
2.4.2 The Gureyev-Nugent Modal Solution	43

2.4.3 Comparing the Properties of the Green's Function and GN Algorithm	45
2.5 Accuracy	46
2.5.1 Accuracy of the Green's function solution	47
2.5.2 Accuracy of the GN algorithm	49
2.5.3 Comparison of the Green's function and GN algorithm accuracies to the Shack- Hartmann wavefront sensor	51
2.6 Limitations of the DPD method	52
2.7.1. Application Considerations	53
2.8 Conclusions	55
2.9 References	56
Chapter 3: Generalised Phase Diversity	59
3.1 Introduction	59
3.2 Image Plane vs. Pupil Plane	60
3.2.1 Image Plane Operation	61
3.2.2 Pupil Plane Operation	64
3.2.3 Comparison of the Image Plane and Pupil Plane Configurations	65
3.2.4 Application Considerations	68
3.3 The Filter Function	68
3.4 The Intensity Images	70
3.5 The Error Signal	72
3.6 Symmetries of the Filter Function	75
3.6.1 Same Symmetry	75
3.6.2 Mixed Symmetry	77
3.6.3 Necessary and Sufficient Conditions	78
3.7 Simulations to verify the Sufficient and Necessary Conditions	79
3.7.1 Plane Wave Input Simulation Results	80
3.7.2 Aberrated Wavefront Simulation Results	82
3.7.2.1 Aberrated Wavefront 1	82
3.7.2.2 Aberrated Wavefront 2	84
3.7.3 Wavefronts with π Phase Steps	86
3.8 The Error Signal	88
3.8.1 Error Sense and Location	89
3.8.2 The Diffraction Grating and the Blur Function	89
3.9 Optimisation	94

3.9.1 Case 1	95
3.9.2 Case 2	98
3.9.3 Optimisation Conclusions	100
3.10 Conclusions	101
3.11 Appendix: Glossary of Commonly Used Functions	103
3.12 References	104
Chapter 4: Phase Retrieval with Generalised Phase Diversity	105
4.1 Introduction	105
4.2 Requirements and Considerations for a GPD Phase Retrieval Algorithm.....	107
4.2.1 Iterative vs. Analytic	108
4.3 Small Angle Expansion.....	109
4.3.1 Definitions.....	111
4.3.2 Manipulating the Error Signal Equation	112
4.3.3 Checking the Equation by Calculating the Difference Image.....	115
4.3.4 Solving for the Phase	118
4.3.5 The Nature of the SAE Phase Solution	120
4.4 The Relationship between the Error Signal and the Phase Difference	122
4.5 PV Wavefront Error and the SAE.....	125
4.6 Simulation Results	127
4.6.1 Hard-Edged Pupil.....	128
4.6.2 Soft-Edged Pupil	131
4.6.3 Factors Affecting the Fit of the Solution for the Hard and Soft-Edge Simulations.....	136
4.6.4 Iterative Refinement.....	141
4.6.5 Iterations and Wavefronts with High Local Slope.....	147
4.6.6 Alternate Approaches.....	150
4.7 Conclusions	151
4.8 References	153
Chapter 5: Further Study of the Small Angle Expansion Phase Retrieval Algorithm	155
5.1 Introduction	155
5.2 The Phase Transfer Function	156
5.2.1 Comparisons of the PTF's Generated by Different Diversity Functions.....	157
5.2.2 Possible Optimisation of the PTF	163

5.2.2.1	Crossed Gratings: A Method of Applying Multiple Diversities	163
5.2.2.2	PTF Engineering By Linear Combination Of Diversity Phase Terms....	166
5.2.3.	The Phase Contrast Microscope.....	170
5.2.3.1	PCM – Its PTF and Performance as a Diversity Function.....	173
5.2.4	PTF Engineering using PCM in Linear Combination with a Zernike Filter...	178
5.2.5	Studying the PTF and PTF Engineering: Conclusions	180
5.3	Extending the SAE.....	181
5.3.1	Use of the SAE with Non-Uniform Illumination.....	182
5.3.1.1	Extending the SAE Equations for Non-Uniform Illumination	183
5.3.1.2	Scintillation by Obscuration.....	187
5.3.2	Use of the SAE with Discontinuous Wavefronts.....	189
5.4	Conclusions	195
5.5	References	197
 Chapter 6: Experimental Testing, Practical Considerations and Applications of the Small Angle Expansion Algorithm		198
6.1	Introduction.....	198
6.2	Experimental Testing of the GPD Sensor and SAE Algorithm.....	200
6.2.1	Experimental Set-Up.....	200
6.2.2	Practical Issues	202
6.2.2.1	Centring Accuracy of the Data Images	204
6.3	Initial Results	210
6.3.1	Comparison with the GN Algorithm Solution	211
6.4	Normalisation and Data Conditioning	215
6.5	Regularisation	220
6.6	Treatment of the Boundary – An Ongoing Issue	226
6.7	Applications of the New Sensor System.....	228
6.7.1	Military Imaging and Free-Space Communications	229
6.7.2	Thin Film Metrology.....	230
6.7.3	Designer Beam Shaping of Femtosecond Lasers.....	232
6.7.4	Polishing Applications: Manufacturing Telescope Mirror Segments and Prosthetic Joints	235
6.8	Conclusions	237
6.9	Appendix: List of Zernike Polynomials Used.....	238
6.10	References	239

Chapter 7: Conclusions and Future Work	242
7.1 References	245

Author's Publications

Journal Publications:

Campbell, H.I., S. Zhang, S. Restaino, and A. Greenaway, *Generalised Phase Diversity for Wavefront Sensing*. Optics Letters, 2004. 29(23): p. 2707-2709.

Towers, C.E., D.P. Towers, H.I. Campbell, S. Zhang, and A.H. Greenaway, *Three-Dimensional Particle Imaging by Wavefront Sensing*. Optics Letters, 2006. 31(9): p. 1220-1222.

Conference Proceedings:

Greenaway, A.H., H.I. Campbell, and S. Zhang. *Generalised Phase Diversity Wavefront Sensor*. in 4th International Workshop on Adaptive Optics in Industry and Medicine. 2003. Munster, Germany: Springer Proceedings in Physics. Vol. 102. p. 177-186

Zhang, S., H.I. Campbell, and A.H. Greenaway. *Generalised Phase Diversity: Initial Tests*. in 4th International Workshop on Adaptive Optics in Industry and Medicine. 2003. Munster, Germany: Springer Proceedings in Physics. Vol. 102. p. 187-196

Campbell, H.I., S. Zhang, A. Brun, and A.H. Greenaway. *Generalised Phase Diversity Wavefront Sensing*. in AMOS Technical Conference. 2004. Wailea, Maui: Published on CD, contact amostech@maui.com or visit www.maui.afmc.af.mil for a copy

Campbell, H.I., S. Zhang, and A.H. Greenaway. *Extended Wavefront Sensing with Novel Diversity Functions*. in The 5th International Workshop on Adaptive Optics in Industry and Medicine. 2005. Beijing, China: SPIE Proceedings. Vol. 6018. p. Article 601800

Dillon, C.E., H.I. Campbell, S. Zhang, and A.H. Greenaway. *Anisoplanatic Atmospheric Measurements for Terrestrial Imaging*. in The 5th International Workshop on Adaptive Optics in Industry and Medicine. 2005. Beijing, China: Proc. SPIE. Vol. 6018. p. Article 60180W60181

Zhang, S., H.I. Campbell, and A.H. Greenaway. *Small Angle Expansion - a Solution to the Phase-Retrieval Problem Using Generalised Phase Diversity in Pupil Space*. in The 5th International Workshop on Adaptive Optics in Industry and Medicine. 2005. Beijing, China: SPIE Proceedings. Vol. 6018. p. Article 60181C

Towers, C.E., H.I. Campbell, S. Zhang, A.H. Greenaway, and D.P. Towers. *Wavefront Sensing for Single View 3-Component 3-Dimensional Particle Imaging*. in The 6th International Symposium on Particle Image Velocimetry. 2005. Pasadena, California

Chapter 1

Introduction

Summary of Chapter 1

This chapter begins with a brief historical introduction and background information on the field of Adaptive Optics. This is followed by a review of wavefront sensing technologies from historical roots to the current state-of-the-art.

1.1 Introduction

Throughout the course of this thesis a novel, phase-diversity based, wavefront sensor system will be described. It will be developed from its theoretical basis to its initial experimental results and this includes the formulation of a new analytic phase retrieval algorithm to invert the data.

As an introduction to this work, and to put this project into context, this chapter will begin with a brief background section on the field of Adaptive Optics (AO) followed by a more detailed review of wavefront sensing technologies. This is not intended to be a comprehensive study of these topics, but an overview. Relevant references are given throughout to identify more detailed sources of information.

1.2 Adaptive Optics and the Earth's Atmosphere ~ Historical Background

The heavens have held a deep fascination for mankind since ancient times. Items such as the Shou-chang planisphere of the West Han dynasty (52BC) and the Mayan Dresden Codex (1200-1250AD) demonstrate the high level of astronomical professionalism and observational skill of these ancient astronomers [1]. With the invention of the telescope astronomers began to see beyond the limits of the human eye, this marked the beginning of modern astronomy. In 1608 Hans Lippershey demonstrated the first refracting telescope, which Galileo used two years later to discover the rings of Saturn, the four

Chapter 1: Introduction

moons of Jupiter, and many other important phenomena. This was an important milestone in the history of observational astronomy which led to the building of larger telescopes in an attempt to see more of the universe around us. This is a trend which continues today, with the European Southern Observatory's project to build a 100m diameter telescope [2].

However, the quality of observations is not purely limited by the size of the telescope used but by "seeing" – the effects of the Earth's turbulent atmosphere on light which passes through it. Wind driven turbulence mixes stratified temperature changes which varies the density of the atmosphere. The effects of this on light that passes through the atmosphere are threefold. Firstly it creates intensity fluctuations (or scintillation) which is observed as the 'twinkling' of the stars. Secondly the position of the star 'wanders' when the varying refractive index of the atmosphere alters the angle of arrival of the starlight. Lastly, there is a spreading effect created by the higher order aberrations which causes stars to appear as small discs of light and not sharply defined point sources [3]. The turbulent nature of the Earth's atmosphere has been known since the time of Newton who observed in *Opticks*:

"For the air through which we look upon the stars is in perpetual tremor...the only remedy is a most serene and quiet air, such as may be perhaps on the tops of the highest mountains above the grosser clouds" [4]

Newton's observations were correct and the improved seeing conditions at high altitudes is one of the main reasons that many of today's world-class telescopes are built on the tops of mountains. Mauna Kea on the island of Hawaii is a popular location for optical, infrared and radio telescopes and at 14,000ft its summit stands above two-thirds of the Earth's atmosphere. However, this is not the complete solution to the problem. To achieve the best possible results requires further mitigation of the atmospheric effects. Despite Newton's early observations it was not until the 20th century that scientists began to work on the problem of correcting these turbulence induced effects.

In 1953 Babcock commented that:

“One may consider himself fortunate to experience one hour out of 1000 of the finest seeing, even at the best locations” [5]

In his paper Babcock proposed a system to compensate for the blurring effects on stellar images caused by atmospheric turbulence [5]. It is from this paper that the field of Adaptive Optics (AO) emerged.

AO provides a means to perform real-time correction of aberrations imposed on light waves as they travel from the source to the imaging system. While AO has its roots in the field of astronomy it is currently used in a wide variety of medical, military and industrial applications. The papers by Milonni [6] and Parenti [7] provide an excellent introduction to the use of AO in astronomy. A comprehensive review of the medical and industrial applications of AO can be found in the technology tracking report by Greenaway and Burnett [8].

1.3 The Basic Principles of Adaptive Optics

There are two main types of AO systems: ones which provide real-time dynamic correction of unwanted optical aberrations, and ones which can manipulate the properties of an optical system to deliver a desired output. The former is used, for example, in applications in astronomy, ophthalmology, military imaging and confocal microscopy. The latter type of AO system can be used, for example, in the measurement and manipulation of laser beams in laser materials processing. In either case both types of AO system will require the same 3 basic components: a wavefront sensor to measure the shape of the optical wavefront, a corrective/manipulative device to alter the shape of the wavefront and a feedback loop to connect these two components to allow real-time adjustment of the wavefront.

Figure 1.1 is an example of an astronomical AO system which shows how the wavefront sensor, corrective element and feedback loop are arranged.

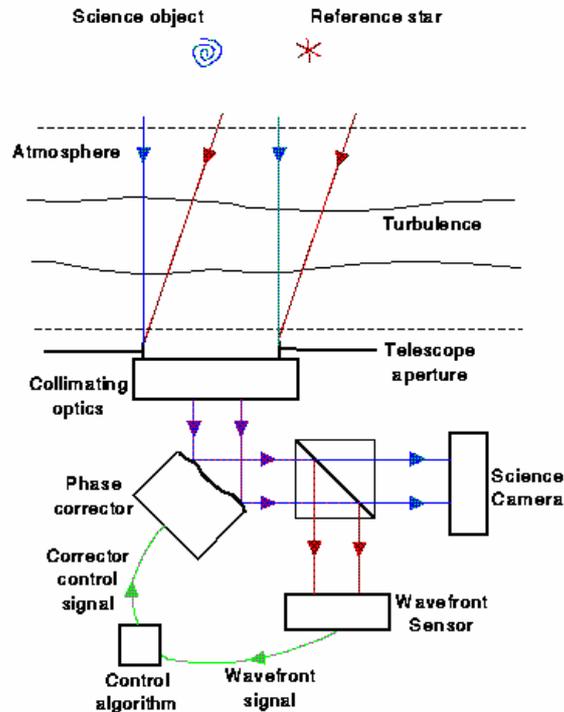


Figure 1.1 A schematic of an astronomical closed-loop AO system [9].

When light enters the system a small amount is taken to the wavefront sensor (via the beam splitter). The wavefront sensor performs an analysis of the aberrations on the wavefront to calculate its shape. This information is sent to the corrective element via the feedback loop. The corrective element flattens the wavefront by a process known as phase conjugation. Using the phase calculated by the wavefront sensor the corrective element applies the complex conjugate of this phase to cancel the distortions on the wavefront. The wavefront sensor continuously monitors the incoming wavefront, sending adjustments to the corrective element as the aberrations change dynamically. In a telescope AO system like the one shown in Figure 1.1 the light used to probe the aberrations comes from a guide star (GS) which can be the observed star, or if it is too dim a bright nearby star or an artificial laser GS is used. A good introduction to the generation and usage of laser guide stars can be found in the article by Parenti [7] and in Tyson's 'Principles of Adaptive Optics' [10].

1.4 Wavefront Sensing

Since the birth of modern Adaptive Optics (AO) in 1953 [5] wavefront sensors have played an important role in the design of AO systems. Wavefront sensing provides the means to measure the shape of an optical wavefront or, in the case of a closed-loop AO system, the deviation of the wavefront from the diffraction-limited case [8]. Depending on the design, wavefront sensors may be used either to generate a signal related to the wavefront deformation, or to provide a full reconstruction of the wavefront shape. In the closed-loop case one seeks to minimise the error signal through manipulation of the wavefront by a corrective element. Full reconstruction of the wavefront is more time consuming, but is sometimes necessary. In metrology applications the shape of the wavefront may represent the shape of a physical surface under test.

The phase/Optical Path Difference (OPD) of a wavefront can be measured directly in a number of ways, which can include; Interferometric methods [10], Fraunhofer diffraction patterns [11], moments of diffraction, multiple intensity measurements [12, 13], or in exceptional cases from a single intensity measurement and the use of *a priori* information [14]. This information can then be fed back to the corrective element. Phase cannot be measured directly, there can be problems with these methods determining the uniqueness of the solution and they often only apply to small angles (i.e. wavefronts which are not greatly aberrated) [15-17]. The indirect approach to wavefront sensing is to compensate for the wavefront error without explicitly calculating the full wavefront reconstruction. The Multi-Dither technique is perhaps the most well known example of an indirect wavefront sensing system [18].

In Zonal sensors the wavefront is expressed in terms of the OPD across a small spatial area, for example a subaperture. If the wavefront is subdivided by N subapertures, then as $N \rightarrow \infty$ the wavefront is fully represented. If wavefront tilts are measured across the subapertures they can be integrated to provide the full wavefront shape. The Shack-Hartmann wavefront sensor is the most commonly used zonal sensor and will be covered in detail later. In Modal sensors the wavefront is described by decomposition into a set of orthogonal polynomials [19-22]. The most common basis set for these sensors is Zernike Polynomials which have the convenient property of forming an

Chapter 1: Introduction

orthonormal set across circular apertures [22]. Curvature sensing and Phase Diversity (PD) are both examples of modal wavefront sensing and are also discussed in some depth in later sections. Tyson proposes [10] that as a general rule, if low order aberrations dominate it is best to use modal wavefront sensing, whereas when higher order aberrations are involved zonal methods will give better performance. For AO in the presence of the atmosphere it is often useful to use a combination of both, employing a zonal method to correct for tilt errors and a modal method to reconstruct the wavefront in terms of Zernike polynomials.

However, for the correct choice of wavefront sensor, practical issues such as hardware limitations, the speed required, computational power available and ease of integration into the existing optical system must be considered. In this chapter a number of different wavefront sensors will be considered; from historically important techniques to the sensors in common use today.

1.4.1 The Shack-Hartmann Wavefront Sensor

The Shack-Hartmann (S-H) wavefront sensor is the most well known and widely used wavefront sensor today. It was developed in 1971 by Roland Shack and Ben Platt [23] in response to the US Air Force's need to improve images of satellites taken from the Earth, which were distorted by the Earth's atmosphere. The S-H sensor is an adaptation of the earlier Hartmann plate method of wavefront sensing [8] and was altered to be more efficient for low-light applications like astronomy.

The S-H is a zonal wavefront sensor which provides a measurement of the local first derivative (slope) of the input wavefront. This is achieved by dividing the wavefront using an array of lenslets, each taking a portion of the beam and focusing it onto a detector array. This use of lenslets was the main innovation on the Hartmann plate design which used an array of holes to separate the wavefront into pencil-like beams. The added light gathering efficiency of the lenslets made the new S-H sensor ideal for photon-starved situations like astronomy [23].

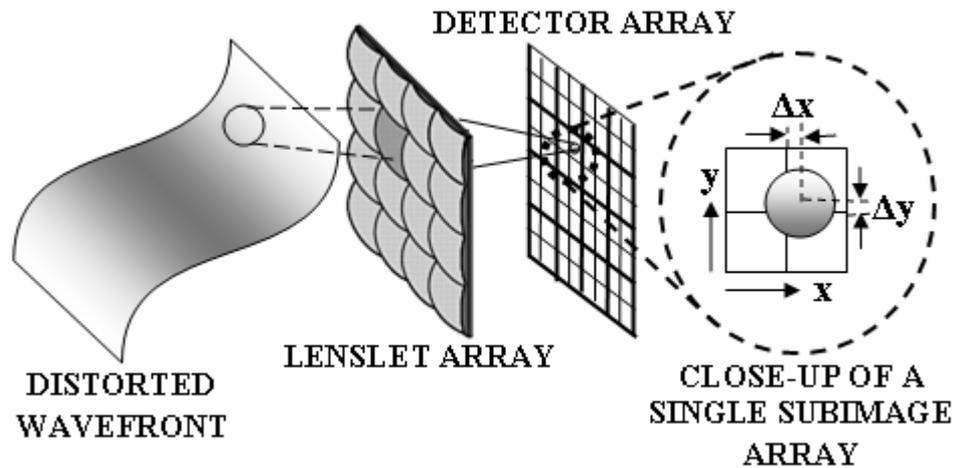


Figure 1.2 Schematic of the basic Shack-Hartmann Wavefront Sensor.

Figure 1.2 shows the basic architecture of the S-H sensor. The detector array is arranged such that a sub-array of pixels is assigned to each subaperture image. In Figure 1.2 this is shown as a simple quad-cell for each lenslet, although more pixels can be used depending on the size of the detector and the number of lenslets. When a plane wave is incident on the lenslet array a tightly focussed spot will appear in the centre of each sub-area of the detector (corresponding to the optic axis of the particular lenslet). When the input wavefront is distorted the sub-images will be shifted in the x and y directions (demonstrated in the close-up in Figure 1.2). The centroid of each image can be used to calculate the slope of the wavefront across its associated subaperture. The wavefront is reconstructed by combining the local slope measurements across the lenslet array. Depending on the type of aberration present in the input wavefront the shape of the sub-images will also be distorted. The S-H is an achromatic wavefront sensor (provided the lenslet array is achromatic) since it will give the same signal for a fixed slope value, irrespective of wavelength.

Figure 1.3 shows examples of simulated Hartmann spot patterns generated by a perfect human eye, and an aberrated one [24]:

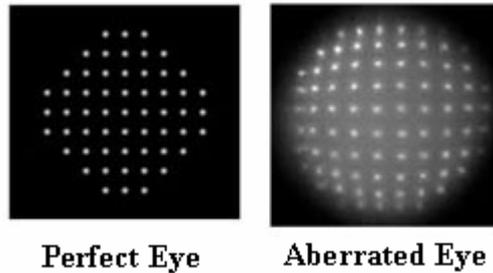


Figure 1.3 Example Hartmann spot patterns for perfect and aberrated eyes. Figure courtesy of David Williams' Laboratory, University of Rochester [24].

The advantages of the S-H sensor are its wide dynamic range, high optical efficiency, white light capability, and ability to use continuous or pulsed sources [10]. The sensitivity and accuracy of the S-H sensor are excellent across the wide dynamic range of the instrument. This makes it a very attractive device where large dynamic ranges are required. As the wavefront aberration increases the sub-images are displaced further from their local optic axis, the dynamic range is limited to aberrations which do not allow the sub-images to be displaced outwith their own sub-array. It is feasible to use time-tracking to observe the sub-images as they move across the detector although this could prove difficult to implement in practice. Therefore the dynamic range of the S-H in most cases is largely limited by the size of the detector. The S-H sensor can also be used to study the intensity and phase information simultaneously by measurement of the spot amplitude and position respectively [25]. One drawback is the precision required in alignment and calibration of this device. Vibration or distortion of the optics could lead to shifts in the spot positions, which in turn would give incorrect measurements of the wavefront slope. There are numerous variations on the basic S-H sensor design to attempt to minimise alignment and calibration problems [26]. A second disadvantage is that this type of wavefront sensor is not well suited to dealing with extended sources. When the object being imaged is large then the shape of the object will be convolved with the diffraction pattern of the subaperture, and optical cross-correlation is required to remove this effect. This takes a lot of computational effort and time [27, 28]. Finally,

Chapter 1: Introduction

the number of pixels required in the detector to create one phase data point is much higher than in curvature sensors [25]. For high spatial resolutions large CCD cameras are required and this can be expensive and adds extra weight to the optical system.

The S-H sensor's white light capability and optical efficiency make it a favourite among astronomers. Also, since astronomy applications mainly involve point sources, the problems encountered with extended sources are less of an issue. This wavefront sensor has also been used extensively in ophthalmic applications. Since the first paper published by Dreher *et al* [29] many researchers have chosen this sensor to improve retinal imaging systems and to map the aberrations of the eye [30-32]. It is also highly suitable for laser testing, and in AO systems [23, 33].

1.4.2 Phase Diversity and Curvature Wavefront Sensing

Phase Diversity (PD) is a phase retrieval algorithm which measures the wavefront phase through intensity images [12, 34, 35]. Gonsalves [34] first proposed PD in the early 1980's as one of a new class of algorithms designed to ensure the uniqueness of the computed phase solution [36, 37]. It is not possible in general to obtain a unique solution from a single intensity image, instead multiple images or *a priori* information must be used to calculate the correct solution [8]. Where multiple images are used these should be captured on a time-scale which is short in comparison to any phase changes. The name 'Phase Diversity' refers to the fact that both images contain a deliberately added (and therefore known) phase term, in addition to the unknown wavefront phase. The wavefront phase can be calculated from the phase diverse intensity images using iterative algorithms. Many such algorithms have been proposed over the years, from versions of the Gerchberg-Saxton [38] to more complicated approaches such as genetic algorithms [16, 39, 40], simulated annealing [40, 41] and neural networks [42]. These solutions are computationally expensive and the time taken to calculate the solution has, in the past, meant that this type of wavefront sensor is not fast enough for real-time AO applications. Curvature Sensing is essentially a variation of the PD method which does not suffer from the same speed limitations as classic PD algorithms.

Roddier first introduced the Curvature sensor in 1988 [43], a special class of PD wavefront sensor which measures the local wavefront curvature using a pair of intensity images with equal and opposite aberration, captured symmetrically about either the image or pupil plane of the optical system (shown in Figure 1.4). The wavefront curvature is mathematically related to the difference of these intensity images by the Intensity Transport Equation (ITE). There exist a great number of algorithms to solve this equation and retrieve the wavefront phase through its curvature [19, 20, 44-48].

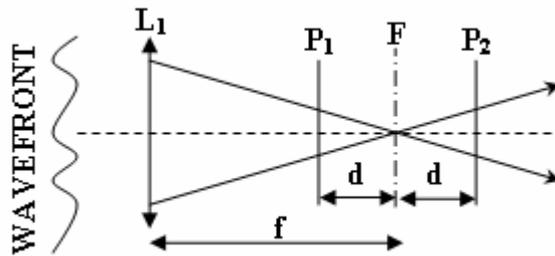


Figure 1.4 A schematic of the Curvature Sensing scheme [43]. Intensity images are obtained on planes P_1 and P_2 symmetrically about the focal plane F .

Figure 1.4 is an example Curvature sensor, as described by Roddier [43]. An input wavefront is focussed by a lens or mirror (labelled L_1), and a pair of intensity images are captured symmetrically about the focal plane (at planes P_1 and P_2). This may be achieved in a number of ways including physical displacement of the image plane [49], use of vibrating mirrors [43], or by beam-splitters and folded paths. Blanchard *et al* described a method of applying defocus phase diversity using a specially designed diffraction grating [50-53]. This method will form the basis of the new wavefront sensor developed in this thesis, and will be covered in detail in Chapter 2.

In Phase Diversity (PD) a pair of intensity images is obtained by adding different amounts of diversity phase (defocus) to the unknown wavefront. This pair could be the in-focus image and one phase-diverse image, or two phase-diverse images. In most cases the diversity phase added is defocus as it may be easily applied by a number of methods, one of which is demonstrated in Figure 1.5. Images recorded symmetrically placed about focus (as shown in Figure 1.4) will contain equal and opposite amounts of

defocus. In this case PD and Curvature sensing are essentially the same technique and the wavefront phase is obtained by solution of the ITE. Where Curvature sensing and PD differ is when the diversity phase used is not defocus, when the phase-diverse images do not contain equal and opposite amounts of defocus, or are not obtained symmetrically about the focus or pupil plane of the system.

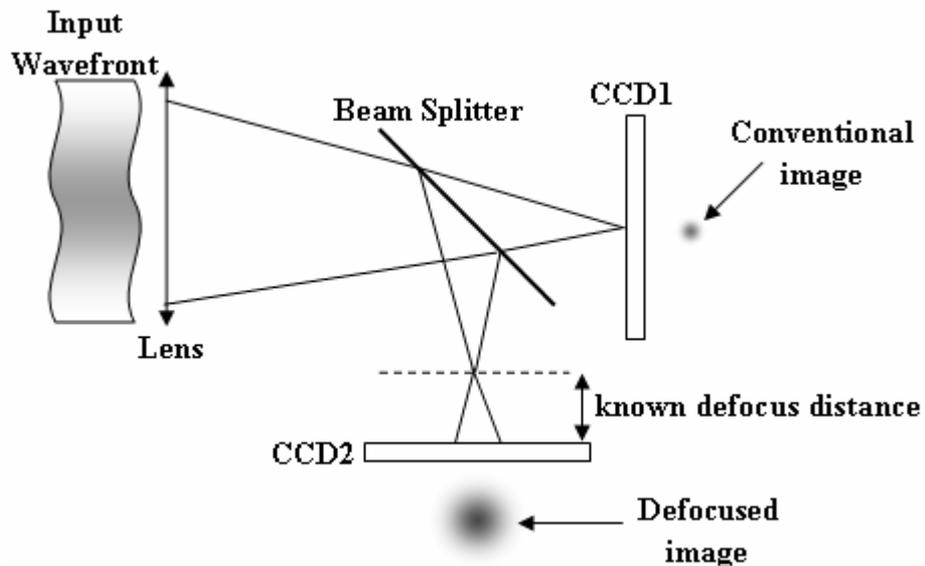


Figure 1.5 An example of a Phase Diversity system. A pair of intensity images are obtained on the CCD cameras, one is the in-focus image and the other is defocused by a known amount.

1.4.2.1 Phase Diversity and Curvature Sensors – Advantages and disadvantages

In comparison to the S-H, Curvature sensing and PD wavefront sensors offer several important advantages. The first is that, since curvature is a scalar field it only requires one sample per point. This means there is a vast reduction in the number of detector pixels required to measure the wavefront, and therefore a large saving in both cost and overall size. Secondly, curvature measurements are more efficient than tilt measurements which are highly correlated [43]. Curvature sensing is better suited to use with extended sources than the S-H for two reasons. Firstly the extended sub-images which would be obtained by the S-H may not be uniform in intensity or overall shape

Chapter 1: Introduction

thus making centroiding more difficult. Secondly, provided the defocus distance is relatively small the error signal produced by the Curvature/PD sensor for an extended source would behave similarly to that for a point source. Therefore, the data analysis for the Curvature/PD sensor need not be altered. However, the sacrifice for operating the Curvature/PD sensor with extended sources is a reduction in the dynamic range. Another advantage of the Curvature and PD sensors is that it's possible to use the signal they generate to directly drive a corrective element (such as a bimorph or membrane mirror [43]). This is much faster, and less computationally expensive, than performing a full reconstruction when it is not strictly needed.

1.4.3 The Paterson-Dainty Hybrid Sensor

In 2000, Paterson and Dainty published their first results with a hybrid wavefront sensor which combines the S-H and Curvature sensors [54]. The configuration of the Hybrid Sensor (HS) is similar to the S-H (see Figure 1.2), the principle difference being the use of astigmatic lenslets to form the lenslet array. The idea is based on the astigmatic focus sensor of Cohen *et al* [55], which measures wavefront curvature using the shape and position of the wavefront's image through an astigmatic lens. The image is detected by a quad-cell detector (see Figure 1.2), when the wavefront is un-aberrated the image will be circular and on axis. If defocus is present in the beam then the image will be elliptical and oriented at 45° to the axis. Therefore, in the HS the shape of each intensity sub-image can be used to measure the local wavefront curvature at that lenslet. The sensor response to the input wavefront aberration will depend on the astigmatic parameters of the lenslet, the quad-cell dimensions and the geometry of the lenslet array [54].

The HS has a limited range over which the defocus and the curvature signal are linearly related, but this range is large enough for closed-loop AO operation [54]. Paterson and Dainty's results have shown that the HS is most sensitive to defocus and spherical aberration. While the S-H (gradient-only) sensor has a wide range of mode sensitivities (and is more sensitive to some modes than others) and Curvature sensors have almost equal sensitivities to all modes, the HS was proven to have improved modal sensitivity compared to either sensor on its own. This is the main advantage of the HS, that whilst

Chapter 1: Introduction

retaining the simple design of the S-H, it has increased modal sensitivity. This may mean that in an AO system the Hybrid Sensor would allow more modes to be properly corrected [54].

1.4.4 Wavefront Shearing Interferometry

Wavefront Shearing Interferometry (WSI), like the S-H, is a wavefront sensing method that can be used to measure the slope of the wavefront. In the 1970's and 1980's the WSI was very much in vogue as a means of wavefront reconstruction [56-58], and found application particularly in the correction of atmospheric distortions on astronomical images, measurement of stellar power spectra, and characterisation of the astronomical 'seeing' [56]. More recently the WSI has been used in ophthalmic research, Licznerski *et al* in 1998 proposed its use in the study of tear-film topography [59]. The tear-film is widely believed to cause a significant amount of variability in wavefront sensing measurements of the eye [60-62]. This is currently a topic of considerable research interest; reducing the variability of the wavefront sensor measurements would be extremely useful in applications such as refractive eye-surgery. Therefore the shearing interferometer, whose use had declined somewhat over the past 20 years, is still an important wavefront sensing technique and its popularity is increasing once more.

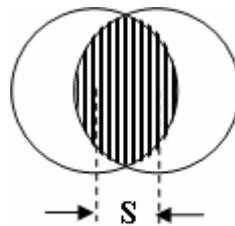


Figure 1.6 A demonstration of lateral shear S between two beams. Interference can only occur where the beams overlap.

In 1947 Bates [63] published a paper that said by tilting the mirrors in a Mach-Zehnder Interferometer by a small amount, and thereby producing a lateral shift between the

signal and reference beams, fringes would only be observed in the overlap region. In the Shearing Interferometer (SI) two copies of the wavefront in the pupil plane are created and one is sheared with respect to the other before placing the interferogram they create onto a 2D detector (see Figure 1.6). By operating in the pupil plane the SI is less sensitive to atmospheric noise than traditional speckle interferometers [57]. There are many different configurations of the SI, which apply the shear a variety of ways the most common being Lateral (LSI), and Rotational (RSI) [64, 65]. In LSI, when the shear distance is small, the interference will depend on the phase difference between points on the wavefront separated by the shear distance. This phase difference, when normalized with the shear distance, provides a measure of the slope of the wavefront (at that point) in the direction of the shear [8]. The wavefront is reconstructed from a pair of interferograms, created by performing the shearing in two orthogonal directions. In this sense it shares something in common with scanning knife edge techniques which will be covered later in section 1.4.7. In RSI, the orthogonal component of the radial shear can be measured by rotating one of the beams by 180° [10]. Another variation on the SI is the radial shearing interferometer where copies of the wavefront with different magnifications are combined co-axially to produce an interference pattern over the area of the smaller diameter beam. Radial shear techniques are particularly useful when the wavefront contains only radial aberrations, for example defocus and spherical aberration [8].

WSI, like S-H wavefront sensors, offer fast computation of the wavefront slopes. There are many variations of the shearing interferometer which have been designed to give white-light capability [66], better operation with extended sources [67], increased optical efficiency [58] and phase closure operation [57]. The WSI remains a versatile wavefront sensing device with applications ranging from Astronomy to Ophthalmology.

1.4.5 The Smartt-Point Diffraction Interferometer

The Smartt or Point Diffraction Interferometer (PDI) was popularised by Smartt in 1972 [68], but was first described by Linnik in 1933 [69]. Like WSI, the PDI is a self-referencing interferometer, largely used in optical shop testing of optical elements and lenses [70]. The common-path design of the PDI makes it compact and versatile. It is

well suited to the testing of large objects, and for use in applications where device size must be kept small, for example in space-borne instruments.

Figure 1.7 shows how the PDI works in principle. It is well known that a point object will diffract light into a perfectly spherical wave. A plate (marked P in Figure 1.7) with a pinhole at its centre is placed into the converging beam from the optical system. The pinhole should be smaller than the size of the Point Spread Function (PSF) of the optical system. This pinhole will thus create a perfect spherical wave, and also allow a portion of the aberrated wavefront to pass through unaltered. The spherical wave and the transmitted part of the wavefront will form an interference pattern on the camera. It is obvious to see that, due to this design, the PDI is optically very inefficient.

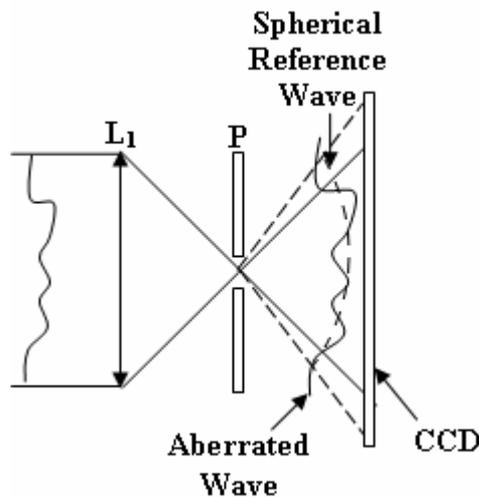


Figure 1.7 Schematic of the PDI. A pinhole mask P, is placed at the focal plane of lens L_1 , creating a spherical reference wave and allowing a small portion of the aberrated wavefront to pass through.

There are a great many variations on the PDI sensor intended to improve its performance in terms of optical efficiency, speed, or convenience of use. An important category of these variations are phase-shifting PDI's. Phase shifting interferometry is the most efficient way of determining the size and direction of wavefront aberration, but the common path design of the PDI makes this a difficult operation to incorporate [71]. It is however possible, Underwood *et al* [72] forced the object and reference beams to

have different polarizations and created a phase shift using an electro-optic modulator, although this was photometrically very inefficient. Kwon [73] described a system with a specially designed diffraction grating to produce 3 interferograms simultaneously, which greatly increased the speed of this method, but also required 3 detectors making it expensive and relatively large. Mercer and Creath [71] demonstrated an interesting phase shifting PDI which uses a liquid crystal layer to introduce the phase shift, with a micro-sphere embedded in its centre to create the spherical reference beam. This device has the advantage of being fully common path and having easily altered variable phase stepping. Love *et al* [74] have also proposed a liquid crystal phase stepping PDI device as a candidate for an Extreme AO (XAO) system, for the imaging of exo-planets and correction of astronomical images from Extremely Large Telescopes (ELTs). Their sensor can be used to give two phase shifted outputs simultaneously, or to drive a phase-only wavefront corrector. It has the added feature of being capable of giving a null output that can be used to calibrate for scintillation effects.

The advantages of using PDI as a wavefront sensor are that its common-path design makes it robust, lightweight and compact and more stable to vibrations and air turbulence. However, the disadvantages are that it is generally photometrically inefficient. Also the addition of phase shifting capabilities adds extra optical elements and therefore extra cost, size and weight thus negating some of the benefits of using a common-path device. While this has not been a greatly popular wavefront sensing technique in the past, in comparison to S-H, PD and Wavefront Shearing sensors, the PDI is still a useful device. The work of Love *et al* shows that it is a viable wavefront sensing option for cutting edge applications like XAO [74].

There are some examples of another variation on common-path interferometry that offer attractive wavefront sensing opportunities. In 2004 Wolfling *et al* demonstrated a generalized wavefront analysis system similar in design to the PDI [75]. In this device the pinhole plate is instead replaced by a phase filter containing a ‘wavefront manipulation function’ which acts over a very small area of the input beam. An iterative algorithm is then employed, which uses minimal approximations, to reconstruct the wavefront. They have shown this to be both fast and accurate and it is intended for use in 3D mapping for metrology applications.

Schlieren imaging, like the S-H, is a technique which involves the division of the input wavefront to gather information about the distortions present in the beam. This technique is a visual process that is used mainly in aeronautical engineering, and for the imaging of turbulent fluid flow. This method is mentioned here, as it provides a good introduction for the knife-edge based wavefront sensors which follow.

There are two main ways in which to implement Schlieren imaging; to study fluid flow or to observe distorted wavefronts. In both cases the image that is formed will be scintillated where the subject has gradients or boundaries, or in the case of fluid flow where there are changes in the density of the medium. Figure 1.8 shows schematically how a Schlieren imaging system works. A distorted wavefront is imaged by a pupil lens and a knife edge is inserted to apodise the focal spot, thus removing certain spatial frequencies from the image. When the truncated image is refocussed onto a CCD camera an image is formed which is bright where the spatial frequencies from the input wavefront were transmitted and dark in regions corresponding to the frequencies blocked by the knife-edge. This variation of the intensity pattern on the camera is essentially a scintillation pattern [76]. Rainbow Schlieren uses a coloured bull's-eye filter instead of the knife edge which allows the strength of the refraction to be quantified [77].

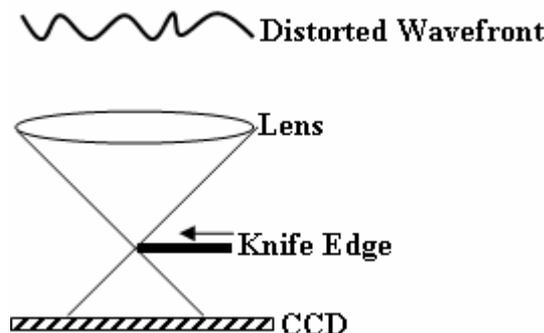


Figure 1.8 Schematic demonstrating the basic principle of Schlieren imaging.

Chapter 1: Introduction

In fluid flow applications a collimated beam is used to illuminate the target object (or area). Where the fluid flow is uniform the intensity pattern will be steady, where there is turbulence a scintillation pattern will be seen. Figure 1.9 shows a particularly demonstrative example of this [78]. In this example the object is a wine goblet, in one case filled with hot water (A) and in the other, ice water (B). The Schlieren images in figure 1.8 clearly show the air turbulence around the goblet.

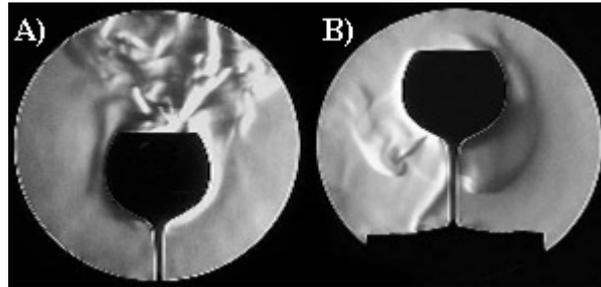


Figure 1.9 A Schlieren photography used courtesy of Prof. Andrew Davidhazy [78] which shows the Schlieren technique being used to study convection. The goblet contains A) hot water and B) cold water.

The principle advantage of Schlieren imaging is that it is a very low cost system to implement, and has high sensitivity. The main disadvantages of the technique are that the field size under study is limited by the sizes of the optical elements, and that it is only a qualitative visualisation process.

1.4.7 Scanning Knife-Edge

The Scanning Knife Edge (SKE) technique is based on the Foucault knife-edge test [79] and is similar in principle to the Schlieren imaging method. In SKE, 2D wavefronts are reconstructed from intensity images which are obtained by scanning 2 knife edges in turn, oriented in orthogonal directions, across the focal plane. As in Schlieren imaging, where the input wave is aberrated, the knife edge will block light entering the pupil plane where the local slope is greater than the value determined by the position of the

knife edge. Combining the data from the scans of the two knife edges allows the wavefront in the pupil plane to be reconstructed from the local slope measurements [8].

Disadvantages of this technique are that diffraction at the knife edge causes blurring in the images making it difficult to distinguish exactly where boundaries lie, and also that the scanning process takes a significant length of time to complete [8]. The latter is by far the most limiting drawback, as it means this method is unsuitable for situations in which the aberrations are dynamic and rapidly changing, for example in ophthalmic applications.

1.4.8 Pyramid or Scanning Prism Knife Edge

The Pyramid Wavefront Sensor (PWS) was proposed by Ragazzoni in 1996 [80] as a new sensor for astronomical applications intended to become a rival of the Shack-Hartmann and Curvature sensors usually used in this field. The principle behind the operation of the Pyramid sensor is similar to the SKE, and also to the sensor described by Horowitz [81] and the concepts illustrated by Sprague and Thompson [82]. Sprague and Thompson demonstrated an early coherent imaging system design which produced an image whose irradiance was directly proportional to the wavefront phase for large phase variations. This was an interesting technique, but involved a time consuming photographic step to create a filter and was therefore not suitable for real-time applications. Horowitz further developed this idea, and instead employed a specially designed filter (in place of the photographic filter) to create an output intensity image which is linear with the derivative (i.e. slope) of the phase function. In the pyramid sensor proposed by Ragazzoni a pyramid is used to fulfil the same role as the filters described by Horowitz, Sprague and Thompson, and this new wavefront sensor also shares characteristics with the modulation contrast microscope [83]. It is an interesting and relatively new type of wavefront sensor which is increasingly popular amongst researchers, but still not as widely used as the S-H or Curvature sensors.

Figure 1.10 shows the basic configuration of the PWS. A prism, or pyramid, is placed in the focal plane of the lens L_1 (which may be the exit pupil of a telescope for example)

so that the incoming light is focussed onto the vertex of the pyramid. The four faces of the pyramid will deflect the portion of the beam incident on them in slightly different directions. The lens relay, simplified and called L_2 in Figure 1.10, is used to conjugate the optical system exit pupil with the focal plane of L_2 where the CCD camera is situated. At the CCD 4 images, one from each face, are seen. The sensor is based on the same quad-cell arrangement as the S-H sensor (Figure 1.2). If the system is un-aberrated and the effects of diffraction are ignored, then the 4 pupil images should be identical [84]. It was mentioned above that the pyramid sensor is similar to the SKE technique, the reason being that by taking the sum of images $a + b$, one exactly obtains the image that would be obtained by a knife edge test. The data required from the orthogonal direction is found by adding $c + d$. In this way the PWS overcomes the main disadvantage of the SKE, as the data is obtained simultaneously and not by scanning the knife edges and it also has better optical efficiency.

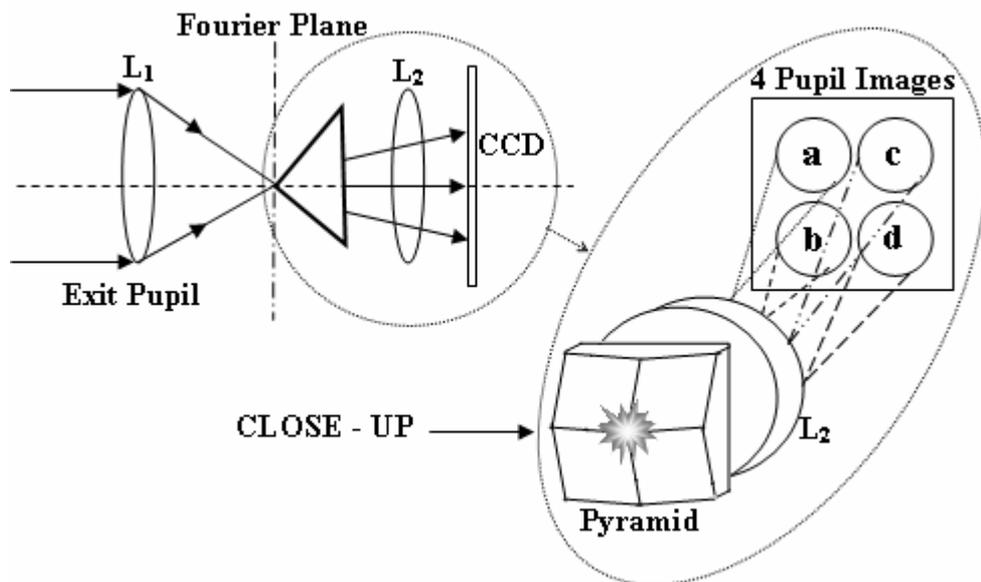


Figure 1.10 The Pyramid Wavefront sensor.

Problems occur if the wavefront slope is so large that all of the incoming light is incident on only one facet of the pyramid. In this situation the signal received is independent of the gradient modulus and the detector quad cell assigned to that facet will be saturated [84]. To avoid this problem Ragazzoni proposed that the pyramid (or the input field) should be oscillated [80, 84].

Chapter 1: Introduction

The advantages of the PWS are that the sampling and the gain of the instrument are easily adjustable. The sampling, which is the size of the pupil on the detector, can be dynamically adjusted using a zoom lens arrangement for the relay L_2 (see figure 1.10), and the gain can be altered by changing the vibration amplitude of the prism [80]. This means that with this sensor it would be possible to perform wavefront sensing on sources of varying brightness and different degrees of aberration [8]. This is important in ophthalmic applications where the eye aberrations of different patients can vary by large amounts. This sensor has been shown to work very well (to estimate the aberrations of the eye) with the human eye as an extended source [84]. This sensor has also been developed for astronomical applications such as the phasing of large telescope mirrors [85] and for Multi-Conjugate AO, where multiple pyramids are used to characterise more than one turbulent atmospheric layer at a time [86].

1.4.9 The Modal Wavefront Sensor

The modal sensor proposed by Neil *et al* [21] in 2000 is the final wavefront sensor which will be considered in this summary. As a modal sensor it does not measure the slope or curvature of the wavefront, but instead directly measures the size of any chosen Zernike mode present in the wavefront [21]. This is a very interesting wavefront sensor, with similarities to the diffraction grating based phase diversity sensor which will be introduced in Chapter 2.

In this modal sensor a diffraction grating is used to produce pairs of spots for each orthogonal mode to be measured. In each spot pair, one image is formed by adding a positive bias and the other by adding a negative bias to the input beam. The pairs of beams are focussed down, and passed through an aperture mask, onto the detector. When the input wavefront is plane there are no offsets in the beams and all beams pass through the aperture mask so that the optical power in each of the spots will be the same. When aberrations are present in the beam the intensity in the spot pairs will vary according to which modes are present, as the beam offsets will affect how much light is passed through the aperture mask.

Chapter 1: Introduction

The modal sensor is of particular use in applications such as confocal microscopy, where the number of aberration modes is relatively few (compared to atmospheric turbulence for example). In this case the modal sensor saves time and effort by not performing a full wavefront reconstruction and instead can be limited to study only the aberrations of greatest interest.

1.4.10 Image Sharpening and Indirect Methods

Indirect wavefront sensing techniques, instead of calculating the wavefront aberration and then applying a correction, seek to apply correction until some chosen error metric has been minimised. This error metric can be any real-time varying quantity that is affected by wavefront aberration. The most common choice is intensity at focus, but image sharpness and scattered field statistical moments are also used [10, 87-91].

Image Sharpening (IS), is an algorithm as opposed to an actual sensor. Historically this technique originates from the first astronomy applications of taking atmospherically degraded images and turning them into sharp images. In its most basic configuration IS works by moving a single phase adjuster at a time and observing the integrated intensity squared in the image plane. Using Parseval's Theorem it is easy to show that this metric (also known as the image sharpness) is maximised when the wavefront is flat. When the maximum is found the phase adjuster is left at this position, another actuator is chosen, and the image is studied in the same way. This can be a time consuming process and uses very little of the image information available. This technique is also only effective on bright point-like objects close to the optic axis, which is why it's used by astronomers for bright stars, but is of limited use elsewhere.

This technique is similar to the multi-dither approach [18], which instead of referring to the intensity at a single point considers the wavefront as a whole. The corrective element is moved and the change to the image sharpness is assessed. If the image sharpness has increased then the change is accepted and the corrective element is moved a little further in the same direction. If however the change is detrimental then the corrective element is moved back and in the opposite direction. This process is repeated

Chapter 1: Introduction

until the wavefront has been flattened. The advantage of this technique is that all of the available light is used, but the disadvantage is that it can be time consuming to arrive at the optimal solution for the corrective element position.

Other indirect methods can include stochastic techniques, which are statistical methods for the effective minimisation of the chosen error metric. These techniques are often formulated to be ‘model-free’ which means that they are independent of the complexity of the optical system used and do not require any *a priori* information about the function under test [90].

Since direct measurement of the wavefront phase can be computationally very expensive, in indirect methods the burden is shifted to the accurate measurement of the chosen quality metric. In applications like astronomy, where one simply seeks to remove the aberrations on the input field, these methods can give faster results than performing a full wavefront reconstruction. In metrology however, where the shape of the wavefront may represent the shape of some test surface, reconstruction is necessary and these methods are of limited use.

1.4.11 Summary of Wavefront Sensing Technologies

Wavefront Sensing is a vast and complex topic which cannot be covered within the scope of this thesis. Therefore the technologies reviewed in this chapter are limited to a selection of historically and currently important wavefront sensor types. The subject of this thesis is the development of a new wavefront sensor; the Generalised Phase Diversity (GPD) sensor. The sensors included in this brief review were chosen to highlight the main competitors to GPD, to describe sensors which share similarities with GPD, and to show where the GPD sensor will offer advantages over currently available technologies.

Figure 1.11 summarises the techniques which have been discussed in this chapter and shows how the technologies are interrelated. Arrows are used to link specific wavefront

sensors, whereas lines are used to group the sensors into categories which share common properties.

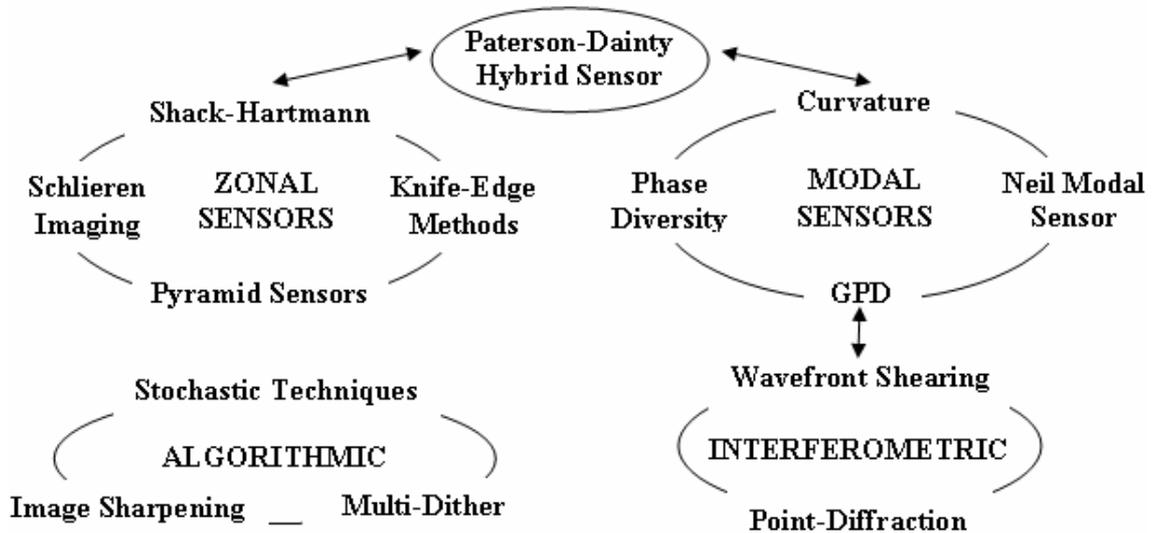


Figure 1.11 Pictorial summary of the wavefront sensing techniques described in this chapter. Arrows illustrate links between specific sensors, lines group the sensors into categories with common properties.

1.5 References

1. Aveni, A.F., *Stairway to the Stars: Skywatching in Three Great Ancient Cultures*. 1st ed. 1997: John Wiley & Sons.
2. Gilmozzi, R., et al. *The Future of Filled Aperture Telescopes: Is a 100m Feasible?* in *Advanced Technology Optical/IR Telescopes VI*. 1998. Kona, Hawaii: Proc. SPIE. Vol. 3352. p. 778-791
3. Lukin, V.P., *Atmospheric Adaptive Optics*. 2nd ed. 1995, Washington: SPIE Optical Engineering Press.
4. Newton, I., *Opticks*. 1704.
5. Babcock, H.W., *The Possibility of Compensating Astronomical Seeing*. Publications of the Astronomical Society of the Pacific, 1953. **65**(386): p. 229-236.
6. Milonni, P.W., *Adaptive Optics for Astronomy*. American Journal of Physics, 1999. **67**(6): p. 476-485.
7. Parenti, R.R., *Adaptive Optics for Astronomy*. The Lincoln Laboratory Journal, 1992. **5**(1): p. 93-113.
8. Greenaway, A. and J. Burnett, *Technology Tracking: Industrial and Medical Applications of Adaptive Optics*. 2004. Institute of Physics Publishing Ltd
9. Tokovinin, A., *Schematic of an Adaptive Optics System*. 2001.

10. Tyson, R.K., *Principles of Adaptive Optics*. 1st ed. 1991, London: Academic Press Ltd.
11. Gonsalves, R.A., *Phase Retrieval by Differential Intensity Measurements*. J.Opt.Soc.Am.A., 1987. **4**(1): p. 166-170.
12. Gonsalves, R.A. *Fundamentals of Wavefront Sensing by Phase Retrieval*. in *Wavefront Sensing*. 1982. San Diego, CA: Proc. SPIE. Vol. 351. p. 56-65
13. Robinson, S.R., *On the Problem of Phase from Intensity Measurements*. J.Opt.Soc.Am.A, 1978. **68**(1): p. 87-92.
14. Dainty, J.C. and M.A. Fiddy, *The Essential Role of Prior Knowledge in Phase Retrieval*. Journal of Modern Optics, 1984. **31**(3): p. 325-330.
15. Burge, R.E., M.A. Fiddy, A.H. Greenaway, and G. Ross, *The Phase Problem*. Proc. Roy. Soc.A, 1976. **350**(1661): p. 191-212.
16. Carmody, M., L.D. Marks, and K.L. Merkle, *Current Transport Along Grain Boundaries in D-Wave Superconductors*. Physica C, 2002. **370**: p. 228-238.
17. Barakat, R. and G. Newsam, *Necessary Conditions for a Unique Solution to Two-Dimensional Phase Recovery*. J.Math. Phys, 1984. **25**(11): p. 3190-3193.
18. Youbin, F., Z. Xuerning, and Y. Jiaxiong, *Research on the Dynamics of a Multidither Adaptive Optical System*. Optical Engineering, 1998. **37**(4): p. 1208-1211.
19. Gureyev, T.E. and K.A. Nugent, *Phase Retrieval with the Transport-of-Intensity Equation. Ii. Orthogonal Series Solution for Nonuniform Illumination*. J.Opt.Soc.Am.A, 1996. **13**(8): p. 1670-1682.
20. Gureyev, T.E., A. Roberts, and K.A. Nugent, *Phase Retrieval with the Transport-of-Intensity Equation: Matrix Solution with Use of Zernike Polynomials*. J.Opt.Soc.Am.A, 1995. **12**(9): p. 1932-1941.
21. Neil, M.A.A., M.J. Booth, and T. Wilson, *New Modal Wave-Front Sensor: A Theoretical Analysis*. J.Opt.Soc.Am.A, 2000. **17**(6): p. 1098-1107.
22. Noll, R.J., *Zernike Polynomials and Atmospheric Turbulence*. J.Opt.Soc.Am.A, 1976. **66**: p. 207-211.
23. Platt, B.C. and R. Shack, *History and Principles of Shack-Hartmann Wavefront Sensing*. Journal of Refractive Surgery, 2001. **17**(5): p. S573-S577.
24. Williams, D.R., *The Shack-Hartmann Wavefront Sensor*. Personal Communication 2005
25. Levecq, X., *Wavefront Sensors Show up Problems with Beams*. Opto and Laser Europe, 2005(131): p. 33-35.
26. Rousset, G., et al. *The Come-on-Plus Adaptive Optics System: Results and Performance*. in *ICO-16 Satellite Conf. on Active and Adaptive Optics*. 1993: ESO Conf. and Workshops Proc. Vol. 48. p. 65
27. Welch, S., A. Greenaway, P. Doel, and G. Love, *Smart Optics in Astronomy and Space*. Astronomy & Geophysics, 2003: p. 26-29.
28. Rao, C.H., W.-h. Jiang, N. Ling, and J.M. Beckers, *Tracking Algorithms for Low-Contrast Extended Objects*. Chinese Astronomy and Astrophysics, 2002. **26**(1): p. 115-124.
29. Dreher, A.W., J.F. Bille, and R.N. Weinreb, *Active Optical Depth Resolution Improvement of the Laser Tomographic Scanner*. Applied Optics, 1989. **28**(4): p. 804-808.
30. Liang, J., B. Grimm, S. Goelz, and J.F. Bille, *Objective Measurement of Wave Aberrations of the Human Eye with the Use of a Hartmann-Shack Wave-Front Sensor*. J.Opt.Soc.Am.A, 1994. **11**(7): p. 1949-1957.
31. Hofer, H., L. Chen, G.Y. Yoon, B. Singer, Y. Yamauchi, and D.R. Williams, *Improvement in Retinal Image Quality with Dynamic Correction of the Eye's Aberrations*. Optics Express, 2001. **8**(11): p. 631-643.

32. Shahidi, M., N.P. Blair, M. Mori, and R. Zelkha, *Optical Section Retinal Imaging and Wavefront Sensing in Diabetes*. Optometry and Vision Science, 2004. **81**(10): p. 778-784.
33. Kudryashov, A., V. Samarkin, A. Alexandrov, A. Rukosuev, and V. Zavalova. *Adaptive Optics for High Powered Laser Control*. in *4th International Workshop on Adaptive Optics in Industry and Medicine*. 2003. Munster, Germany: Springer Proceedings in Physics. Vol. 102. p. 237-248
34. Gonsalves, R.A., *Phase Retrieval and Diversity in Adaptive Optics*. Optical Engineering, 1982. **21**(5): p. 829-832.
35. Jefferies, S.M., M. Lloyd-Hart, E.K. Hege, and J. Georges, *Sensing Wavefront Amplitude and Phase with Phase Diversity*. Applied Optics, 2002. **41**(11): p. 2095-2102.
36. Schiske, P., *Phase Determination from a Focal Series and the Corresponding Diffraction Pattern in Electron Microscopy for Strongly Scattering Objects*. Journal of Physics D: Applied Physics, 1975. **8**: p. 1372-1386.
37. Teague, M.R., *Irradiance Moments: Their Propagation and Use for Unique Retrieval of Phase*. 1982. **72**(9): p. 1199-1209.
38. Gerchberg, R.W. and W.O. Saxton, *A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures*. Optik (Stuttgart), 1972. **35**: p. 237-246.
39. Give'on, A., N.J. Kasdin, R.J. Vanderbei, D.N. Spergel, M.G. Littman, and P. Gurfil. *Stochastic Optimal Phase Retrieval Algorithm for High-Contrast Imaging*. in *Astronomical Adaptive Optics Systems and Applications*. 2003: Proc. SPIE. Vol. 5169. p. 276-287
40. Thust, A., M. Lentzen, and K. Urban, *The Use of Stochastic Algorithms for Phase Retrieval in High Resolution Transmission Electron Microscopy*. Scanning Microscopy, 1997. **11**: p. 437-454.
41. Nieto-Vesperinas, M., R. Navarro, and F.J. Fuentes, *Performance of a Simulated-Annealing Algorithm for Phase Retrieval*. J.Opt.Soc.Am.A, 1988. **5**(1): p. 30-38.
42. Kendrick, R.L., D.S. Acton, and A.L. Duncan, *Phase-Diversity Wave-Front Sensor for Imaging Systems*. Applied Optics, 1994. **33**(27): p. 6533-6546.
43. Roddier, F., *Curvature Sensing and Compensation: A New Concept in Adaptive Optics*. Applied Optics, 1988. **27**(7): p. 1223-1225.
44. Teague, M.R., *Deterministic Phase Retrieval: A Green's Function Solution*. J.Opt.Soc.Am.A, 1983. **73**(11): p. 1434-1441.
45. Van Dam, M.A. and R.G. Lane, *Extended Analysis of Curvature Sensing*. J.Opt.Soc.Am.A, 2002. **19**(7): p. 1390-1397.
46. Van Dam, M.A. and R.G. Lane, *Theoretical Performance of Phase Retrieval on a Subdivided Aperture*. Optical Engineering, 2002. **41**(6): p. 1387-1395.
47. Woods, S.C. and A.H. Greenaway, *Wave-Front Sensing by Use of a Green's Function Solution to the Intensity Transport Equation*. J.Opt.Soc.Am.A, 2003. **20**(3): p. 508-512.
48. Quiroga, J.A., J.A. Gomez-Pedrero, and J.C. Martinez-Anton, *Wavefront Measurement by Solving the Irradiance Transport Equation for Multifocal Systems*. Optical Engineering, 2001. **40**(12): p. 2885-2891.
49. Barty, A., K.A. Nugent, D. Paganin, and A. Roberts, *Quantitative Optical Phase Microscopy*. Optics Letters, 1998. **23**(11): p. 817-819.
50. Blanchard, P.M., D.J. Fisher, S.C. Woods, and A.H. Greenaway, *Phase-Diversity Wave-Front Sensing with a Distorted Diffraction Grating*. Applied Optics, 2000c. **39**(35): p. 6649-6655.

Chapter 1: Introduction

51. Blanchard, P.M. and A.H. Greenaway. *Multiplane Imaging and Wavefront Sensing Using Distorted Diffraction Gratings*. in *Trends in Optics and Photonics. Diffractive Optics and Micro-Optics. Vol.41. Technical Digest. Postconference Edition, 18-22 June 2000*. 2000b. Quebec City, Que., Canada: Opt. Soc. America. p. 250-252
52. Blanchard, P.M. and A.H. Greenaway, *Broadband Simultaneous Multiplane Imaging*. Optics Communications, 2000a. **183**(1-4): p. 29-36.
53. Blanchard, P.M. and A.H. Greenaway, *Simultaneous Multiplane Imaging with a Distorted Diffraction Grating*. Applied Optics, 1999. **38**(32): p. 6692-6699.
54. Paterson, C. and J.C. Dainty, *Hybrid Curvature and Gradient Wave-Front Sensor*. Optics Letters, 2000. **25**(23): p. 1687-1689.
55. Cohen, D.K., W.H. Gee, M. Ludeke, and J. Lewkowicz, *Automatic Focus Control: The Astigmatic Lens Approach*. Applied Optics, 1984. **23**(4): p. 565-570.
56. Roddier, C., *Measurements of the Atmospheric Attenuation of the Spectral Components of Astronomical Images*. J.Opt.Soc.Am.A, 1976. **66**(5): p. 478-482.
57. Roddier, F. and C. Roddier, *Phase Closure with Rotational Shear Interferometers*. Optics Communications, 1986. **60**(6): p. 350-352.
58. Greenaway, A. and C. Roddier, *A 180 Degree Rotation Shearing Interferometer with Increased Optical Efficiency*. Optics Communications, 1980. **32**(1): p. 48-50.
59. Licznarski, T.J., H.T. Kasprzak, and W. Kowalik. *Two Interference Techniques for in Vivo Assessment of the Tear Film Stability on a Cornea and Contact Lens*. in *Proceedings of the 1997 10th Slovak-Czech-Polish Optical Conference on Wave and Quantum Aspects of Contemporary Optics, Sep 16-20 1996*. 1996. Karpacz, Poland: The International Society for Optical Engineering. Vol. 3320. p. 183-186
60. Hofer, H., P. Artal, B. Singer, J.L. Aragon, and D.R. Williams, *Dynamics of the Eye's Aberration*. J.Opt.Soc.Am.A, 2001. **18**(3): p. 497-506.
61. Smirnov, H.S., *Measurement of Wave Aberration in the Human Eye*. Biophys., 1961. **6**: p. 52-66.
62. Liang, J. and D.R. Williams, *Aberrations and Retinal Imaging Quality of the Normal Human Eye*. J.Opt.Soc.Am.A, 1997. **14**(11): p. 2873-2883.
63. Bates, W.J., *A Wave-Front Shearing Interferometer*. Proc.Phys.Soc, 1947. **59**: p. 940-950.
64. Armitage, J.D. and A. Lohmann, *Rotary Shearing Interferometry*. Optica Acta, 1965. **12**: p. 185-192.
65. Ribak, E. *Interferometry Following Adaptive Optics*. in *New Frontiers in Stellar Interferometry*. 2004. Bellingham, WA: Proc. SPIE. Vol. 5491. p. 1154-1159
66. Wyant, J.C., *White Light Extended Source Shearing Interferometer*. Applied Optics, 1974. **13**(1): p. 200-202.
67. Wyant, J.C., *Double Frequency Grating Lateral Shear Interferometer*. Applied Optics, 1973. **12**(9): p. 2057-2060.
68. Smartt, R.N. and J. Strong, *Point Diffraction Interferometer*. J.Opt.Soc.Am.A, 1972. **62**(5): p. 737-737.
69. Linnik, V.P., *A Simple Interferometer for the Investigation of Optical Systems (Article in Russian)*. C.R.Acad Sci. USSR, 1933. **1**: p. 208-210.
70. Malacara, D., *Optical Shop Testing*. 2nd ed. 1991, New York: Wiley.
71. Mercer, C.R. and K. Creath, *Liquid-Crystal Point Diffraction Interferometer for Wave-Front Measurements*. Applied Optics, 1996. **35**(10): p. 1633-1642.

Chapter 1: Introduction

72. Underwood, K., J.C. Wyant, and C.L. Koliopoulos. *Self-Referencing Wavefront Sensor*. in *Wavefront Sensing*. 1982. San Diego, California: Proc. SPIE. Vol. 351. p. 108-114
73. Kwon, O.Y., *Multi-Channel Phase Shifted Interferometer*. Optics Letters, 1984. **9**: p. 59-61.
74. Love, G., T.J.D. Oag, and A.K. Kirby, *Common Path Interferometric Wavefront Sensor for Extreme Adaptive Optics*. Optics Express, 2005. **13**(9): p. 3491-3499.
75. Wolfling, S., N. Ben-Yosef, and Y. Arieli, *Generalized Method for Wave-Front Analysis*. Optics Letters, 2004. **29**(5): p. 462-464.
76. Born, M. and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. 7th (expanded) ed. 2001: Cambridge University Press.
77. Howes, W.L., *Rainbow Schlieren and Its Applications*. Applied Optics, 1984. **23**(14): p. 2449-2460.
78. Davidhazy, A., *Schlieren Photography Principles*. Personal Communication 2005
79. Smith, W.J., *Modern Optical Engineering*. 3rd Edition ed. 2000: SPIE Press and McGraw Hill.
80. Ragazzoni, R., *Pupil Plane Wavefront Sensing with an Oscillating Prism*. Journal of Modern Optics, 1996. **43**(2): p. 289-293.
81. Horwitz, B.A., *Phase Image Differentiation with Linear Intensity Output*. Applied Optics, 1978. **17**(2): p. 181-186.
82. Sprague, R.A. and B.J. Thompson, *Quantitative Visualization of Large Variation Phase Objects*. Applied Optics, 1972. **11**(7): p. 1469-1479.
83. Hoffman, R. and L. Gross, *Modulation Contrast Microscope*. Applied Optics, 1975. **14**: p. 1169-1176.
84. Iglesias, I., R. Ragazzoni, Y. Julien, and P. Artal, *Extended Source Pyramid Wavefront Sensor for the Human Eye*. Optics Express, 2002. **10**(9): p. 419-428.
85. Gonte, F., et al. *Ape: The Active Phasing Experiment to Test New Control System and Phasing Technology for a European Extremely Large Optical Telescope*. in *Advanced Wavefront Control: Methods, Devices, and Applications III*. 2005. San Diego: Proc.SPIE, **5894**, p. 306-317. Vol. **5894**. p. p. 306-317
86. Diolaiti, E., et al. *Some Novel Concepts in Multipyramid Wavefront Sensing*. in *Adaptive Optical System Technologies II*. 2002. Kona, Hawaii: Proc.SPIE, **4839**, p. 299-306. Vol. **4839**. p. p. 299-306
87. Vorontsov, M.A., G.W. Carhart, V. Pruidze, J.C. Ricklin, and D.G. Voelz, *Image Quality Criteria for an Adaptive Imaging System Based on Statistical Analysis of the Speckle Field*. J.Opt.Soc.Am.A, 1996. **13**(7): p. 1456-1466.
88. Polejaev, V.I. and M.A. Vorontsov. *Adaptive Active Imaging System Based on Radiation Focusing for Extended Targets*. 1997: Proc. SPIE. Vol. 3126. p. 216-220
89. Hardy, J.W. *Active Optics: A New Technology for the Control of Light*. 1978: Proc. IEEE. Vol. 66. p. 651-697
90. Cohen, M. and G. Cauwenberghs, *Image Sharpness and Beam Focus Vlsi Sensors for Adaptive Optics*. IEEE Sensors Journal, 2002. **2**(6): p. 680-690.
91. Muller, R.A. and A. Buffington, *Real Time Correction of Atmospherically Degraded Telescope Images through Image-Sharpness*. J.Opt.Soc.Am.A, 1974. **64**(9): p. 1200-1210.

Chapter 2

Phase Diversity Wavefront Sensing

Summary of Chapter 2

In Chapter 1 a basic history and review of wavefront sensors was given. This chapter begins with an in-depth study of one of the techniques mentioned; the diffraction grating-based defocus-only phase diversity (DPD) wavefront sensor. The principles behind the operation of the DPD sensor are the basis of the Generalised Phase Diversity (GPD) wavefront sensor, the development of which forms the core work of this thesis. Results obtained using the DPD sensor with both a Green's function and Modal algorithm to reduce the data will be presented to demonstrate the general properties and accuracy of this wavefront sensing method. Finally, the limitations of this wavefront sensor will be discussed to reveal the motivation behind the development of the GPD wavefront sensor.

2.1 Introduction

In Chapter 1 the principle of Phase Diversity (PD) wavefront sensing was introduced as a phase retrieval algorithm which uses a pair of intensity images recorded with a known and deliberately added phase term. The addition of this phase term, known as the diversity phase, with careful data reduction allows a unique solution for the wavefront phase to be calculated. Traditionally defocus is used as the diversity phase; the phase diverse data may be one in-focus and one defocused intensity image, or a pair of intensity images captured about the image or pupil plane of the optical system. As discussed in Chapter 1, this is relatively simple to achieve using translation of the image plane [1], vibrating mirrors [2], beamsplitters or folded paths. In his original paper on phase diversity wavefront sensing Gonsalves suggested that defocus may not be the best diversity phase to use in all cases [3]. The investigation of the possible use of other diversity functions and the benefits this may entail motivated the development of the

Generalised Phase Diversity (GPD) wavefront sensor which will be discussed in following chapters.

In this chapter the concept of Defocus PD (DPD) will be studied in more depth. This will include a description of a diffraction-grating based method for application of the defocus diversity (referred to as the DPD sensor) and the issue of data reduction to retrieve the phase of the input wavefront. The limitations inherent in the data reduction process will be described to provide motivation for the development of this method for use with other diversity phases. Results will be presented that will serve as a comparison to assess the performance of the Generalised Phase Diversity (GPD) wavefront sensor.

2.2 The Defocus Only Phase Diversity Wavefront Sensor

In Chapter 1 Roddier's Curvature Sensor (CS) was introduced as a means of measuring the curvature of a wavefront by obtaining a pair of intensity images placed symmetrically about the focal plane of an optical system (see Figure 1.4 for details). CS may be considered a special class of PD wavefront sensor in which equal and opposite amounts of defocus are applied to an image of the object. Where CS and PD differ is when the diversity phase used is not defocus, when the phase-diverse images do not contain equal and opposite amounts of defocus, or are not obtained symmetrically about the focus or pupil plane of the system.

When the intensity images do contain equal and opposite amounts of defocus then these images can be thought of as opposite ends of a cylinder or volume at the centre of which is the focal or pupil plane. This provides a convenient physical picture which may be used to understand how the shape of the wavefront is linked to the distribution of intensity in the phase diverse images. Figure 2.1A demonstrates how the intensity in the 2 recording planes can be related to the local curvature of the wavefront. Portions of the wavefront which are locally concave will converge as they propagate to form a smaller, brighter, patch at the corresponding position on Plane 2 than is seen on Plane 1. The opposite is true of portions of the wavefront which are convex. The difference of the

two intensity images, divided by their axial separation, provides an approximation to the axial intensity gradient between the image planes. This approximation will hold if the wavefront curvature is not too strong. If a focal point occurs within the volume an ambiguity arises as it is impossible to know whether light which is diverging at Plane 2 is coming from a focus, or from a convex portion of the wavefront (demonstrated in Figure 2.1B). The maximum allowable distance between the planes will therefore depend on the strength of the wavefront curvature. A highly aberrated wavefront will have greater curvature, therefore concave portions of it will focus strongly and the distance between the sampling planes must be kept small to avoid ambiguity. If the plane separation is kept small to accommodate severe curvature then smaller aberrations will generate only small signals. This therefore provides a limitation on the dynamic range of the system.

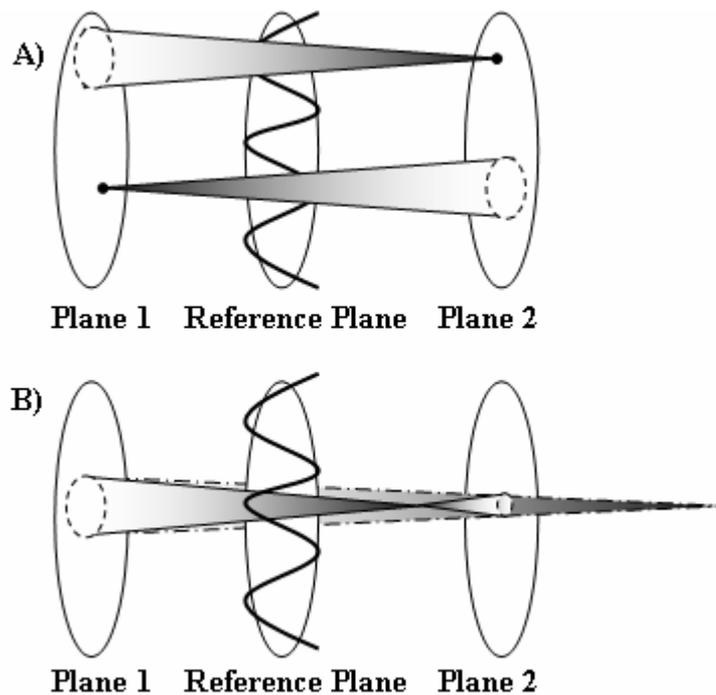


Figure 2.1 A) Schematic showing the relationship between intensity images (formed on planes P_1 and P_2) and wavefront curvature (in the focal, F, or pupil plane). B) Shows the ambiguity caused by a focus between the two planes (the same output could be generated for the two different situations pictured in (B)).

Figure 2.1 summarises the physical description of the CS/DPD principle, the next task is to process the intensity images to retrieve the phase of the wavefront in the centre plane. The relationship between the intensity and the curvature must now be defined mathematically for the data to be inverted.

2.2.1 The Intensity Transport Equation

The Intensity (or Irradiance) Transport Equation (ITE) is a second-order differential equation linking the axial intensity derivative with the slope and curvature of the wavefront which produces the change in intensity as the wavefront propagates [4, 5]. The ITE is given in Equation (2.1):

$$-k \frac{\partial}{\partial z} I_z(\mathbf{r}) = \underbrace{I_z(\mathbf{r}) \nabla^2 \phi_z(\mathbf{r})}_{\text{Curvature}} + \underbrace{\nabla I_z(\mathbf{r}) \cdot \nabla \phi_z(\mathbf{r})}_{\text{Slope}}, \quad (2.1)$$

where $I_z(\mathbf{r})$ is the intensity and $\phi_z(\mathbf{r})$ is the phase of $u_z(\mathbf{r})$, a complex amplitude distribution propagating along the z-axis ($u_z(\mathbf{r}) = (I_z(\mathbf{r}))^{1/2} \cdot e^{i\phi_z(\mathbf{r})}$, $\mathbf{r} = (x, y)$). The wave number k is given by $2\pi / \lambda$ and the Laplacian ∇ represents $\nabla = (\partial/\partial x) + (\partial/\partial y)$, therefore $\nabla^2 = (\partial^2/\partial x^2) + (\partial^2/\partial y^2)$.

Equation (2.1) shows that the ITE contains terms dependent on the first (∇) and second (∇^2) derivative of the phase. Therefore, a measurement of the axial intensity, $I_z(\mathbf{r})$, and its derivative, $(\partial/\partial z)I_z(\mathbf{r})$, should allow the slope and curvature of the wavefront $u_z(\mathbf{r})$ to be calculated using Equation (2.1). In practice the solution of (2.1) is a complex task requiring approximations to separate the terms on the r.h.s.

If the assumption is made that the intensity profile is uniform (i.e. unscintillated) then (2.1) can be simplified to measure only the wavefront curvature [6].

The slope term vanishes (since $\nabla I_z(\mathbf{r}) = 0$) and Equation (2.1) becomes:

$$\frac{-k}{I_z(\mathbf{r})} \left(\frac{\partial}{\partial z} I_z(\mathbf{r}) \right) = \nabla^2 \phi_z(\mathbf{r}) . \quad (2.2)$$

The intensity images described in Section 2.2 (and shown in Figure 2.1), captured symmetrically about the focal or pupil plane, can be used to provide an approximation of the axial intensity derivative in the following way:

$$\frac{\partial}{\partial z} I_z(\mathbf{r}) \approx \frac{I_1 - I_2}{z_1 - z_2} , \quad (2.3)$$

where I_1 is the intensity image recorded in the first plane, at axial position z_1 , and similarly I_2 is the second intensity image recorded at z_2 .

The description and analysis thus far is applicable to both CS and DPD sensors regardless of their optical configuration. It is only when the method of obtaining the intensity images is considered that the sensors' operation may be different.

2.3 The Quadratically Distorted Diffraction Grating

The Fresnel Zone Plate (FZP), shown in Figure 2.2 (a), is a well known optical device which can be used to focus light using diffraction rather than refraction in the case of the classic lens [7]. Unlike a lens the zone plate creates not just one, but multiple copies of the focussed image distributed along the axis of the plate. The FZP consists of a set of radially symmetric rings, or Fresnel zones as they are also known, that (in an amplitude zone plate) alternate between opaque and transparent. Light is diffracted around the opaque zones and then interferes, coming to a focus where constructive interference occurs. The Off-Axis FZP (OAFZP), shown in Figure 2.2 (b), is an interesting extension of this basic Diffractive Optic Element (DOE) which can be used

to image multiple object planes onto a single image plane. In this way it can be used to obtain the pair of defocused images required for DPD wavefront sensing. The basic principle of the OAFZP will be covered in this section to show how this multi-plane imaging is achieved and how it can be used to create a useful PD wavefront sensor.

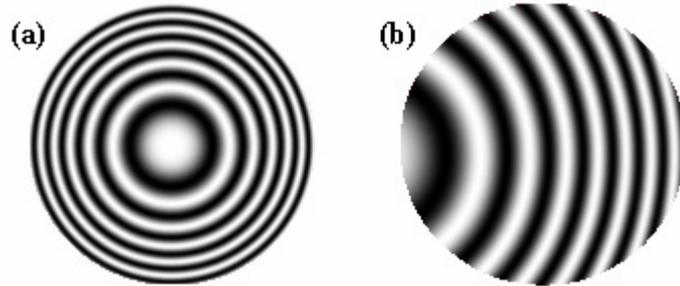


Figure 2.2 (a) An example of a standard FZP (b) an OAFZP.

The curved grating rulings of the OAFZP (see Figure 2.2(b)) apply a phase term to the input wavefront by a process known as detour phase [8]. The added phase shift is different for each diffraction order and can be calculated using the following equation [9]:

$$\phi_m(x, y) = \frac{2\pi m \Delta_x(x, y)}{d}, \quad (2.4)$$

where $\phi_m(x, y)$ is the phase shift applied to diffraction order m in a grating of period d . The term $\Delta_x(x, y)$ is the displacement of the grating rulings with respect to a straight line (i.e. the form of an undistorted, straight-line, grating). This is illustrated in Figure 2.3.

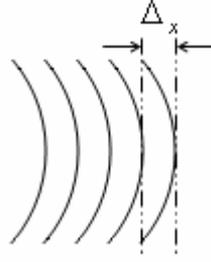


Figure 2.3 Schematic of a set of distorted grating rulings, shifted from the undistorted straight-line grating position by an amount given by $\Delta_x(x, y)$.

For CS and DPD wavefront sensing the grating should be designed to introduce a defocus phase term into the diffraction order images. To achieve this $\Delta_x(x, y)$ should be quadratic in form, as described by Blanchard *et al* [8-11]:

$$\Delta_x(x, y) = \frac{W_{20}d}{\lambda R^2} (x^2 + y^2) , \quad (2.5)$$

where W_{20} is the coefficient of the defocus term (Z_2^0) equivalent to the extra path length introduced at the edge of the aperture by the grating, R is the radius of the grating aperture and x and y are Cartesian co-ordinates with origins on the optic axis. The combination of Equations (2.4) and (2.5) produces an equation for the diversity phase added in each diffraction order of the Quadratically Distorted (QD) diffraction grating:

$$\phi_m(x, y) = m \left[\frac{2\pi W_{20}}{\lambda R^2} \cdot (x^2 + y^2) \right] . \quad (2.6)$$

Blanchard *et al* [8-11] have shown that diffraction gratings designed according to Equation (2.6) can be used to image multiple object planes onto a single image plane. This is due to the focussing property created by adding a defocus phase term to the non-zero diffraction orders. The effective focal length, f_m , that they calculated for each diffraction order is given by (2.7):

$$f_m = \frac{R^2}{2mW_{20}} \quad (2.7)$$

To ensure that each diffraction order has a focal length > 0 the grating must be combined with a lens. This is also more efficient as the lens can perform most of the focussing required so the grating does not require huge levels of distortion to create a high focussing power.

Blanchard *et al* gave the combined power of the QD grating and focussing lens in each non-zero diffraction order as:

$$f_m = \frac{f_L R^2}{R^2 + 2f_L m W_{20}} \quad (2.8)$$

where f_L is the focal power of the lens. Equation (2.7) shows that the ± 1 diffraction orders will have equal and opposite focal power. Figure 2.4 [9] shows how the grating-lens combination may be used to either create multiple images of a single object plane (2.4(a)) or to image multiple object planes onto a single image plane (2.4(b)). It is the latter configuration which is of most use in PD wavefront sensing, and it is this set-up which will be referred to from now on as the DPD wavefront sensor. In both configurations the separation of the multiple image/object planes can be calculated from the following equation:

$$\delta z_m = -\frac{2mz^2 W_{20}}{R^2 + 2mz W_{20}} \quad (2.9)$$

where δz_m is referred to as the ‘defocus distance’, and z is the distance from the central image/object plane to the primary/secondary principle plane of the optical system [9].

The DPD wavefront sensor, shown in Figure 2.4(b), can be used to provide the necessary data to solve the ITE and retrieve the phase of the wavefront under test. In Section 2.3.2 some examples of typical data are shown and their properties discussed. Before continuing to look at the data and methods for the solution of the ITE there are a few final points to be discussed concerning the QD grating and DPD sensor.

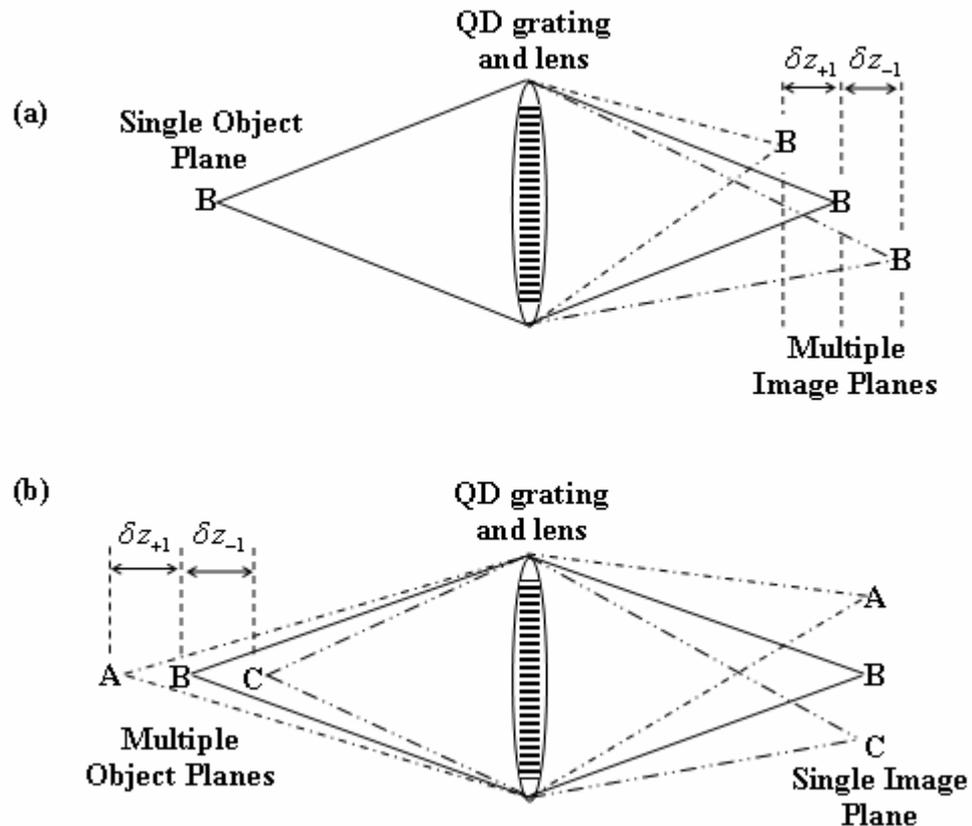


Figure 2.4 The QD grating and lens combination is used to (a) image a single object plane onto multiple image planes or (b) image multiple object planes into a single image plane [9].

Firstly, although all of the figures in this section have suggested that the QD grating is a binary amplitude grating it is possible to manufacture phase gratings using the same principles. The mathematical description of the grating properties and operation in this section are unchanged when using a phase grating. The main advantage of a phase grating is that all the available light is used making it far more photometrically efficient.

Secondly, although Figure 2.4 shows the QD grating and lens combination as being coincident this does not have to be the case. The combined focal length of the grating and lens is dependent on their separation as are the positions of the image and the 1st Principle Plane (PP). The combination focal length is easily derived by considering the grating as another lens (with focal length of each non-zero order given by f_m see Equation (2.7)), and is shown in Equation (2.10):

$$f_c = \frac{f_L \cdot f_G}{f_L + f_G - s}, \quad (2.10)$$

where f_c is the combined focal length of the lens (focal length f_L) and grating (focal length $f_G = f_m$), their separation is given by s . The positions of the 1st and 2nd PP for the ± 1 diffraction orders can be tuned by manipulating the separation s [12]. Their separation is equal when the grating and lens are coincident (as in Figure 2.4) or separated by f_L . When this is the case the system is telecentric [13] meaning that the magnification in the ± 1 diffraction orders is equal.

Finally, the common path design of this wavefront sensor coupled with its simplicity make this a very robust, compact wavefront sensor which does not suffer from the same alignment issues as the Shack-Hartmann for example (see Chapter 1). It requires fewer detector pixels than the Shack-Hartmann and the error signal is formed simply by taking the difference of the intensity images in the ± 1 diffraction orders.

2.3.2 Examples of Data

The phase diverse data is formed by the pair of intensity images in the ± 1 diffraction orders. Some examples of the data which can be seen in the image plane are given in Figure 2.5(a)-(e). The size, shape, and distribution of the intensity seen in these images is directly affected by the aberrations present in the wavefront. This is similar to the sub-images in the Shack-Hartmann wavefront sensor whose size and shape is related to the local wavefront aberration across the lenslet which formed them. However, Shack-

Hartmann spots are typically a small number of pixels in diameter which makes this effect difficult to observe.

If the input wavefront to the DPD sensor is plane (i.e. has constant phase and therefore no curvature) then the images in the ± 1 diffraction orders will be identical in size and intensity, provided the grating has not been blazed or has a variable substrate. Therefore, for a plane wave input, the difference between the two images will be zero. Based on this fact the DPD could be operated as a simple null sensor; that is a wavefront sensor which only provides a signal when the input wavefront is aberrated. By simply looking at the difference between the intensity images, using a multi-dither approach [14], a corrective element could be driven to reduce the difference between the images until the wavefront is flat.

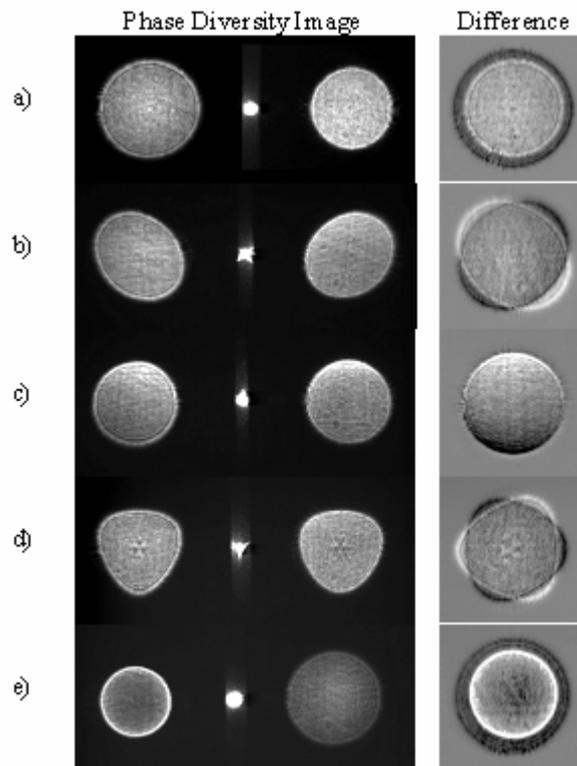


Figure 2.5 Examples of the data produced by the QD grating and lens which show the characteristic features of the following common aberrations (a) Defocus (b) Astigmatism (c) Coma (d) Trefoil (e) Spherical Aberration [11].

Note that for phase diverse images captured about the focal plane (as in Figure 1.4) one image is effectively formed before focus and is therefore a virtual image which must be rotated by 180° before the difference of the two intensity images is taken. Figure 2.5 gives examples of the DPD sensor's output when the input wavefront is aberrated. As these figures show the aberrations present in the wavefront produce easily recognisable features within the data that allow the user to see, with the naked eye, which aberrations are causing the greatest problems.

As Equation (2.7) showed, the QD diffraction grating has positive focussing power in the +1 diffraction order and an equal negative focussing power in the -1 order. Therefore if a spherical wavefront is incident on the grating then the focussing power of the wavefront and the grating will add in one diffraction order and partially cancel in the other. The visual result of this is that, in the ± 1 orders, one image will be smaller and (due to energy conservation) brighter than the other. This is shown in Figure 2.5(a), and demonstrates how easily a defocus term in the input wavefront can be seen simply from a difference in size of the two intensity image spots. Similarly astigmatism (2.5(b)) and trefoil (2.5(d)) distort the overall shape of the intensity images in a distinctive manner. Coma (2.5(c)) and spherical aberration (2.5(e)) affect the distribution of intensity in the pair of images. When coma is present in the input wavefront the intensity spots contain a bright crescent shaped edge (see 2.5(c)). This can cause problems when attempting to locate the centre of the spot using a centre of mass calculation since the increased intensity at the edge will weight the centre calculation towards that edge. Spherical aberration in the input wavefront is identified by a bright spot in the centre of one diffraction order image and a bright ring around the edge of the other (see 2.5(e)). This is a useful fact for the experimentalist as it provides a simple visual method to determine whether any of the achromat lenses in the optical set-up are the wrong way round. Achromats, which typically have one plane and one convex face, should be oriented with the plane face towards the closest conjugate (e.g. the closest diverging source). If the lens is placed the wrong way round this induces extra spherical aberration in the wavefront.

2.4 Data Reduction

It has been shown that a QD diffraction grating can be used to provide a pair of intensity images in the ± 1 diffraction orders which contain an equal and opposite defocus phase diversity term. In Section 2.2.1 it was discussed that the difference of these intensity images could be used to form an approximation of the axial intensity derivative which in turn can be used in the ITE to solve for the curvature of the input wavefront. The ITE, shown in Equation (2.2), is a second order differential equation and as such there are many ways in which to solve it for the unknown phase $\phi_2(\mathbf{r})$ [5, 6, 15-18]. In this section two different methods for the solution of the ITE will be discussed briefly and their relative merits outlined.

2.4.1 The Green's Function Solution

The Green's function solution was originally proposed by Teague [17]. Different choices of boundary conditions have resulted in there being different Green's function solutions to the ITE [6, 17], the algorithm used within the course of this project was the version proposed by Woods and Greenaway [6] which uses Neumann boundary conditions. The use of Neumann boundary conditions was studied by Roddier [19] and has the advantage that the slope of the phase at the boundary can be measured by intensity measurements at this boundary [5, 19].

In their paper Woods and Greenaway [6] proved that the wavefront phase, $\phi(\mathbf{r})$ can be calculated to within an arbitrary additive constant using the following equation:

$$\phi(\mathbf{r}) = \int S(\mathbf{r}')G(\mathbf{r},\mathbf{r}')d^2\mathbf{r}' , \quad (2.11)$$

where $G(\mathbf{r},\mathbf{r}')$ is the Green's function. $S(\mathbf{r}')$, given in Equation (2.12), is found by experimental measurements:

$$S(\mathbf{r}') = - \left[\frac{k}{I_0} \right] \cdot \frac{\partial}{\partial z} I_z(\mathbf{r}') , \quad (2.12)$$

where k is the wavenumber ($2\pi/\lambda$), I_0 is the constant value of the intensity within the uniformly illuminated aperture under consideration and $(\partial/\partial z)I_z(\mathbf{r}')$ is the axial intensity derivative approximated by the difference of the intensity images in the ± 1 diffraction orders. In this way the Green's function algorithm is able to retrieve the phase of the wavefront and using this it performs a modal decomposition into a chosen set of orthonormal basis functions. Strictly speaking the algorithm uses a 4-D function to solve for the phase at each point on a 2D plane.

If the test wavefront is considered to be the linear combination of terms ($u_i(\mathbf{r})$) from an orthonormal series (Zernike Polynomials for example), with the size of each term given by its coefficient a_i , Woods and Greenaway have shown:

$$a_i = \int \phi(\mathbf{r}) u_i(\mathbf{r}) d^2 \mathbf{r} . \quad (2.13)$$

Substituting Equation (2.11) into (2.13) gives $a_i = \int S(\mathbf{r}') G'(\mathbf{r}, \mathbf{r}') d^2 \mathbf{r}'$ where $G'(\mathbf{r}, \mathbf{r}') = \int G(\mathbf{r}, \mathbf{r}') u_i(\mathbf{r}) d^2 \mathbf{r}$. Therefore, $G'(\mathbf{r}, \mathbf{r}')$ is a matrix which can be computed in advance and the phase retrieval and decomposition into Zernike polynomials is achieved using a simple matrix multiply with the measured data $S(\mathbf{r}')$ [6]. This solution is therefore a fast and efficient way to retrieve a modal solution for the input wavefront.

In the course of formulating this solution the assumptions that have been made are that the illumination is uniform, and that the wavefront and its slope are continuous. These assumptions and the implications that arise from them will be discussed in more detail in Section 2.7.

2.4.2 The Gureyev-Nugent Modal Solution

The second algorithm that was used during the course of this project for phase retrieval through solution of the ITE was based on the algorithm proposed by Gureyev and Nugent (the GN algorithm) [5, 15]. This algorithm and the Green's function solution outlined in Section 2.4.1, although they both solve the ITE, achieve this in quite different ways.

The GN algorithm solves the ITE by decomposing the measured data and the solution for the wavefront phase into a series of orthogonal polynomials. The solution provided by the GN algorithm is already expressed in the chosen polynomial series and does not have to be converted using a separate step as in the Green's function solution (Equation (2.13)). This means that when using the GN algorithm, the choice of the area over which to compute the polynomial set (the unit disc) is critical to the accuracy of the solution. Performing the phase retrieval in a single modal decomposition step means there is no way to extract the phase information if the unit disc has been wrongly defined. This will be discussed in more detail later.

To derive the GN algorithm Gureyev and Nugent expressed the ITE in a slightly different way:

$$k \frac{\partial}{\partial z} I_z(\mathbf{r}) = -\nabla \cdot (I_z(\mathbf{r}) \nabla \phi_z(\mathbf{r})) . \quad (2.14)$$

A 2D disc function, called Ω , is defined as a finite sized illuminated aperture and the intensity function $I_z(r)$ is assumed negligible outwith this area. Thus Equation (2.14) need only be solved over the area of Ω to retrieve the phase $\phi_z(r)$ [5]. Multiplying both sides of (2.14) by $Z_j(r/R, \theta)$, which forms an orthonormal set over Ω (the unit disc), with radius R and integrating over all of Ω they derive Equation (2.15):

$$\frac{1}{R^2} \int_0^{2\pi} \int_0^R F Z_j r dr d\theta = -\frac{1}{R^2} \int_0^{2\pi} \int_0^R \nabla \cdot (I_z \nabla \phi) Z_j r dr d\theta , \quad (2.15)$$

where $F = k(\partial/\partial z[I_z])$ which is the measured data formed by taking the difference of the two intensity images in the ± 1 diffraction orders. The Left Hand Side (LHS) of Equation (2.15) is the expression for the j^{th} Zernike coefficient of F_j (the measured data decomposed into Zernike polynomials). Using this substitution, decomposing the phase solution into a series of Zernike terms and integrating by parts they derive:

$$F_j = \sum_{i=1}^{\infty} \phi_i R^{-2} \int_0^{2\pi} \int_0^R I \nabla Z_i \cdot \nabla Z_j r dr d\theta . \quad (2.16)$$

Defining a new matrix, $M_{ij} = \int_0^{2\pi} \int_0^R I(r, \theta) \nabla Z_i(r/R, \theta) \cdot \nabla Z_j(r/R, \theta) r dr d\theta$ this becomes:

$$F_j R^2 = \sum_{i=1}^{\infty} M_{ij} \phi_i, j = 1, 2, 3, \dots, . \quad (2.17)$$

To implement this algorithm in practice Gureyev and Nugent said that the reconstruction should be truncated to N modes (instead of $i, j = 1, \dots, \infty$). They also state that since the GN algorithm solution is only unique to within an additive constant that the retrieval of the piston term (itself a constant) could have no meaning and therefore the solution of Equation (2.17) is calculated for $i, j = 2, 3, \dots, N$ [15]. In their paper [15] they proved that the matrix M_{ij} is always invertible so that the final solution computed by the GN algorithm is given by Equation (2.18):

$$\phi_N = N_F \mathbf{M}_{(N)}^{-1} F_{(N)} , \quad (2.18)$$

Chapter 2: Phase Diversity Wavefront Sensing

where N_F is a scaling factor given by $N_F = (2\pi R^2 / \lambda \delta z)$, R is the radius of the unit disc (defined by the user), and δz is the defocus distance defined as the total distance between the 2 planes in which the intensity images are recorded. $F_{(N)}$ is the difference between the intensity images recorded in the ± 1 diffraction orders. If the input illumination is constant the matrix $\mathbf{M}_{(N)}^{-1}$ can be pre-computed so that solution of Equation (2.18) becomes a simple matrix multiply (as is the case with the Green's function solution). However, uniform illumination is not a prerequisite for the GN algorithm and it is therefore suitable for use with scintillated wavefronts. It just requires some extra computational effort to calculate $\mathbf{M}_{(N)}^{-1}$.

2.4.3 Comparing the Properties of the Green's Function and GN Algorithm

The details presented in Sections 2.4.1-2 have shown that both the Green's function and GN algorithm should be capable of providing results for fast modal wavefront sensing. Both are solutions of the ITE, and both use the same data (the difference between the spots in the ± 1 diffraction orders) to compute the solution. There are however advantages and disadvantages of each in comparison with the other.

The Green's function solution, while requiring a lot of work initially to define the Green's function matrix, is a very fast algorithm to use in practice. However, assumptions made in the formulation of this solution mean that it is designed for use with uniform illumination (unscintillated wavefronts) and that the wavefront and its slope must be continuous. These are assumptions that place limitations on the applications that this algorithm is suitable for (discussed in Section 2.7). The GN algorithm on the other hand is specifically formulated to be capable of dealing with non-uniform illumination [15] and therefore has one advantage over the Green's function solution.

In the GN algorithm the output is intrinsically limited to be a modal decomposition of the wavefront, in this case into Zernike polynomials. As is the case for any algorithm that performs this decomposition, the accuracy is heavily reliant on the accurate choice

of R , the size of the unit disc, which in practice is given by the boundary of the spot image in each diffraction order. Therefore the GN algorithm is particularly prone to human error in the judgement of where the boundary of the data is found. A difference of just a few pixels will significantly alter the values of the Zernike coefficients retrieved. Even using a computer automated process for determining the boundary will result in some level of error (although this is likely to be less than the human error if the process is carefully designed). While the Green's function algorithm has been designed to output the wavefront shape in terms of Zernike coefficients it would be possible to alter the algorithm to output just the phasemap of the wavefront thus avoiding the difficulty of accurately defining the unit circle. Therefore this is one advantage that the Green's function has over the GN algorithm.

In conclusion it would seem that both algorithms are good solutions to the ITE, their basic properties suggesting that the GN algorithm may be more versatile but the Green's function solution could be more accurate. Both these algorithms were used during the course of this project and results will now be presented in the next section to allow a comparison of the accuracy of each algorithm.

2.5 Accuracy

In later chapters the details of the new GPD wavefront sensor will be given (in Chapter 3) and a new algorithm for phase retrieval with GPD data will be described (in Chapter 4). The GPD sensor coupled with the new algorithm will be suitable for use with, but not limited to, defocus as the phase diversity function. The new algorithm will therefore be a non-ITE based solution for the DPD wavefront sensor. As such it will be interesting to be able to make comparisons between the two ITE methods described in this chapter and the new algorithm. In this section the accuracy of the retrieved phase from the Green's function and GN algorithm will be studied to allow for useful comparison later.

2.5.1 Accuracy of the Green's function solution

The best recorded accuracy for this particular Green's function solution was presented by Djidel and Greenaway [20]. They measured the relative curvature of a spherical wavefront using the DPD wavefront sensor to an accuracy of $\lambda/900$, using a Helium-Neon (HeNe) laser source with $\lambda = 632.8\text{nm}$ this is an accuracy of 7\AA . The spherical wavefront was created using a single-mode fibre as a point source mounted at a distance of between 5-7m from the DPD sensor. The sensor was initially positioned such that the object (the end of the fibre) was sharply focussed in the 0th order, at this 'focal position' the defocus coefficient will be zero. Relative curvature measurements were made by monitoring the change in the defocus coefficient as the source was translated about this focal position.

The quoted accuracy of the curvature measurement is the root mean square (rms) fit of the data to a quadratic best-fit, and only the better half of the data set is considered. Taking the whole sequence of data into account gave an rms fit of $\lambda/500$ to the best-fit curve, which is an accuracy of 12.7\AA in the surface measurement [20].

The Green's function software is the intellectual property of QinetiQ (formerly DERA) and as such access to the algorithm was not available, except for a brief period at the outset of this project. During this time the main purpose of the research was to investigate how well the DPD sensor could cope with extended scene imaging since this is a fundamental advantage of this method over the Shack-Hartmann for example. Experiments were conducted to test the accuracy of the defocus measurement when placing diffusers or pinholes in front of an extended source (a multi-mode fibre bundle with white light and a red filter). The intent was to apodise the source or image to simulate what would happen if this technique were used to select a region of particular interest in an extended scene. Using the diffusers (pieces of ground glass) placed directly against the source created smooth illumination from the fibre bundle which would otherwise look like a segmented source. The content of these findings is not relevant to the rest of this thesis and will not be discussed here in any detail. However, there were some experiments conducted which measured the change in the defocus coefficient as the single mode fibre point source was translated about the 'focal

position'. In this context the focal position' is taken to be the object distance at which the wavefront sensor will sense a plane wave. Figure 2.6 is a plot of the results from one of these experiments, it shows the change in defocus coefficient as a function of the source displacement.

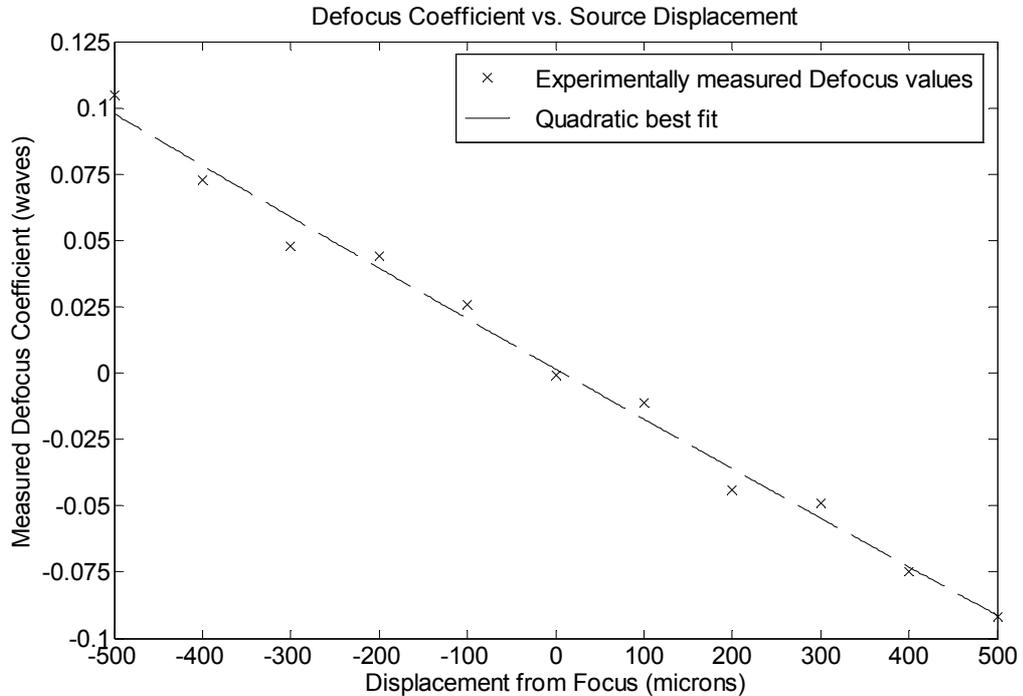


Figure 2.6 The experimentally measured defocus coefficients (using the Green's function algorithm) fitted to a quadratic best-fit curve.

The rms error of the fit of the data to the quadratic best-fit was calculated to be $\lambda/167$, giving an accuracy in the measurement of the surface shape of approximately 3.8nm. This result is much less than the highest reported accuracy of this algorithm, there are 3 main reasons which could help to explain this. The first is that the experimental setup used for the Djidel-Greenaway experiment consisted of only a diverging source, a mirror to fold the beam, and the wavefront sensor. To obtain the results presented in Figure 2.6 a pupil lens was used to bring the light to an aerial focus (and magnify the beam), the light then diverged from this point and was collected by a collimating lens before reaching the wavefront sensor. This increase in the number of intervening optical elements will have added extra error to the measurement. The second reason is

that the total path length of the propagation in the Djidel-Greenaway experiment was more than double the propagation distance used in Figure 2.6 (5m compared to 2m). The change in the wavefront sag is directly proportional to the reciprocal of the propagation distance squared, therefore the greater this distance the smaller a change in the sag which can be measured. Finally, in their experiment Djidel and Greenaway used a 150mm focal length lens to image the diffraction orders onto the camera. This created images 90 pixels in diameter, 3x larger than the images used in the experimental results in Figure 2.6. For very small changes in wavefront curvature the change in intensity between successive measurements is easier to detect than the change in spot size, which may be less than 1 pixel. Therefore the increased image size used by Djidel and Greenaway allows an increase in accuracy due to better sampling and sensitivity to intensity changes.

The proprietary nature of the Green's function software has made it impossible to repeat the accuracy experiment of Djidel and Greenaway. Therefore for comparison purposes their published result of 7\AA (or $\lambda/900$) will be taken as the accuracy of this algorithm.

2.5.2 Accuracy of the GN algorithm

After the version of the GN algorithm which was used in this project had been created an experiment was set up to test its accuracy in as similar a set-up to the Green's function accuracy experiment as possible. A single mode fibre was mounted on a computer controlled translation stage to be used as a point source. A QD grating, with a focal length of $\pm 4\text{m}$, was mounted together with a 60mm focal length achromat, to form the DPD sensor. A CCD camera was positioned after the grating-lens combination so that the point source (seen in the 0^{th} order) was sharply focussed. The point source was then translated about this focal position and the intensity images in the ± 1 diffraction orders were recorded.

These images were then analysed using the GN algorithm and the defocus coefficient was measured. Instead of fitting the measured defocus coefficient (equivalent to the sag of the spherical wavefront) to a best fit through the data (as in the Green's function

accuracy experiment [20]), the GN solution data was compared to the theoretical sag value as calculated from Equation (2.19):

$$s = -\frac{r^2 \Delta z}{2z^2} + \frac{r^2 (\Delta z)^2}{2z^3} - \dots, \quad (2.19)$$

where s is the sag produced by the displaced source, r is the radius of the lens, Δz is the small axial shift of the source from focus and z is the distance between the object and the lens when the source is in focus. Higher order terms than the ones shown in Equation (2.19) have been neglected. Figure 2.7 is a plot of the calculated sag from Equation (2.19) and the measured defocus values (from the GN algorithm). The focal distance (z) was found to be 783.5 mm from the lens, this distance is marked on the graph by the black dotted lines.

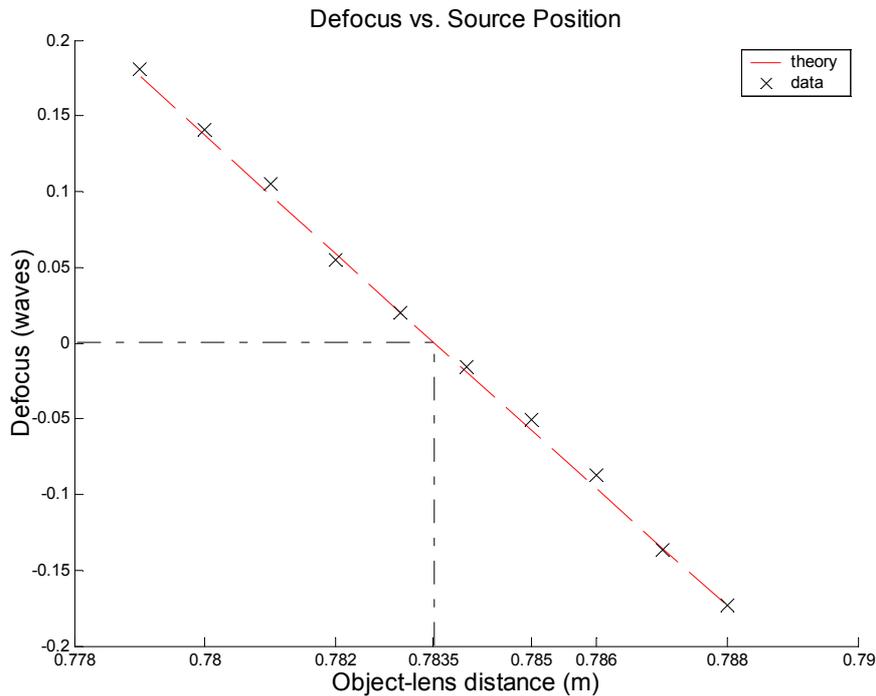


Figure 2.7 The experimentally measured defocus coefficients (using the GN algorithm) fitted to the theoretical straight line calculated using Equation (2.19).

Chapter 2: Phase Diversity Wavefront Sensing

As Figure 2.7 clearly shows, use of the GN algorithm has given phase retrieval results which are very close to the theoretical values. The total rms error between the calculated and measured defocus values is 0.005257 waves which corresponds to an accuracy of $\lambda/190$. At the HeNe wavelength used (632.8nm) this is an accuracy in the relative curvature measurement of 3.3nm.

To make a better comparison with the accuracy measurements of Djidel and Greenaway [20] the quadratic best-fit of this data was calculated and the rms fit of the data to this curve was calculated. Fitting to the best-fit curve, over the whole dataset, results in an improvement of the Nugent algorithm accuracy from $\lambda/190$ to $\lambda/248$ (3.3nm to 2.5nm). Fitting to the same curve, using only the better half of the dataset (as Djidel did) gives a further improvement to $\lambda/364$ (1.7nm).

2.5.3 Comparison of the Green's function and GN algorithm accuracies to the Shack-Hartmann wavefront sensor

Diameter	Focal Length	Array Size	Sensitivity	Dynamic Range
252 μm	25 mm	25 x 19	$\lambda / 150$	30 λ
198 μm	15.5 mm	31 x 24	$\lambda / 100$	40 λ
144 μm	8 mm	44 x 33	$\lambda / 50$	50 λ
108 μm	4.6 mm	58 x 44	$\lambda / 30$	70 λ
72 μm	2 mm	88 x 66	$\lambda / 10$	120 λ

Table 2.1 A comparison of several “Off-The-Shelf” Shack Hartmann wavefront sensors [21].

In Chapter 1 the Shack-Hartmann was introduced as currently the most popular and widely used wavefront sensor. Table 2.1, taken from the paper by Rammage *et al* [21], gives examples of the accuracy of several “Off-The-Shelf” Shack Hartmann wavefront sensors. As discussed in Chapter 1, and as this table demonstrates, the accuracy of the Shack-Hartmann depends largely on the size of the array (number of sub-apertures). Accuracies of $\lambda / 100$ to $\lambda / 150$ are not uncommon and ImagineOptic have one sensor (the HASO HP 26, a 26 x 26 sub-apertures with aperture diameter of 11.7mm) which

they claim to have an accuracy of $\lambda / 1000$ [22]. These figures give some indication of the accuracy of commonly used Shack-Hartmann sensors and puts the measured accuracies of the Green's function and GN algorithms into context. Even the lowest measured accuracy ($\lambda / 167$, Green's function solution) compares very favourably to the Shack-Hartmann, which is the PD sensors largest competitor.

2.6 Limitations of the DPD method

The DPD method itself, the use of a QD grating to apply defocus phase diversity to a test wavefront, is fundamentally a very versatile wavefront sensing technique. Its common path design makes it a very compact device, and using the diffraction grating to create the data removes the alignment complications of other techniques (in particular the Shack-Hartmann). The dynamic range of this device depends largely on the effective propagation distance between the two defocused sample planes and the strength of the wavefront aberration (as discussed in Section 2.2, see Figure 2.1).

The main limitations of the DPD sensor, as used previously, arise from the data reduction method. As discussed in Section 2.4.1 the Green's function solution relies on the limiting assumptions that the illumination must be uniform, and that both the wavefront and its slope must be continuous. This means that the reconstruction will be poor when dealing with scintillated and/or discontinuous wavefronts. The implications of this will be discussed further in the next section. The DPD sensor itself would be able to deal with scintillated wavefronts if used with a different data reduction algorithm, although the dynamic range would be reduced. If the propagation distance between the two planes (images in the ± 1 orders) is small then a scintillated, plane, wavefront can still give a null output. Over short distances the dark-patches/minima in the data from the scintillation will be in roughly the same place in the two planes, therefore the difference will still be zero.

The GN algorithm does not make the same limiting assumptions but its accuracy is highly dependent on the correct definition of the unit disc. This is a problem inherent in any algorithm which decomposes the data into a series of Zernike polynomials, and is

therefore not unique to the GN algorithm. The DPD sensor with the Green's function algorithm, which was the starting point for this project, is far more limited in terms of possible applications which provided motivation for the generalisation of this method and the creation of a new algorithm to analyse the data.

2.7.1. Application Considerations

The limitations of the Green's function solution, which make it unsuitable for scintillated and/or discontinuous wavefronts, preclude its use in a number of interesting applications.

In polishing applications extra errors in the finished form are caused by demounting the polished object for surface shape measurement then remounting it in the machine for further polishing. To increase polishing accuracy a compact, light-weight, wavefront sensing device like the diffraction grating-based sensor could be mounted within the polishing machine. Monitoring the surface shape of an sample in real-time would involve laser illumination of a rough surface thus creating a scintillated wavefront. A wavefront sensor capable of dealing with scintillated wavefronts would help to increase the accuracy of the polishing process, which is vital for several important current applications. Proposals for future Extremely Large Telescopes (ELTs), with primary mirrors of 30-100m in diameter, will require major advances in the manufacture and metrology of individual mirror segments. For example, each of the 618 hexagonal mirror segments for the EURO50 50m diameter telescope is required to have a surface form accuracy of better than 18nm peak to valley [23]. In medical applications, the lifetime of a prosthesis is highly dependent on the accuracy to which it has been polished, micron level defects in the form of prosthetic joints (for example hips and knees) can cause significant pressure concentrations and lead to premature failure [24, 25].

Scintillated wavefronts are also common in applications involving atmospheric propagation. As discussed in Chapter 1, variations in the atmosphere result in intensity variations of light which is propagated through it. This effect is more pronounced when

the light used is coherent, thus the propagation of lasers through the atmosphere creates significant problems. This is of particular interest in the fields of free-space optical communications and military imaging/targeting which often involve the use of lasers. A wavefront sensor capable of dealing with these scintillated wavefronts would therefore be of great use in these areas [26-29]

The second limitation of the Green's function solution, excluding its use with discontinuous wavefronts, is also very important. One disadvantage of this is that, if the DPD sensor with this algorithm is used in an Adaptive Optics system, in principle it precludes the use of pixellated Liquid Crystal (LC) wavefront correctors. LC correctors are lightweight, versatile and cost-effective [30], therefore a wavefront sensor capable of working with these new devices would be an attractive option for many applications. There are many research areas in which discontinuous wavefronts are unavoidable which would benefit from an improved wavefront sensor. In ELT's, where mirror segmentation creates piece-wise discontinuous wavefronts, the co-phasing of the individual mirror segments is a serious issue. ESO (European Southern Observatory) is currently developing OWL, the Overwhelmingly Large telescope, whose primary mirror it is proposed will be 100m in diameter containing 3048 hexagonal segments. The secondary mirror will also be segmented and is planned to contain 216 segments. Currently the worlds largest optical and infrared telescopes are the twin Keck telescopes, each 10m in diameter and comprising 36 segments [31-33]. This gives some idea of the increased technological challenge of phasing the OWL telescope segments to within the $0.5\mu\text{m}$ tolerance required for visible wavelengths [34]. The 618 segments of the EURO50 telescope primary mirror are to be phased and aligned to 10-20nm accuracy [23, 35]. These stringent requirements have led to a flurry of activity in segmented mirror co-phasing research.

In the computer and electronics industry the metrology of integrated circuits and the masks (called 'reticles') from which they are created (using lithography techniques) is an area of significant research interest [36-42]. As feature sizes continue to shrink the challenges faced in metrology of these components is increased. The diffraction-grating based PD wavefront sensor, if capable of dealing with discontinuous wavefronts, would be an ideal device for mounting on a robotic arm for example to provide real-time quality assessment measurements of integrated circuitry on the production line. It could

also perhaps be used to study the reticle pre-production where the detection of defects as small as 100nm could potentially save millions of pounds in production costs [42].

These applications are just a small selection of those which could benefit from the development of a more versatile phase diversity sensor. The development of one such sensor is the core work of this thesis and a more detailed discussion of possible applications will be given in Chapter 6, after the properties of this new sensor have been discussed.

2.8 Conclusions

In this chapter the basic principles underlying the design and operation of a diffraction-grating based DPD wavefront sensor have been discussed. It has been shown that a wavefront sensor of this type can be used to perform simultaneous multi-plane imaging and thus generate a pair of intensity images in the ± 1 diffraction orders which contain equal and opposite amounts of defocus as the phase diversity. The difference of these images divided by their separation is then approximately the axial intensity derivative along the axis of propagation and can be used to solve the ITE and retrieve the phase of the input wavefront. In this configuration the DPD sensor is analogous to the Curvature sensor described by Roddier [2].

Two methods of performing the data inversion and solving for the unknown phase were discussed; the Green's function and GN algorithm. The Green's function was shown to be a very accurate method, but its formulation required limiting assumptions to be placed on the input illumination and the wavefront under test. The implications of these assumptions and the motivation for developing a new method to overcome them were discussed. Experimental results demonstrating the accuracy of the GN algorithm were presented and, although less than the Green's function accuracy, this algorithm was shown to give excellent results when compared to theoretically calculated values.

In conclusion, this chapter has outlined the prior art and motivation for the development of the diffraction grating based PD sensor, which is the basis of this project. The aim of

this project is to generalise this method to use diversity functions other than defocus, and design a new algorithm to retrieve the phase which is capable of handling scintillated and discontinuous wavefronts. This will provide a new and improved wavefront sensing method which will benefit many applications and, it is hoped, be a viable alternative to the popular wavefront sensing techniques in use today.

2.9 References

1. Barty, A., K.A. Nugent, D. Paganin, and A. Roberts, *Quantitative Optical Phase Microscopy*. Optics Letters, 1998. **23**(11): p. 817-819.
2. Roddier, F., *Curvature Sensing and Compensation: A New Concept in Adaptive Optics*. Applied Optics, 1988. **27**(7): p. 1223-1225.
3. Gonsalves, R.A., *Phase Retrieval and Diversity in Adaptive Optics*. Optical Engineering, 1982. **21**(5): p. 829-832.
4. Teague, M.R., *Irradiance Moments - Their Propagation and Use for Unique Retrieval of Phase*. J.Opt.Soc.Am.A, 1982. **72**(9): p. 1199-1209.
5. Gureyev, T.E., A. Roberts, and K.A. Nugent, *Phase Retrieval with the Transport-of-Intensity Equation: Matrix Solution with Use of Zernike Polynomials*. J.Opt.Soc.Am.A, 1995. **12**(9): p. 1932-1941.
6. Woods, S.C. and A.H. Greenaway, *Wave-Front Sensing by Use of a Green's Function Solution to the Intensity Transport Equation*. J.Opt.Soc.Am.A, 2003. **20**(3): p. 508-512.
7. Hecht, E., *Optics*. 3rd ed. 1997: Addison Wesley Publishing Co.
8. Blanchard, P.M. and A.H. Greenaway. *Multiplane Imaging and Wavefront Sensing Using Distorted Diffraction Gratings*. in *Trends in Optics and Photonics. Diffractive Optics and Micro-Optics. Vol.41. Technical Digest. Postconference Edition, 18-22 June 2000*. 2000. Quebec City, Que., Canada: Opt. Soc. America. p. 250-252
9. Blanchard, P.M. and A.H. Greenaway, *Simultaneous Multiplane Imaging with a Distorted Diffraction Grating*. Applied Optics, 1999. **38**(32): p. 6692-6699.
10. Blanchard, P.M. and A.H. Greenaway, *Broadband Simultaneous Multiplane Imaging*. Optics Communications, 2000. **183**(1-4): p. 29-36.
11. Blanchard, P.M., D.J. Fisher, S.C. Woods, and A.H. Greenaway, *Phase-Diversity Wave-Front Sensing with a Distorted Diffraction Grating*. Applied Optics, 2000. **39**(35): p. 6649-6655.
12. Djidel, S., *Accuracy in 3d Imaging Using a Distorted Diffraction Grating*. 2001.Heriot-Watt University, Undergraduate Project Report
13. Greenaway, A.H. and S. Djidel, *Telecentric Microscope*. Patent No. GB0421570.3 2004 U.K
14. Youbin, F., Z. Xuerning, and Y. Jiexiong, *Research on the Dynamics of a Multidither Adaptive Optical System*. Optical Engineering, 1998. **37**(4): p. 1208-1211.
15. Gureyev, T.E. and K.A. Nugent, *Phase Retrieval with the Transport-of-Intensity Equation. Ii. Orthogonal Series Solution for Nonuniform Illumination*. J.Opt.Soc.Am.A, 1996. **13**(8): p. 1670-1682.

16. Quiroga, J.A., J.A. Gomez-Pedrero, and J.C. Martinez-Anton, *Wavefront Measurement by Solving the Irradiance Transport Equation for Multifocal Systems*. Optical Engineering, 2001. **40**(12): p. 2885-2891.
17. Teague, M.R., *Deterministic Phase Retrieval: A Green's Function Solution*. J.Opt.Soc.Am.A, 1983. **73**(11): p. 1434-1441.
18. Van Dam, M.A. and R.G. Lane, *Extended Analysis of Curvature Sensing*. J.Opt.Soc.Am.A, 2002. **19**(7): p. 1390-1397.
19. Roddier, F., *Wavefront Sensing and the Irradiance Transport Equation*. Applied Optics, 1990. **29**: p. 1402-1406.
20. Djidel, S. and A.H. Greenaway. *Nanometric Wavefront Sensing*. in *3rd International Workshop on Adaptive Optics in Industry and Medicine*. 2002: Starline Printing Inc. p. 213-219
21. Rammage, R.R., D.R. Neal, and R.J. Copland. *Application of Shack-Hartmann Wavefront Sensing Technology to Transmissive Optic Metrology*. in *Advanced Characterization Techniques for Optical, Semiconductor, and Data Storage Components*. 2002: Proc. SPIE. Vol. 4779. p. 161-172
22. Ballesta, J., *Schwarzschild Alignment with Hp26*. 2005.Imagine-Optic Report (not published - received by email)
23. Dae Wook Kim and S.-W. Kim, *Static Tool Influence Function for Fabrication Simulation of Hexagonal Mirror Segments for Extremely Large Telescopes*. Optics Express, 2005. **13**(3): p. 910-917.
24. Walker, D., et al. *First Results on Free-Form Polishing Using the Precessions Process*. in *'Free-Form Optics: Design, Fabrication, Metrology, Assembly, ASPE 2004 Winter Topical Meeting*. 2004. Chapel Hill, North Carolina. Vol. 31. p. 29-34
25. Davis, P.J., *Artificial Hip Joints: Applying Weapons Expertise to Medical Technology*. Energy and Technology Review (October issue), 2004: p. 16-21.
26. Weyrauch, T. and M.A. Vorontsov, *Atmospheric Compensation with a Speckle Beacon in Strong Scintillation Conditions: Directed Energy and Laser Communication Applications*. Applied Optics, 2005. **44**(30): p. 6388-6401.
27. Zocchi, F.E., *A Simple Analytical Model of Adaptive Optics for Direct Detection Free-Space Optical Communication*. Optics Communications, 2005. **248**(4-6): p. 359-374.
28. Primmerman, C.A., T.R. Price, R.A. Humphreys, B.G. Zollars, H.T. Barclay, and J. Herrmann, *Atmospheric-Compensation Experiments in Strong-Scintillation Conditions*. Applied Optics, 1995. **34**(12): p. 2081-2088.
29. Levine, B.M., E.A. Martinsen, A. Wirth, A. Jankevics, M. Toledo-Quinones, F. Landers, and T.L. Bruno, *Horizontal Line-of-Sight Turbulence over near-Ground Paths and Implications for Adaptive Optics Corrections in Laser Communications*. Applied Optics, 1998. **37**(21): p. 4553-4560.
30. Dayton, D.C., S.L. Browne, S.P. Sandven, J.D. Gonglewski, and A.V. Kudryashov, *Theory and Laboratory Demonstrations on the Use of a Nematic Liquid-Crystal Phase Modulator for Controlled Turbulence Generation and Adaptive Optics*. Applied Optics, 1998. **37**(24): p. 5579-5589.
31. Medwadowski, S.J., *Structure of the Keck Telescope - an Overview*. Astrophysics and Space Science, 1989. **160**(1-2): p. 33-43.
32. Waldrop, M.M., *Keck Telescope Ushers in a New Era*. Science, 1990. **249**(4974): p. 1244-1246.
33. Chanan, G., M. Troy, and C. Ohara. *Phasing the Primary Mirror Segments of the Keck Telescopes: A Comparison of Different Techniques*. 2000: Proc. SPIE. Vol. 4003. p. 188-201

34. Gilmozzi, R., et al. *The Future of Filled Aperture Telescopes: Is a 100m Feasible?* in *Advanced Technology Optical/IR Telescopes VI*. 1998. Kona, Hawaii: Proc. SPIE. Vol. 3352. p. 778-791
35. Andersena, T., A. Ardeberga, J. Beckersb, A. Goncharova, M. Owner-Petersena, and H. Riewaldta. *Euro 50*. in *2nd Bäckaskog Workshop on Extremely Large Telescopes*. 2003. Bäckaskog, Sweden: Proc. SPIE. Vol. 5382. p. 169-182
36. Yuan, C.-M. and A.J. Strojwas. *Modeling of Optical Alignment and Metrology Schemes Used in Integrated Circuit Manufacturing*. in *Optical/Laser Microlithography III, Mar 7-9 1990*. 1990. San Jose, CA, USA: Publ by Int Soc for Optical Engineering, Bellingham, WA, USA. Vol. 1264. p. 203-218
37. Griffith, J.E., D.A. Grigg, M.J. Vasile, P.E. Russell, and E.A. Fitzgerald, *Scanning Probe Metrology*. *Journal of Vacuum Science & Technology a-Vacuum Surfaces and Films*, 1992. **10**(4): p. 674-679.
38. Gale, D.M., M.I. Pether, and J.C. Dainty, *Linnik Microscope Imaging of Integrated Circuit Structures*. *Applied Optics*, 1996. **35**(1): p. 131-148.
39. Chim, S.S. and G.S. Kino. *Optical Metrology for Integrated Circuit Fabrication*. in *Integrated Circuit Metrology, Inspection, and Process Control V, Mar 4-5 1991*. 1991. San Jose, CA, USA: Publ by Int Soc for Optical Engineering, Bellingham, WA, USA. Vol. 1464. p. 138-144
40. Bridges, G.E., R.A. Said, and D.J. Thompson, *Heterodyne Electrostatic Force Microscopy for Noncontact High-Frequency Integrated-Circuit Measurement*. *Electronics Letters*, 1993. **29**(16): p. 1448-1449.
41. Andriamonje, G., V. Pouget, Y. Ousten, D. Lewis, B. Plano, and Y. Danto. *Structural Analysis of Integrated Circuits Using Scanning Laser Ultrasonics*. in *2004 IEEE International Reliability Physics Symposium Proceedings, 42nd Annual, Apr 25-29 2004*. 2004. Phoenix, AZ., United States: Institute of Electrical and Electronics Engineers Inc., Piscataway, United States. p. 375-380
42. Kincade, K., *Mask Inspection Rises to the Subwavelength Challenge*. *Laser Focus World*, 2004. **40**(7): p. 92-98.

Chapter 3

Generalised Phase Diversity

Summary of Chapter 3

In this chapter Generalised Phase Diversity (GPD) will be presented as an extension of the Defocus only Phase Diversity (DPD) wavefront sensor described in Chapter 2. GPD will use aberration kernels other than Defocus to apply diversity to the phase of the input wavefront. The configuration of this wavefront sensor and the theory governing its operation will be detailed. The Necessary and Sufficient conditions an aberration kernel must satisfy for use in a null wavefront sensor will be derived, and verified through computer simulation. The form of the error signal will be considered and the implications this has for use of the wavefront sensor will be discussed. This discussion will be continued in Chapter 4.

3.1 Introduction

Generalised Phase Diversity (GPD) is the next step in diffraction grating based phase diversity wavefront sensors. As the previous chapter showed, the Defocus-only Phase Diversity (DPD) wavefront sensor proved to be a very accurate means of measuring wavefront curvature. However, this method was not without disadvantages.

The main problem with the DPD method was caused by the data inversion. Solution of the Intensity Transport Equation (ITE), by use of a Green's function, involves placing certain assumptions on the input wavefront [1]. These assumptions, detailed in Section 2.4.1, mean that the DPD wavefront sensor is unable to reconstruct discontinuous or scintillated wavefronts with sufficient accuracy.

As discussed in Chapter 2 there are many applications which would benefit from a wavefront sensor capable of handling discontinuous and scintillated wavefronts. Initially it would be enough to develop a sensor capable of dealing with significant

intensity variations, the problem of branch points (points where the wavefront phase is indeterminate at points where scintillation has caused the intensity to fall to zero) can be studied as a separate issue later. Even if a sensor could be developed which can deal with branch points, correction of such a wavefront is not a trivial problem. Continuous wavefront modulators such as membrane deformable mirrors are fundamentally incapable of correcting discontinuities in the phase. Therefore, while the problems caused by scintillation are of significant research interest, a PD sensor immune to the intensity variations this causes would be a reasonable first step forward.

To overcome the limitations of the DPD sensor and develop a wavefront sensor capable of dealing with discontinuous and scintillated wavefronts, it was first necessary to look at the basic theory behind the operation of the wavefront sensor. Defocus has been shown to work successfully as the PD function which is added to the unknown wavefront. This then raises certain questions: What, if anything, is unique about defocus? Could other aberration kernels be used? What conditions must an aberration function satisfy to create a successful wavefront sensor signal? What would this error signal look like? Could this be used to build a useful wavefront sensor?

The analysis presented in this chapter will begin to answer some of these questions. Firstly the configuration of the wavefront sensor will be considered, as this will impact its operation. Then, the conditions an aberration filter function must satisfy to create a null wavefront sensor will be derived. This forms the basis for GPD and in this chapter it will be shown that it is possible to create a PD wavefront sensor using aberration kernels other than defocus.

3.2 Image Plane vs. Pupil Plane

There are several methods which could be used to apply diversity phase to an unknown wavefront. A convenient method is the diffraction grating based system described in Chapter 2. Using this system as its basis the GPD wavefront sensor will contain a carefully chosen diffraction grating in combination with a lens to apply diversity to the input wavefront. The design of the grating will include the use of an aberration kernel that may not be defocus. The theory developed in this chapter essentially details how to

choose the grating to be used. The wavefront sensor can incorporate other elements in addition to this to increase the flexibility of the wavefront sensor's operation. For example, several gratings could be mounted in a filter wheel which could be rotated in front of the camera to allow the user to change the diversity phase to suit their measurement.

Initially let us consider the wavefront sensor in its simplest form; a grating and lens. There are two different optical configurations we will consider for the GPD wavefront sensor. These are referred to as the 'image plane' and 'pupil plane' configurations and will be studied separately.

3.2.1 Image Plane Operation

A schematic of an 'image plane' wavefront sensor is given below in Figure 3.1:

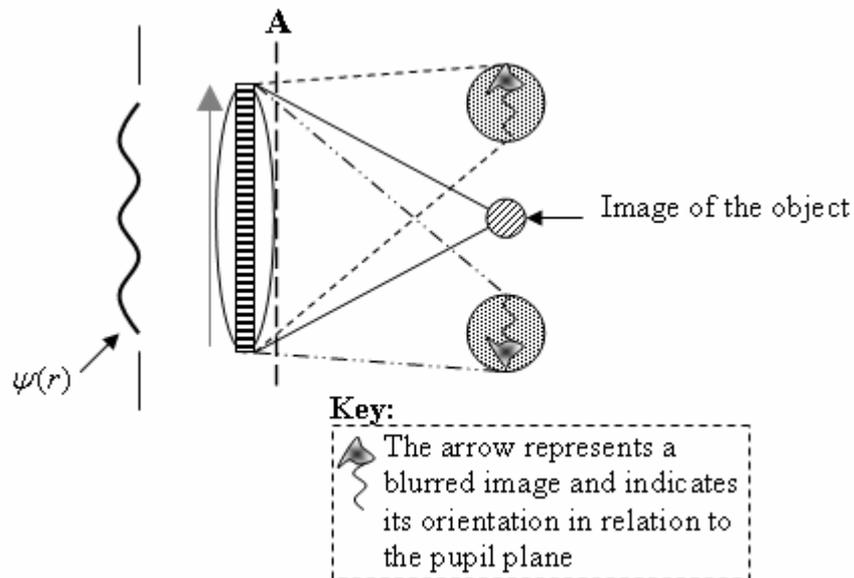


Figure 3.1 Schematic of a GPD wavefront sensor designed to operate in the image plane. The grey arrows are used to indicate the orientation of the formed images with respect to each other and to the pupil plane.

Chapter 3: Generalised Phase Diversity

An input wavefront $\Psi(r)$, generated by radiation from a distant object, is incident on the input pupil of the wavefront sensor. In the image plane configuration this is simply a limiting aperture which truncates the wavefront and may be the edge of the imaging lens. The next element the input wavefront will encounter is the grating and lens combination. In Figure 3.1 this is represented as a single element, actually the two elements are placed as close to each other as possible and can be approximated by a single element whose focal length is a combination of the lens and grating focal lengths. The orientation of the images relative to each other and to the pupil plane is indicated by the grey arrows in Figure 3.1.

It is well known that a lens produces a two-dimensional Fourier Transform (FT) of an object at a distance f after the lens. Using the concept of Fractional FT's [2, 3] this can be described as a FT of degree 1 ($\mathfrak{F}^{(1)}$). Fractional FT's of order 0-1 exist between the plane of the lens and its' focal plane [4, 5]. If the grating is placed directly against the lens it can be approximated that the order of the FT is 0 at this point (shown as plane 'A' in Figure 3.1). Therefore at point A the distribution is given by:

$$\psi(r).e^{i\varphi_d}.pupil , \quad (3.1)$$

where $\psi(r)$ is the input wavefront and $e^{i\varphi_d}$ is the function applied by the grating (φ_d is the diversity phase).

Equation (3.1) shows that the phase diversity is applied by multiplying the input wavefront and grating function. This is then multiplied by the pupil function of the lens. If the lens is considered to be a perfect, infinitely sized, thin lens then the pupil function is simply unity.

Let the input wavefront be described:

$$\psi(r) = Ae^{i\varphi} . \quad (3.2)$$

The intensity image formed in the 0th order is given by the following equation:

$$I_{0^{\text{th}}} = \left| \mathfrak{F}\{Ae^{i\varphi}\} \right|^2 = |\Psi(\xi)|^2, \quad (3.3)$$

therefore, the image formed in the 0th order is the image of the object which produced the wavefront $\psi(r)$. It is for this reason that this is named the image plane configuration.

The DPD wavefront sensor, introduced in Chapter 2, is usually built in an image plane configuration. If the amount of defocus added to the wavefront is large enough, then the planes that are imaged in the ± 1 diffraction orders approach the pupil plane of the system. In this case the 0th order image will still contain an image of the source, but the phase diverse images will more closely resemble the pupil plane. The advantage of imaging the pupil plane will be discussed in the following sections where the pupil plane configuration is introduced.

Finally, note that there is a 180° rotation between the intensity images in the ± 1 diffraction orders when using this configuration (as shown by the grey arrows in Figure 3.1). This is due to the fact that one image is effectively formed before and one after focus. The image formed before focus is therefore a virtual image and has the same orientation as the pupil plane. This means that care must be taken when using the image plane configuration to rotate one of the intensity images by 180° before taking the difference between them.

3.2.2 Pupil Plane Operation

In the ‘pupil plane’ configuration a pupil lens replaces the limiting aperture. This is shown in Figure 3.2.

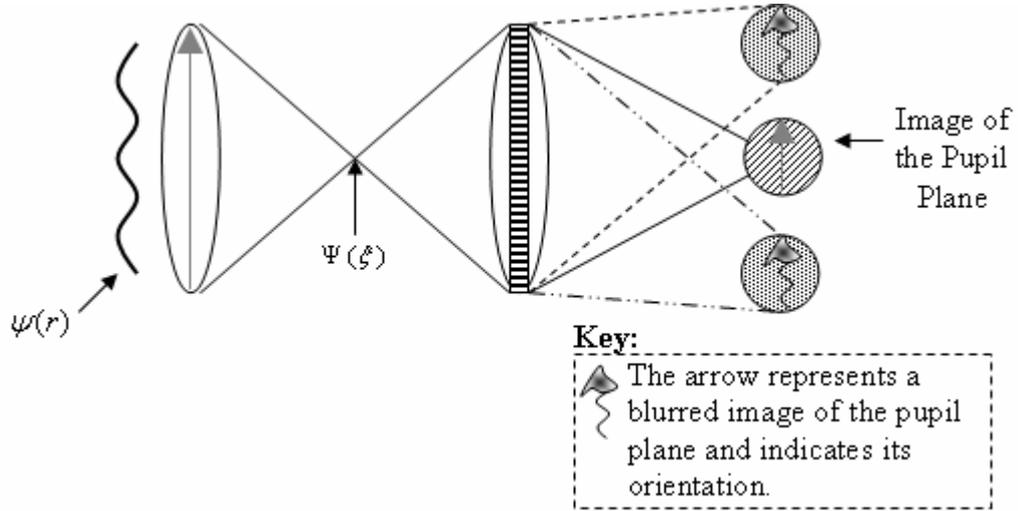


Figure 3.2 Schematic of the Pupil Plane configuration of the GPD wavefront sensor. The grey arrows are used to indicate the orientation of the formed images with respect to each other and to the pupil plane.

At a distance f from the pupil lens the FT of the input wavefront is formed [5], shown in Figure 3.2 as $\Psi(\xi)$. Goodman [6] has shown that the complex amplitude distribution in the back focal plane of a lens, in this case $\Psi(\xi)$, is the Fraunhofer diffraction pattern of the incident field $\psi(r)$. This is calculated as the FT of the input distribution multiplied by a quadratic phase factor and is shown in Equation (3.4):

$$\Psi(\xi) = \frac{\exp\left[i\frac{k}{2z}\cdot\xi^2\right]}{i\lambda z} \int \int_{-\infty}^{\infty} \psi(r) \exp\left[-i\frac{2\pi}{\lambda z}r\xi\right] dr . \quad (3.4)$$

Chapter 3: Generalised Phase Diversity

Amplitude and phase co-ordinates in the focal plane are determined by the amplitude and phase of the input Fourier components at $f_r = \xi/\lambda z$, where z in this case is the focal length of the pupil lens. The quadratic phase term in Equation (3.4), it will be shown in Section 3.4, can be neglected due to the properties of the lens-grating combination. Therefore for this analysis it will be assumed that the field at the focal plane of the pupil lens ($\Psi(\xi)$) is simply the FT of the input wavefront $\psi(r)$.

The optical set-up from this point to the image plane of the grating and lens is identical to the image plane configuration. The analysis is therefore the same, with the input to this system now replaced by $\Psi(\xi)$. The effect of this is that the intensity image in the 0th order is now given by:

$$I_{0^{\text{th}}} = |\mathfrak{I}\{\Psi(\xi)\}|^2 = |\psi(r)|^2 . \quad (3.5)$$

Therefore, in the 0th order the image of the wavefront in the pupil plane is obtained. In the next section it will be shown that, when using the pupil plane configuration, there is a direct 1-1 mapping between the positions of the error on the input wavefront at the pupil, with the position it is detected using the intensity images from the ± 1 and 0th diffraction orders. This will prove to be an important property of the pupil plane system and the theory behind it will be studied in more depth in Chapter 4.

Comparison of Equations (3.3) and (3.5) shows that the main distinction between the image and pupil plane configuration is whether the 0th order contains an image of the source ((3.3)), or the wavefront amplitude ((3.5)).

3.2.3 Comparison of the Image Plane and Pupil Plane Configurations

A Matlab simulation was written to demonstrate the dramatic difference in the form of the error signal generated by image plane and pupil plane systems. The wavefront sensor error signal is defined as the difference of the intensity images in the ± 1 diffraction orders. In this simulation an aberrated input wavefront, with a Peak to

Chapter 3: Generalised Phase Diversity

Valley (PV) distortion of 0.3λ , was created from a combination of 3 Zernike polynomials ($Z_2^{-2}, Z_4^{-4}, Z_6^2$). Spherical aberration ($Z_4^0, 0.75\lambda$ PV) was chosen as the PD filter function to clearly differentiate between the image and pupil plane results. As mentioned before defocus diversity phase of sufficient strength would also image close to the pupil plane and therefore the error signal would closely resemble the pupil plane system equivalent. The results are shown in Figure 3.3.

Note that in this example, and throughout this chapter, the PV distortions of the diversity phase and input wavefronts are chosen arbitrarily. There is no special significance at this point of the values chosen except where a specific comparison is being calculated, in which case this will be pointed out in the accompanying text.

Figure 3.3(a) shows the distorted input wavefront, the scale of this subplot is in waves of distortion. The error signals for the pupil plane and image plane configuration are shown in Figure 3.3(b) and (c) respectively and the scales of these subplots are a measure of contrast. Throughout this thesis error signal scales will be described as ‘a measure of contrast’. The reason for this is, as a difference of intensity images, the units for these should be a measure of intensity and ultimately a number of photons. As all of these results are simulated this ‘number of photons’ has little meaning, and would simply be a multiplying constant for the scales of these difference images.

Note that the integrated energy in the difference images shown in 3.3(b) and (c) is the same. The 10^2 difference between the scales of these images is due to the fact that image plane simulation requires one less FT than the pupil plane example, and the FT has caused a scaling difference of N^2 .

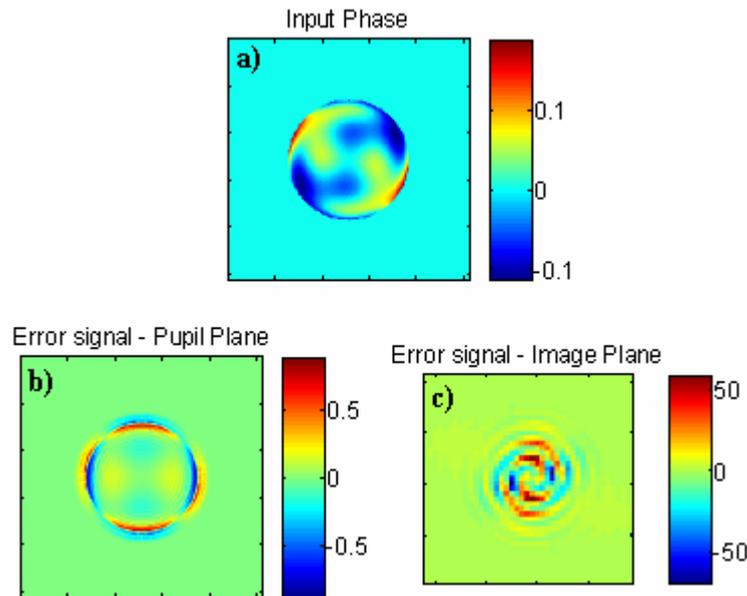


Figure 3.3 Matlab simulation results showing the error signal generated when a distorted input wavefront (a), is incident on a pupil plane (b) and image plane (c) wavefront sensor [scale of (a) is in waves, scales of (b) and (c) are a measure of contrast]

Figure 3.3 (b) clearly shows the 1-1 mapping between the position of the wavefront error on the pupil and its position on the error signal. This property makes the pupil plane configuration ideal for a null wavefront sensor. The information in this error signal can be used to drive a corrective element in an Adaptive Optics (AO) system directly, as the error signal gives the location, magnitude and sense of the error. As mentioned in Section 3.2.1 the image plane sensor could be constructed, with a large enough defocus, to image close to the pupil plane in the ± 1 diffraction orders. In this case the 1-1 mapping would also hold. However, this would only be true when the diversity phase used is defocus, whereas the pupil plane configuration retains this property regardless of the diversity function used.

Figure 3.3(c) shows that using the image plane system an error signal is obtained, indicating that the input wavefront is distorted, so this is also suitable for a null wavefront sensor. The error signal shown in 3.3(c) is localised at the origin, and this plot has been ‘zoomed in’ so that it can be seen clearly. The disadvantage of using the image plane configuration, as this plot shows, is that the error location information has

been lost and therefore correction would have to be applied using a multi-dither approach [7].

3.2.4 Application Considerations

It has been shown that a GPD wavefront sensor can be constructed to operate in the image or the pupil plane. The choice of configuration will depend on the final application of the sensor.

For AO systems where the main objective is to correct errors on the input wavefront the pupil plane system property of 1-1 mapping between the position of the error on the wavefront and the point it appears in the error signal is very useful. The wavefront sensor may be used as a null sensor without the need to reconstruct the wavefront. This increases the speed of operation and reduces the complexity of the system. An image plane system and multi-dither corrective approach could also be used, but would not be as fast due to the loss of the error location information (unless defocus is used to image the pupil plane as discussed earlier).

In metrology applications it is the surface shape which is often of interest. In this case a reconstruction of the wavefront should be used and so neither configuration has a clear advantage over the other. The choice would then depend on the preference of the user and other considerations such as the sensitivity, accuracy and algorithmic speed required.

3.3 The Filter Function

The GPD wavefront sensor, like its predecessor the DPD wavefront sensor, will comprise a grating in combination with a lens, and a CCD camera. The major differences between the GPD and DPD sensors are the design of the diffraction grating, and the data inversion method. In both cases the aberration kernel within the grating is combined with the input wavefront to produce the phase diverse data required for phase retrieval. How this combination takes place will depend on the configuration of the

optical system. If the wavefront sensor is built to operate in the image plane then the wavefront is taken in product with the grating, i.e. the recorded data is the convolution of the wavefront and the FT of the diversity function. In the pupil plane configuration the recorded image data is the image of the pupil convolved with different aberration functions.

In the DPD sensor it is defocus which is applied to the intensity images in the ± 1 diffraction orders. Generalised Phase Diversity is so called because the aberration kernel may include, but will not be limited to, defocus. The choice of permitted aberration functions will be limited by the way in which the user chooses to operate the GPD wavefront sensor, and by the symmetries of the filter function and input wavefront.

In the description given below the GPD wavefront sensor will be considered to be a simple null sensor, which is designed to give an error signal only when the input wavefront is distorted. Equations for the diversity filter function and the generated error signal will be derived in this chapter, and used to formulate a phase retrieval algorithm in Chapter 4. A glossary is included in the Appendix at the end of this chapter to summarise all the terms introduced here and their definition, for quick reference as they will be used extensively both here and in following chapters.

The aberration kernel chosen to apply the diversity will be encoded into the complex filter function $f_{\pm}(r)$ whose Fourier Transform $F_{\pm}(\xi)$ will be convolved with the input wavefront. $F_{\pm}(\xi)$ can be written as the sum of its real and imaginary parts:

$$F_{\pm}(\xi) = R(\xi) \pm iI(\xi) , \quad (3.6)$$

where $R(\xi)$ is the real part, and $I(\xi)$ is the imaginary part of the FT of the filter function.

As discussed in Chapter 1, PD often refers to systems in which the data is created by comparing a pair of images where only one of those images contains the diversity phase (for example, an in-focus and defocused image). In Curvature Sensing (CS), a special case of PD, the data images are encoded with equal and opposite amounts of defocus aberration. As in the DPD sensor described in Chapter 2 the GPD sensor outlined here will use the CS approach of applying equal and opposite diversity to the input wavefront. In both $f_{\pm}(r)$ and $F_{\pm}(\xi)$ the \pm is used to represent these equal and opposite aberrations. As discussed in Chapter 2, use of detour phase with a Diffractive Optic Element (DOE) conveniently achieves this automatically. The image in the +1 diffraction order has been created using $F_{+}(\xi)$, and similarly $F_{-}(\xi)$ is used to generate the -1 intensity image.

3.4 The Intensity Images

GPD, like the DPD wavefront sensor, will require a pair of intensity images to reconstruct the wavefront phase. These intensity images, instead of being formed under different defocus conditions, will be produced by the convolution of the input wavefront with the FT of the filter function ($F_{\pm}(\xi)$). As in the DPD case it is the difference of these two intensity images that will provide the information required for phase retrieval. In this section the mathematical expression for the intensity images, whose difference will form the basis for the error signal, will be derived. Chapter 4 will look at the formation of this error signal in a little more detail and will then go on to describe how it may be used to retrieve the unknown phase of the input wavefront.

Consider a simple optical system with an input pupil, a grating in combination with a lens, and an output image plane. This situation is shown schematically in Figure 3.4. In both the Fraunhofer (described earlier in Section 3.2.2) and Fresnel (shown in this case) approximations an additional curvature term multiplies the field found after the lens. In forming an image an idealised lens must correct any curvature terms on the wavefront formed in the lens. A lens may be described as performing two Fourier Transforms in the same direction (i.e. both forward) with any lens aberrations being applied to the object FT by a multiplicative phase function within the lens. In this example, shown in Figure 3.4, the diffraction grating is shown as being coplanar with the lens and may

therefore be considered as the ‘lens aberration’ from this description. If the lens-grating combination is modelled as an idealised lens then the additional phase curvature term (described in Section 3.2.2.) will be cancelled by the grating and will not be included in any of the analysis which follows.

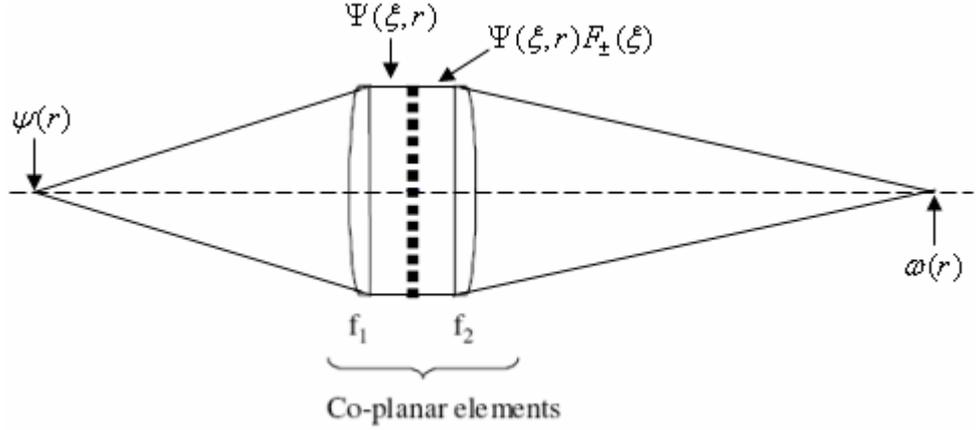


Figure 3.4 Schematic of the GPD optical system [this is a modified version of a figure from Dijdel and Greenaway [8]].

$\Psi(\xi)$ is the Fourier Transform (FT) of the input wavefront $\psi(r)$. $\Psi(\xi)$ is multiplied by $F_{\pm}(\xi)$, which is equivalent to a convolution of $\psi(r)$ and $f_{\pm}(r)$. The image $\omega(r)$ formed in the output plane is the FT of the product $\Psi(\xi)F_{\pm}(\xi)$, and is given in Equation (3.7):

$$\omega_{\pm}(r) = \int d\xi \cdot \Psi(\xi) F_{\pm}(\xi) \exp[-i\xi r] . \quad (3.7)$$

The intensity images in the ± 1 diffraction orders may therefore be written as the modulus squared $\omega_{\pm}(r)$:

$$\begin{aligned} j_{\pm}(r) &= |\omega_{\pm}(r)|^2 \\ &= \left| \int d\xi \cdot \Psi(\xi) F_{\pm}(\xi) \cdot \exp[-i\xi r] \right|^2 , \end{aligned} \quad (3.8)$$

where $j_+(r)$ is the image intensity in the +1 diffraction order, and $j_-(r)$ is the intensity in the -1 diffraction order. For a properly constructed null sensor, with a plane wave input, the images in the ± 1 diffraction orders will be identical so that their difference is zero. This fact will be used later as a test of different diversity functions to prove their suitability (or unsuitability) for use in a null sensor.

3.5 The Error Signal

The error signal, $d(r)$ is produced by taking the difference between the intensity images in the ± 1 diffraction orders. Since the modulus squared of a complex function can be calculated by multiplying the function by its conjugate, expanding (3.8) the following is obtained:

$$\begin{aligned}
 j_{\pm}(r) &= \left| \int d\xi \cdot \Psi(\xi) F_{\pm}(\xi) \cdot \exp[-i\xi r] \right|^2 \\
 &= \left\{ \int d\xi \cdot \Psi(\xi) F_{\pm}(\xi) \cdot \exp[-i\xi r] \right\} \cdot \left\{ \int d\xi \cdot \Psi(\xi) F_{\pm}(\xi) \cdot \exp[-i\xi r] \right\}^* \quad (3.9) \\
 &= \iint d\xi d\xi' \Psi(\xi) \Psi^*(\xi') F_{\pm}(\xi) F_{\pm}^*(\xi') \exp[-ir(\xi - \xi')]
 \end{aligned}$$

Therefore the difference between the two images can be written:

$$\begin{aligned}
 d(r) &= j_+(r) - j_-(r) \\
 &= \left\{ \iint d\xi d\xi' \Psi(\xi) \Psi^*(\xi') F_+(\xi) F_+^*(\xi') \exp[-ir(\xi - \xi')] \right\} \cdot \quad (3.10) \\
 &\quad - \left\{ \iint d\xi d\xi' \Psi(\xi) \Psi^*(\xi') F_-(\xi) F_-^*(\xi') \exp[-ir(\xi - \xi')] \right\}
 \end{aligned}$$

Chapter 3: Generalised Phase Diversity

By substituting (3.6) for the FT of the filter function a new expression for the error signal is derived as shown in Equation (3.11):

$$d(r) = \iint d\xi d\xi' \Psi(\xi) \Psi^*(\xi') [F_+(\xi) F_+^*(\xi') - F_-(\xi) F_-^*(\xi')] \exp[-ir(\xi - \xi')] \quad (3.11)$$

$$\Rightarrow d(r) = \iint d\xi d\xi' \Psi(\xi) \Psi^*(\xi') [-2i \{R(\xi) I(\xi') - I(\xi) R(\xi')\}] \exp[-ir(\xi - \xi')]$$

Equation (3.11) is then rearranged, collecting terms of the same variable, to obtain (3.12):

$$d(r) = 2i \left[\int d\xi \Psi(\xi) I(\xi) \exp(-ir\xi) \int d\xi' \Psi^*(\xi') R(\xi') \exp(ir\xi') \right. \\ \left. - \int d\xi \Psi(\xi) R(\xi) \exp(-ir\xi) \int d\xi' \Psi^*(\xi') I(\xi') \exp(ir\xi') \right] \quad (3.12)$$

This expression for $d(r)$ is purely real since the term in the [] brackets is the difference of two complex conjugates and is therefore purely imaginary. Thus far no limiting assumptions have been placed on the input wavefront so (3.12) is generally valid.

The input wavefront in the pupil plane is defined to be:

$$\psi(r) = |\psi(r)| \exp[i\phi(r)] \quad (3.13)$$

where r is the co-ordinate in the pupil plane. The FT of $\psi(r)$ can be written:

$$\Psi(\xi) = H(\xi) + A(\xi) \quad (3.14)$$

$H(\xi)$ is the Hermitian component of $\Psi(\xi)$, formed by the FT of the purely real parts of $\psi(r)$, and $A(\xi)$ is the Anti-Hermitian component of $\Psi(\xi)$ and therefore the FT of the

imaginary parts of $\psi(r)$. The symmetries of these components will be required later and can be expressed:

$$H(\xi) = H^*(-\xi) \text{ and } A(\xi) = -A^*(-\xi) . \quad (3.15)$$

The form of the input wavefront given in (3.14) may now be substituted into (3.12), fully expand and the terms grouped appropriately to create:

$$\left[\int d\xi H(\xi) I(\xi) \exp(-ir\xi) \int d\xi' H^*(\xi') R(\xi') \exp(ir\xi') - \int d\xi H(\xi) R(\xi) \exp(-ir\xi) \int d\xi' H^*(\xi') I(\xi') \exp(ir\xi') \right] \quad (3.16)$$

$$\left[\int d\xi H(\xi) I(\xi) \exp(-ir\xi) \int d\xi' A^*(\xi') R(\xi') \exp(ir\xi') - \int d\xi A(\xi) R(\xi) \exp(-ir\xi) \int d\xi' H^*(\xi') I(\xi') \exp(ir\xi') \right] \quad (3.17)$$

$$\left[\int d\xi A(\xi) I(\xi) \exp(-ir\xi) \int d\xi' H^*(\xi') R(\xi') \exp(ir\xi') - \int d\xi H(\xi) R(\xi) \exp(-ir\xi) \int d\xi' A^*(\xi') I(\xi') \exp(ir\xi') \right] \quad (3.18)$$

$$\left[\int d\xi A(\xi) I(\xi) \exp(-ir\xi) \int d\xi' A^*(\xi') R(\xi') \exp(ir\xi') - \int d\xi A(\xi) R(\xi) \exp(-ir\xi) \int d\xi' A^*(\xi') I(\xi') \exp(ir\xi') \right] . \quad (3.19)$$

$d(r)/2i$ is the sum of Equations (3.16) to (3.19). These equations have been arranged such that in each one there is a difference of two complex conjugates. Also we note that Equations (3.16) and (3.19) contain terms only in either $H(\xi)$ or $A(\xi)$. Equations (3.17) and (3.18) contain 'cross terms', meaning that in each pair of integrals there is one containing $H(\xi)$ and one containing $A(\xi)$.

It is immediately apparent that every term in (3.16)-(3.19) contains both $R(\xi)$ or $I(\xi)$, the real and imaginary parts of the filter function FT (as described in (3.6)). This means

that the filter function must be complex. If either the real or the imaginary parts are allowed to go to zero, the wavefront sensor will produce a null output for any input wavefront. This is therefore the first necessary condition for operation as a null sensor.

The sum of Equations (3.16) to (3.19) creates the full form of the error signal $d(r)$. The different symmetries of $F_{\pm}(\xi)$ will now be examined to see if this expression can be simplified and to find further necessary and sufficient conditions to supply a useful error signal for a null wavefront sensor.

3.6 Symmetries of the Filter Function

A null sensor is a specific type of wavefront sensor which will only provide an error signal when the wavefront is aberrated. Therefore, the first sufficient condition for null sensor operation is that it should give no signal when the input wavefront is a plane wave, but should produce an error signal when the input wavefront is aberrated.

To formulate the necessary conditions the form of the error function when the real and imaginary parts of $F_{\pm}(\xi)$ have the same symmetry (i.e. both are odd or even functions of ξ) and when they have mixed symmetry (i.e. one is an even function while the other is odd) will be studied.

3.6.1 Same Symmetry

First consider the case when both $R(\xi)$ and $I(\xi)$ are even functions of ξ , and are therefore both real-valued and symmetric. In Equation (3.16), recalling the symmetry properties of $H(\xi)$ given in equation (3.15), it can be seen that each of the integrals is purely real and so this difference of complex conjugates will be $0 \forall \psi$. Similarly for Equation (3.19) each integral is purely imaginary and therefore this equation will also reduce to $0 \forall \psi$. The difference between the intensity images, which is the error signal, is now reduced to the sum of Equations (3.17) and (3.18). This is shown in Equation (3.20):

$$\begin{aligned} \frac{d(r)}{2i} = & \left[\int d\xi H(\xi) I(\xi) \exp(-ir\xi) \int d\xi' A^*(\xi') R(\xi') \exp(ir\xi') - \right. \\ & \left. \int d\xi A(\xi) R(\xi) \exp(-ir\xi) \int d\xi' H^*(\xi') I(\xi') \exp(ir\xi') \right] + , \quad (3.20) \\ & \left[\int d\xi A(\xi) I(\xi) \exp(-ir\xi) \int d\xi' H^*(\xi') R(\xi') \exp(ir\xi') - \right. \\ & \left. \int d\xi H(\xi) R(\xi) \exp(-ir\xi) \int d\xi' A^*(\xi') I(\xi') \exp(ir\xi') \right] \end{aligned}$$

To determine whether this is a suitable form of error signal for a null sensor let us consider the result of Equation (3.20) when a plane wave is input.

A plane wave can be defined simply as a wavefront with a constant phase. Without any loss of generality this phase ($\varphi(r)$) can be defined to be zero. When $\varphi(r) = 0$ then $a(r)$, the imaginary part of $\psi(r)$ $\left[a(r) = \text{Im}\{|\psi(r)|e^{-i.2\pi.\varphi(r)}\} \right]$, will also be zero. Therefore the wavefront, $\psi(r)$, is purely real and its FT will be purely Hermitian ($A(\xi) \equiv 0$). When this is the case Equation (3.20) shows that the error signal, $d(r)$, will be zero. This is due to the cross terms described earlier, since each integral pair contains $A(\xi)$ each term will therefore reduce to $0 \forall \xi$. Any non-plane wavefront will contain both Hermitian and Anti-Hermitian components and will produce an error signal as described by (3.20). Therefore, this shows that when $R(\xi)$ and $I(\xi)$ are even functions the error signal produced will obey the sufficient condition for a null sensor described earlier.

Now consider the case when both $R(\xi)$ and $I(\xi)$ are odd functions of ξ . The same argument as presented for the even symmetry case will hold. Now each integral in (3.16) is purely imaginary, and each integral in (3.19) is purely real. Therefore these equations will again be $0 \forall \psi$ and the error signal will be given by Equation (3.20).

The analysis thus far has placed no limiting assumptions on the input wavefront. From this analysis it is concluded that filter functions whose real and imaginary parts share

the same symmetry will produce useful error signals for the operation of a GPD null wavefront sensor. The next step is to consider what form the error signal will take when the real ($R(\xi)$) and imaginary ($I(\xi)$) parts of $F_{\pm}(\xi)$ do not share the same symmetry.

3.6.2 Mixed Symmetry

Let us examine the form of the error signal when one of $R(\xi)$ and $I(\xi)$ is an odd function and the other is an even function of ξ . Each equation containing cross terms is affected by the mixed symmetry. When $R(\xi)$ is even and $I(\xi)$ is odd, all the integrals in (3.17) are purely imaginary and all the integrals in (3.18) are purely real. When $R(\xi)$ is odd and $I(\xi)$ is even, all the integrals in (3.17) are purely real and all the integrals in (3.18) are purely imaginary. This is the same cancellation effect experienced by Equations (3.16) and (3.19) in the ‘same symmetry’ case.

Equations (3.17) and (3.18) therefore reduce to zero in both cases, and the error signal is given by the sum of (3.16) and (3.19):

$$\begin{aligned} \frac{d(r)}{2i} = & \left[\int d\xi H(\xi) I(\xi) \exp(-ir\xi) \int d\xi' H^*(\xi') R(\xi') \exp(ir\xi') - \right. \\ & \left. \int d\xi H(\xi) R(\xi) \exp(-ir\xi) \int d\xi' H^*(\xi') I(\xi') \exp(ir\xi') \right] + . \quad (3.21) \\ & \left[\int d\xi A(\xi) I(\xi) \exp(-ir\xi) \int d\xi' A^*(\xi') R(\xi') \exp(ir\xi') - \right. \\ & \left. \int d\xi A(\xi) R(\xi) \exp(-ir\xi) \int d\xi' A^*(\xi') I(\xi') \exp(ir\xi') \right] \end{aligned}$$

A plane wave ($A(\xi) \equiv 0$) incident on a mixed symmetry filter would therefore still produce an error signal, given by the upper half of equation (3.21). This violates the sufficient condition outlined earlier for operation of a null sensor; that it produce a null output when the input wavefront is unaberrated. Therefore filter functions with mixed symmetries are unsuitable for use in a GPD wavefront sensor since an error signal will be produced for all input wavefronts.

3.6.3 Necessary and Sufficient Conditions

In this section the conditions identified by the analysis in Sections 3.6.1 and 3.6.2 will be summarised into a set of sufficient and necessary conditions for the GPD null sensor. The next step will then be to test these conditions in simulation with plane and aberrated wavefronts to check that there have been no mistakes in the analysis and assess the performance of the GPD sensor.

Sufficient condition for operation of a null wavefront sensor is that it should only produce an error signal when the input wavefront is non-plane. The necessary conditions which have been identified are that the FT of the filter function ($F_{\pm}(\xi)$) must be complex and furthermore that the real and imaginary parts of this function must have the same symmetry. Any function, or combinations of functions, which satisfy these conditions are suitable for use in a GPD null wavefront sensor.

As a first check of the analysis presented thus far the symmetry of a defocus filter function was calculated and it was found that this possessed the ‘same symmetry’ which produces a useful error signal. This is a useful initial test since defocus is known to work well in the DPD sensor (described in Chapter 2). Had it not met the necessary and sufficient conditions it would have been obvious that something was wrong in the analysis. Throughout this thesis the validity of many new findings are tested by making sure they confirm that defocus would work well.

A programme was written to calculate the symmetry conditions of Zernike polynomials from piston (Z_0^0) to the high order aberration Z_{10}^{10} (some 65 possible filter functions in total) to find which ones would satisfy the conditions for a GPD null sensor. The filter function was created in the form $e^{-i.2\pi.\varphi_d}$, where φ_d is the diversity phase described by a chosen Zernike polynomial. It was discovered that Zernike polynomials whose indices (n and $n - 2m$ i.e. Z_n^{n-2m}) are positive and even have suitable symmetry for use in a GPD sensor, e.g. Z_6^2 is suitable but Z_6^{-2} and Z_3^2 are not. In the suitable polynomials it was observed that the polar radius ρ is raised to an even power and that there is an even

multiple of the polar angle θ . For example Defocus ($Z_2^0 = 2\rho^2 - 1$) satisfies the symmetry conditions whereas Triangular Astigmatism ($Z_3^{-3} = \rho^3 \sin(3\theta)$) does not. This provides a quick and easy method to identify suitable Zernike functions. In total it was found that 21 of the 65 polynomials tested would be suitable for a GPD sensor. This represents a large range of possibilities to be explored within the set of Zernike's alone; although functions other than Zernike polynomials, that satisfy the conditions, could also be used.

In the following section the performance of several Zernike filter functions will be tested to look at the form of the error signal, and confirm that the filter functions identified as having 'same symmetry' do provide a null signal when the wavefront is plane and a significant error signal when it is aberrated. Simulations with mixed symmetry filters will also be tested to confirm that a significant error signal is produced for un-aberrated wavefronts.

3.7 Simulations to verify the Sufficient and Necessary Conditions

Matlab simulations were conducted to verify the analysis which led to the sufficient and necessary conditions detailed in Section 3.5.3. A range of diversity functions were created with the form:

$$e^{-i.2\pi.\varphi_d} , \quad (3.22)$$

the diversity phase, φ_d , is a Zernike polynomial or a combination of Zernike's. As mentioned before, the symmetry of the real and imaginary parts of each diversity function was tested to identify which had same, and which had mixed symmetry. These functions were then used in further simulations to study the error signals, $d(r)$, generated by a pupil plane GPD wavefront sensor for plane and aberrated input wavefronts. A schematic of this process is shown in Figure 3.5.

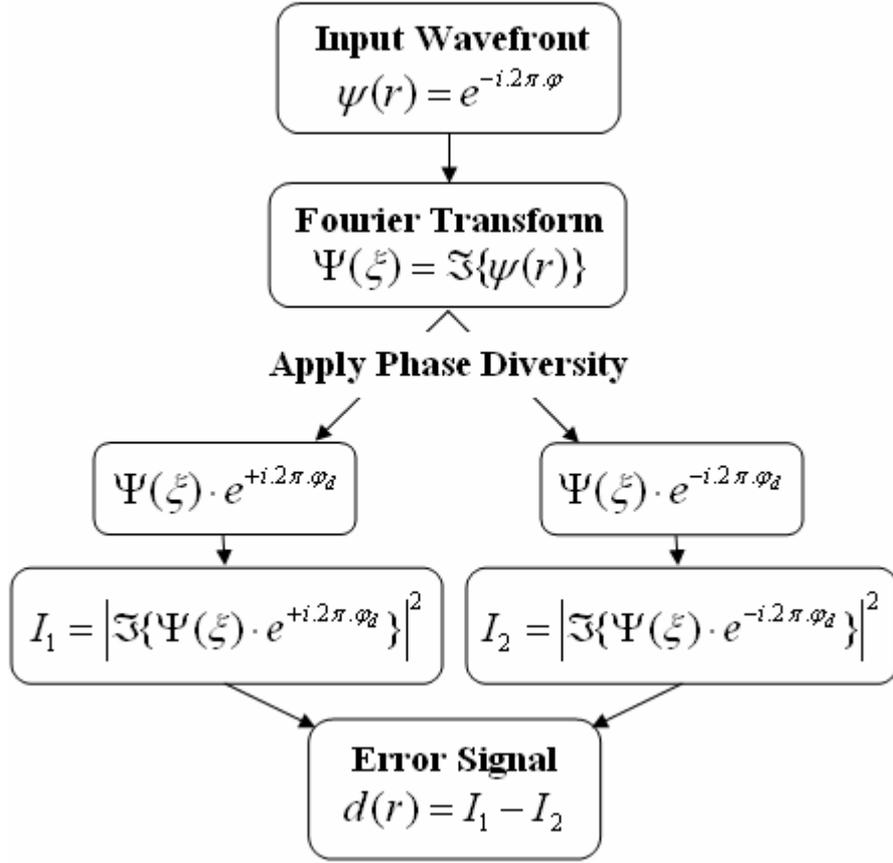


Figure 3.5 Schematic of the GPD simulation used to verify the Symmetry conditions.

If the symmetry conditions described in 3.5.3 are correct then, for same symmetry filter functions $d(r)$ should be zero for plane wave inputs and have significant value for aberrated inputs. Mixed symmetry filter functions should give non-zero output for plane and aberrated input wavefronts.

3.7.1 Plane Wave Input Simulation Results

In this simulation single Zernike filter functions, and ones created by a combination of two Zernike's were used to test the symmetry conditions identified in Section 3.5.3. The diversity functions used to generate the error signals shown in Figure 3.6 (a,b) have even, same symmetry. The filter functions chosen were defocus ($Z_2^0 \sim$ shown in 3.6(a)) and a combination of defocus and astigmatism ($Z_2^0 + Z_2^2 \sim$ 3.6(b)). The error signals

depicted in Figure 3.6 (c,d) were generated with mixed symmetry filter functions, these were tilt ($Z_1^{-1} \sim 3.6$ (c)) and a combination of triangular astigmatism and the higher order Zernike Z_5^{-5} ($Z_3^{-3} + Z_5^{-5} \sim 3.6$ (d)).

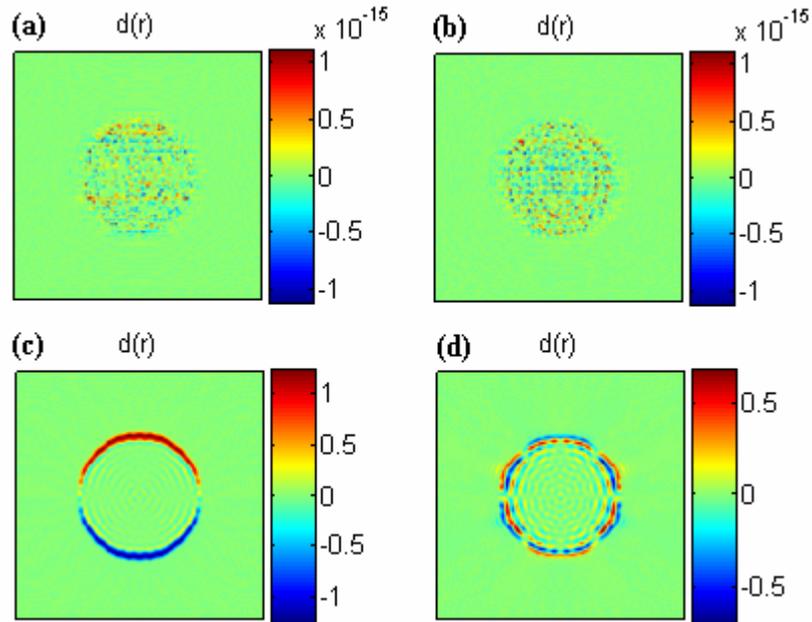


Figure 3.6 Simulation results showing $d(r)$ for a plane wave input with: a) Z_2^0 Defocus b) $Z_2^0 + Z_2^2$ Defocus + Astigmatism c) Z_1^{-1} Tilt and d) $Z_3^{-3} + Z_5^{-5}$ diversity phase functions. [Scale in each plot is a measure of contrast].

Note that the internal structure evident in Figure 3.5 (c,d) (and in subsequent figures) is caused by diffraction effects from the hard edged pupil function used throughout.

As the results in Figure 3.6 show, for the even (same) symmetry filter functions tested the detected error signal was zero, to within computer rounding error which satisfies the sufficient condition for a null sensor. The results in Figure 3.6 (c,d) show that, when using these mixed symmetry filters, a significant error signal was generated for a plane wave input.

Chapter 3: Generalised Phase Diversity

These simulations are representative of a large selection of similar tests which all supported the analysis presented thus far; that functions with same symmetry provide a null output for a plane wave input and mixed symmetry filters are unsuitable for use in a null sensor.

3.7.2 Aberrated Wavefront Simulation Results

The next step is to test the performance of some same symmetry filter functions with different aberrated wavefronts to study the form of the error signal produced. In each case the simulation was set to produce the error signal from a pupil plane sensor configuration (see Section 3.2.2 for details).

3.7.2.1 Aberrated Wavefront 1

The first aberrated wavefront, with the form $\psi(r) = e^{-i.2\pi.\varphi(r)}$, was created using linear a combination of coma, astigmatism and triangular astigmatism ($\varphi(r) = Z_2^{-2} + Z_3^{-1} + Z_3^3$). This test wavefront may be seen in Figure 3.7 below, the scale is given in waves of error:

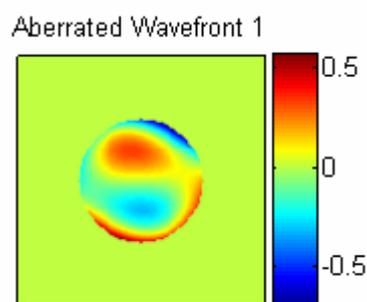


Figure 3.7 Aberrated Wavefront 1, created using a linear combination of Z_2^{-2} , Z_3^{-1} & Z_3^3 .

[Scale is in waves of error.]

Using Aberrated Wavefront 1 as the test wavefront, the error signal generated using several same symmetry filters was tested. The filter functions were created as described in Section 3.7. In all the examples presented in Figure 3.8 the strength of the diversity was set at 1λ PV. The same symmetry functions used were defocus (Z_2^0), astigmatism (Z_2^2), the higher order aberration (Z_6^0), and a linear combination of Spherical aberration and astigmatism ($Z_4^0 + Z_2^2$). Results for these are shown in Figure 3.8 (a-d) respectively.

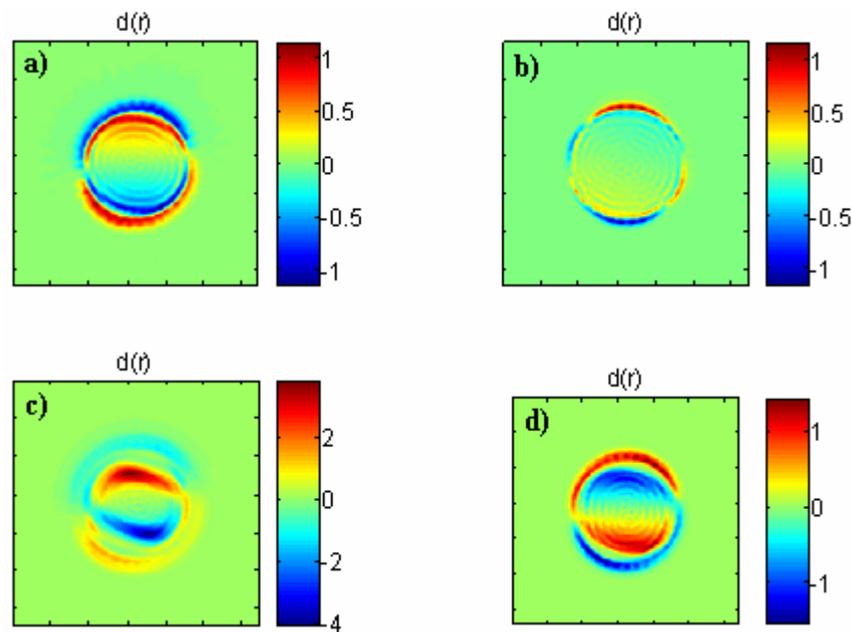


Figure 3.8 Simulation results showing $d(r)$ for: a) Z_2^0 defocus b) Z_2^2 astigmatism c) higher order aberration Z_6^0 d) $Z_4^0 + Z_2^2$ A combination of spherical aberration and astigmatism diversity phase functions. [Scale is a measure of contrast].

Figure 3.8 clearly shows that in each case a significant error signal is generated. This agrees well with expectations that the same symmetry filters will produce a useful error signal. Many same symmetry filter functions were studied and the same conclusion was made in each case. Figure 3.8 demonstrates that the error signal obtained for each different filter function will have a different form, some more faithfully representing the input wavefront shape than others. The amplitudes of the error signals also vary when using different diversity filters. Discussion of this effect and its significance is the subject of Section 3.8.2 and will continue in Chapter 4.

3.7.2.2 Aberrated Wavefront 2

The process described in Section 3.7.2.1 was repeated for a second aberrated input wavefront, whose phase was given by $\varphi(r) = Z_7^{-3} + Z_3^3 + Z_3^{-1}$. This wavefront is shown in Figure 3.9.

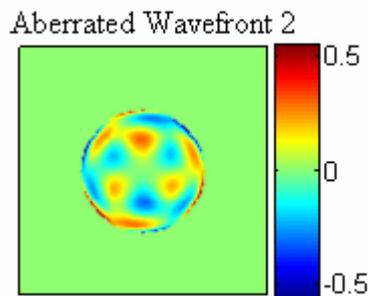


Figure 3.9 Aberrated Wavefront 2, created using a linear combination of Z_7^{-3} , Z_3^3 & Z_3^{-1} . [Scale is in waves of error.]

Figure 3.10 shows the error signals generated for a second set of same symmetry filter functions. In this case the same symmetry functions used were: defocus (Z_2^0), spherical aberration (Z_4^0), a higher order radially symmetric aberration (Z_8^0) and a linear combination of defocus and another higher order aberration ($Z_2^0 + Z_{10}^0$). The results of these simulations are plotted in Figure 3.10 (a-d) respectively. As before, in each case the strength of the diversity phase was set at 1λ PV. As in the first aberrated wavefront example it is seen that a significant error signal is generated in each case, in accordance with expectations.

It is interesting to note in this example that the best performance, in terms of similarity between the error signal and the actual wavefront error shape, is given by the combination between spherical aberration and defocus with a reversal in the contrast between the two. In each simulation the performance of defocus diversity was carefully

studied since this was known to perform well in practice. Defocus was seen to work very well in each example, and it was also found that spherical aberration performed in a similar way. This may point to an advantage in using radially symmetric Zernike polynomials as the diversity phase. The issue of choosing the correct diversity function, and optimisation for different applications will be considered in Section 3.9. The apparent change in contrast between the error signals generated by defocus and spherical aberration (Figure 3.10 (a) & (b)) is also caused by the difference in structure between these two functions and this is a topic that will be revisited and explained later when the form of the filter function is studied in more detail.

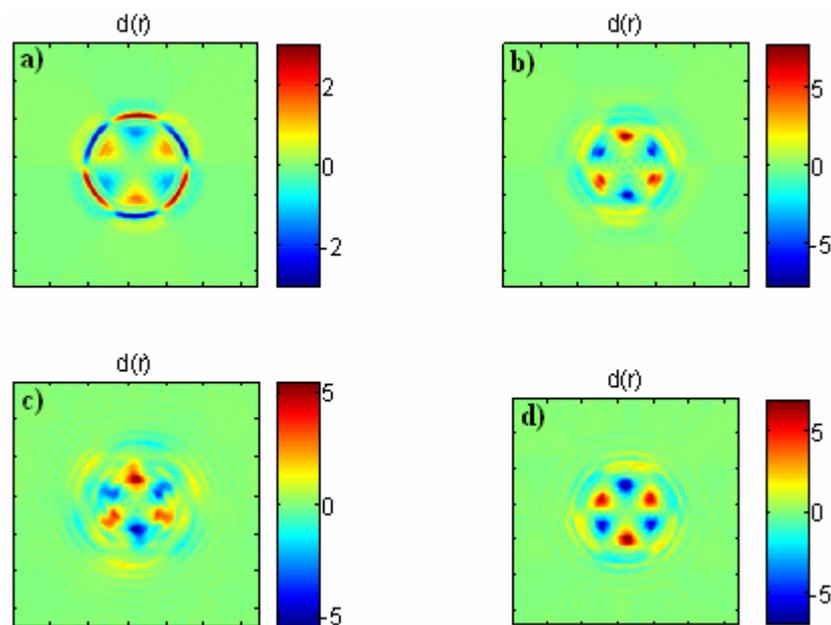


Figure 3.10 Simulation results showing $d(r)$ for: a) Z_2^0 defocus b) Z_4^0 spherical aberration c) Z_8^0 the higher order aberration d) $Z_2^0 + Z_{10}^0$ a combination of defocus and the higher order aberration Z_{10}^0 [Scale is a measure of contrast].

In conclusion, the simulations presented in Sections 3.7.2.1-2 have shown that a significant error signal is created when an aberrated wavefront is present, thus supporting the last of the conditions identified in Section 3.6. Now that a method for choosing allowable diversity functions has been demonstrated the next step is to study these in more detail. Some questions which should be addressed include: How much

information is encoded in this error signal? How does performance vary when using different functions? What effect does the physical structure of the function have on this performance? Can a diversity function be chosen to optimise sensitivity in particular applications? These are important questions and some will be revisited in Chapter 4 where a phase retrieval algorithm is developed and the ‘performance’ of the wavefront sensor can be tested in terms of its ability to produce a signal from which the original phase can be accurately reconstructed.

Before continuing to look at the answers to some of these questions, an interesting problem will be demonstrate which occurs when the input wavefront contains phase steps of exact multiples of π .

3.7.3 Wavefronts with π Phase Steps

An interesting ambiguity occurs when there is a phase step of exactly π ($\lambda/2$) in the wavefront under test. A change of π corresponds to a multiplication of the wavefront by -1, the wavefront is still real valued and its FT therefore has Hermitian symmetry. Thus the GPD wavefront sensor will be inherently insensitive to π phase steps in the test wavefront and the error signal, $d(r)$, will be zero for every exact multiple of the phase by $\lambda/2$. Presumably this would mean that the error signal is maximised for phase errors of $\lambda/4$, thus giving the error signal a sinusoidal response to the size of the phase step. This hypothesis will be studied in more detail in Chapter 4.

Figure 3.11 is a 1D cross section through a pupil plane phase profile with piston errors used to create a phase step of 0.5λ in the central portion of the circular phase profile. The input wavefront, whose cross section is shown in Figure 3.11 was used in a pupil plane wavefront sensor simulation, with an even symmetry filter function (Z_2^0 , defocus) to study the error signal created. Figure 3.12 shows the results of this simulation.

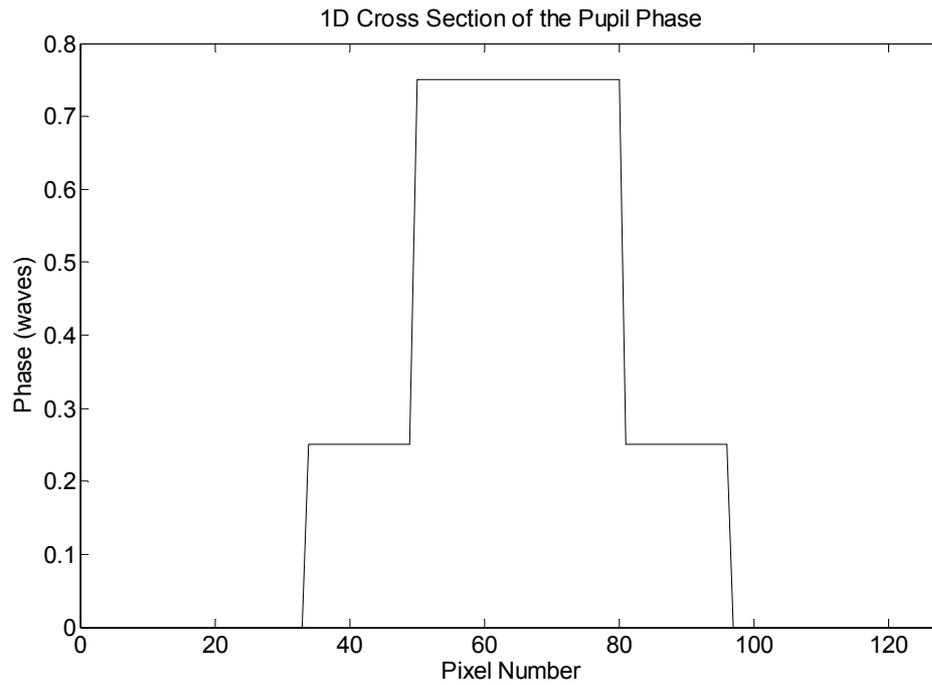


Figure 3.11 1D cross section of a circular pupil plane phase profile with piston errors

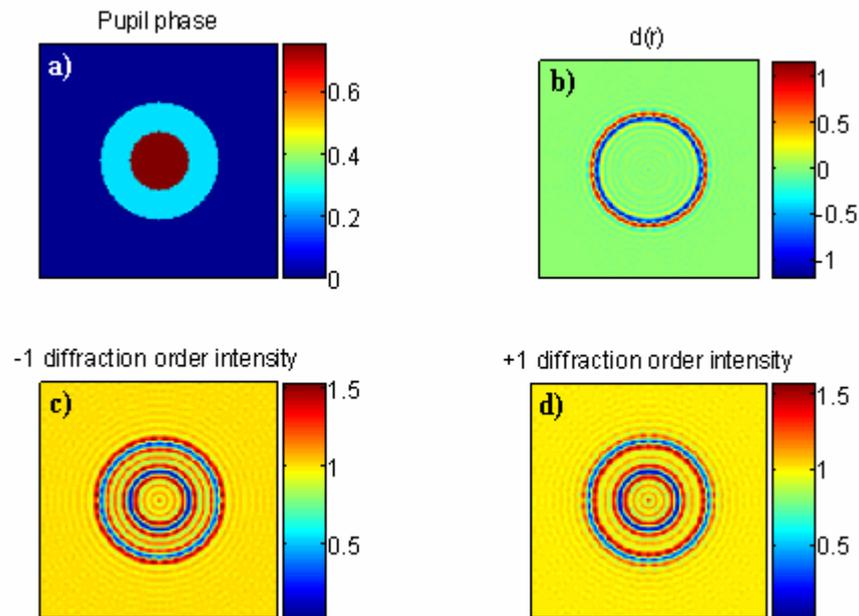


Figure 3.12 The results of a pupil plane GPD simulation for input phase (a), using a defocus diversity function. The simulated phase diverse images are shown in (c) and (d). The error signal, $d(r)$ ($= (c)-(d)$), is given in (b).

As Figure 3.12 clearly shows the error signal contains no evidence of the π phase step. However, the phase step is visible in the phase diverse intensity images. Since the

images are identical, except around the edges, taking the difference will subtract out the features caused by the phase step. Repeat simulations were conducted, and in every case where a phase step of any exact multiple of π was used the result was the same. This shows that there is a fundamental blindness in the GPD error signal to piston errors of exact multiples of π , in any input wavefront.

For astronomy applications an insensitivity of this kind does not present a fundamental problem. In metrology, this insensitivity would be a problem. However, in any real system, the chance of obtaining a piston error of exactly π is very small.

3.8 The Error Signal

It has been shown that aberration functions which satisfy the conditions outlined in Section 3.6 may be used to construct a null GPD wavefront sensor. From the analysis presented in Section 3.7 it may be supposed that the error signal is periodic with π phase errors. This is confirmed later by simulation in Chapter 4, and its significance studied in more detail. At this point it may be said that signal interpretation is not trivial and has some context dependence. If, for example, the wavefront errors are continuous then the π phase errors will not occur. In cases where the errors are discontinuous, for example the phase errors associated with misalignment of large segmented telescope mirrors, the π phase ambiguity may occur and will be missed by this method.

In the following section the information encoded in the error signal, the sense and location of the error, will be discussed. Also, the form of the aberration function and its effect on the error signal will be introduced. This will be continued in Chapter 4 where we consider extending the null sensor to full wavefront sensing by developing a phase retrieval algorithm.

3.8.1 Error Sense and Location

The sense of the error is encoded in the error signal. This can be seen by considering Equation (3.20). This equation gives the form of the error signal generated by a suitable GPD filter function. From this it may be seen that the error signal, $d(r)$, will change sign if the sense of the wavefront error is reversed. This is due to the symmetry of $A(\xi)$, given in Equation (3.15). This means that the ‘sense’ (or direction) of the wavefront error is directly linked to the error signal.

The location of the wavefront error can also be found from the error signal. The error location is directly related to $a(r)$, the imaginary part of the input wavefront and the transform of $A(\xi)$. If the filter function is concentrated or peaked around the origin (as it is for the radially symmetric aberrations like defocus for example) then the position of the wavefront error signal will be localised around the point that $a(r)$ is non-zero. In Chapter 4 the filter function and the formation of the error signal is considered in much more detail and this question of error location will be revisited.

Therefore it can be concluded that the error signal alone, without full reconstruction, provides a versatile tool for wavefront sensing. Since the sense and location of the wavefront error is included in the signal this information could be used to tell a deformable mirror whether to ‘push’ or ‘pull’, and at which position, to flatten the wavefront. As discussed in Section 3.2.4, full reconstruction of the wavefront is often unnecessary for AO applications. Therefore null sensor operation could prove to be a faster more efficient method for correcting the wavefront distortions.

3.8.2 The Diffraction Grating and the Blur Function

In the GPD wavefront sensor the error signal, $d(r)$, is generated by the difference of a pair of intensity images formed in the ± 1 diffraction orders. These intensity images are themselves created by the convolution of the input wavefront with the filter function programmed into a diffraction grating. To look into this further and begin to study the effect of the form of the filter function on the performance of the GPD sensor,

simulations were conducted to compare the filter function for defocus and spherical aberration.

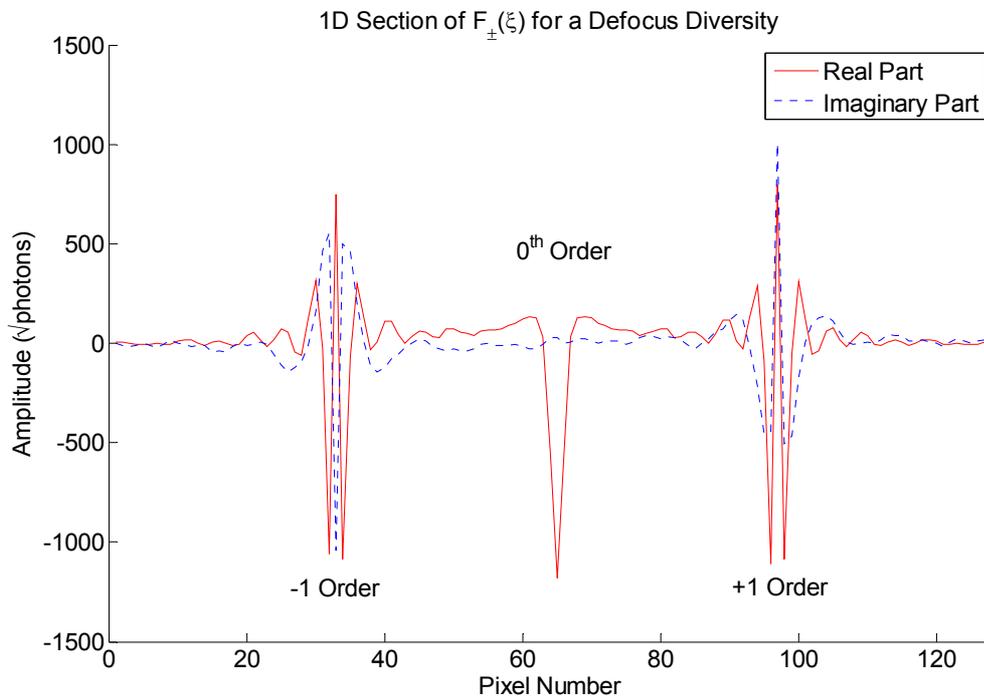


Figure 3.13 Cross section of the FT of a defocus diffraction grating.

Figure 3.13 shows a cross section through $F_{\pm}(\xi)$, the FT of a diffraction grating programmed with a defocus filter function. This figure shows the behaviour of the real and imaginary parts of the filter function. In the GPD sensor the intensity images in the ± 1 diffraction orders are created by the convolution of the side lobes of this function (labelled -1 order and +1 order in this figure) and the input wavefront. These side lobe functions will be referred to as the ‘blur’ function in reference to the effect they have when convolved with the input wavefront. The form of this blur function will have implications for the sensitivity of the sensor and this discussion will be continued in the following chapters. GPD allows a wide choice of diversity functions and by examining the differences between the blur functions from two different aberration kernels the effect on the wavefront sensor sensitivity can be studied.

Figure 3.14 is a comparison of the profile of two different blur functions, defocus and spherical aberration. Note that the PV distortion of the diversity phase in both cases was the same (1.5λ) for a meaningful comparison.

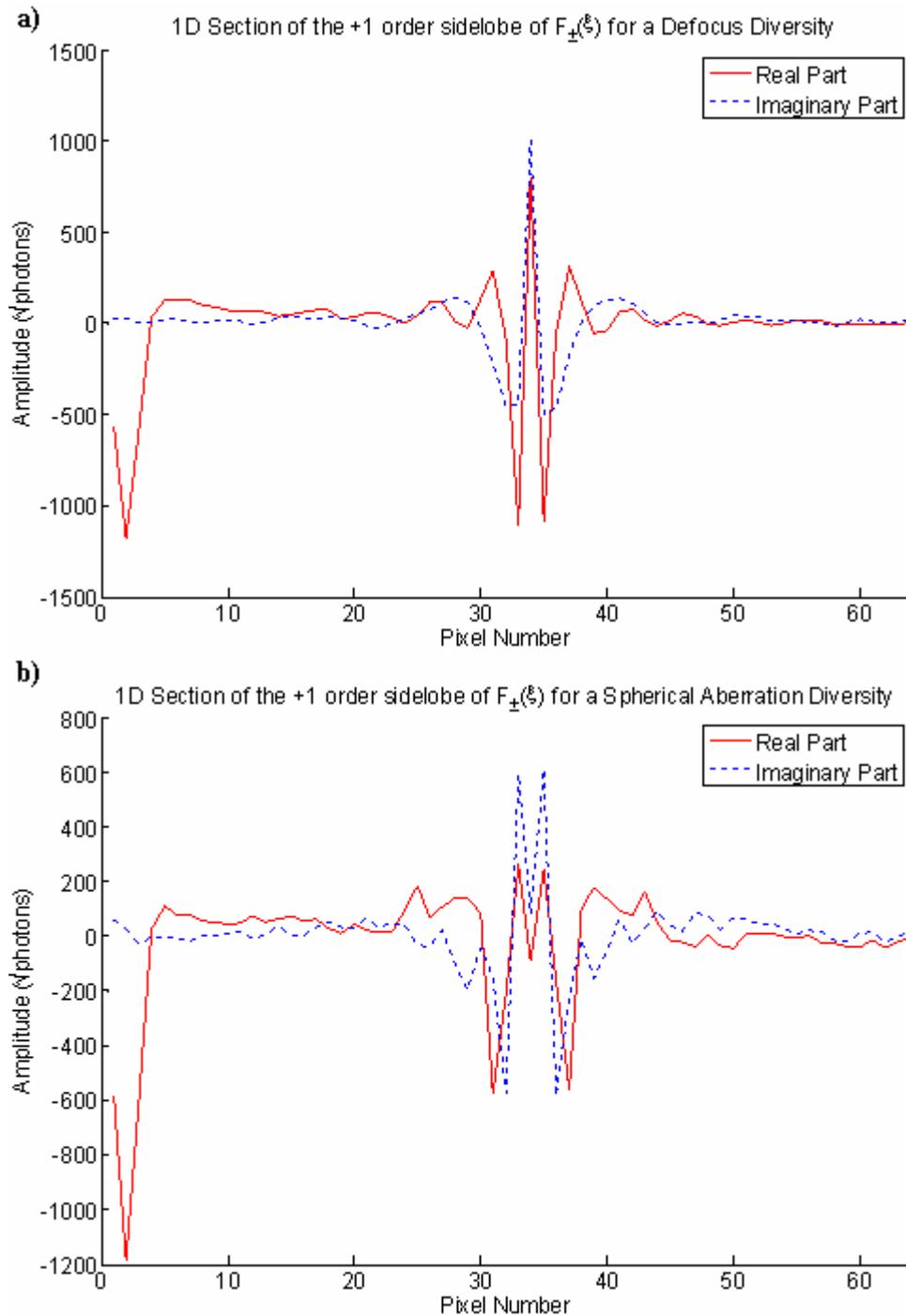


Figure 3.14 The profile of the blur function for (a) a defocus filter function and (b) a spherical aberration filter function.

Firstly, note that both the blur functions plotted in Figure 3.14 are for the same sidelobe (+1) but there is a ‘reversal’ between the real and imaginary parts of the two examples. This can be used to help explain the change in contrast between the error signals shown earlier in Figure 3.10. In this figure the sense of the error signal generated by the spherical aberration filter (Figure 3.10(b)) matches the sense of the input wavefront (Figure 3.9) and the Defocus error signal is the exact opposite. We can see now that this is due to the reversal in sign between the real and imaginary parts of $F_{\pm}(\xi)$. Since the grating can be rotated there is no unique way to determine which intensity image is the +1 order and which is the -1 order. Therefore it is simply a question of which order the intensity images are subtracted that determines which example, defocus or spherical aberration, will match the input wavefront and which will have the opposite contrast.

The width of the blur function determines the area over which the convolution of the input wavefront and the diversity function occurs. The wider this blur function is, the larger the area used for the convolution. Wider blur functions should therefore be more sensitive to small slopes on the input wavefront. Whereas narrow blur functions will be more suitable for use with input wavefront with high slope values. Therefore, given the blur function shown in Figure 3.14 it could be supposed that the spherical aberration diversity function should prove to be the more sensitive of the two.

To test this hypothesis a computer simulation was conducted to compare the sensitivity of the two filter functions. The simulations were all conducted for a pupil plane configured wavefront sensor. The same aberrated wavefront was input into both simulations. It was created using a mixture of 3 Zernike polynomials (Z_2^{-2}, Z_5^5, Z_9^1), and had a peak to valley (PV) distortion of 1.4λ . The input pupil phase is shown in Figure 3.15(a) and the scale for this plot is in waves of wavefront error. In 3.15(b) the error signal for a defocus filter function, with PV distortion of 1λ is plotted. This signal shows good contrast and has a PV difference of 6.1 (measured in arbitrary units). Figure 3.15(c) is the error signal generated by a spherical aberration diversity filter with the same PV (1λ), however in this case the PV error signal generated is much larger, 15.7 (arbitrary units). This is a significantly higher signal for the same input wavefront and diversity strength. As a final comparison the amplitude of the spherical aberration filter function was adjusted until an error signal with the same PV as the error signal in

3.15(b) (which was generated using the defocus filter function) was created. This is shown in 3.15(d) and the filter function used has a PV distortion of 0.35λ . Comparison of (b) and (d) shows that, although the overall PV contrast of the error signals is the same the spherical aberration filter function has generated a signal with sharper contrast.

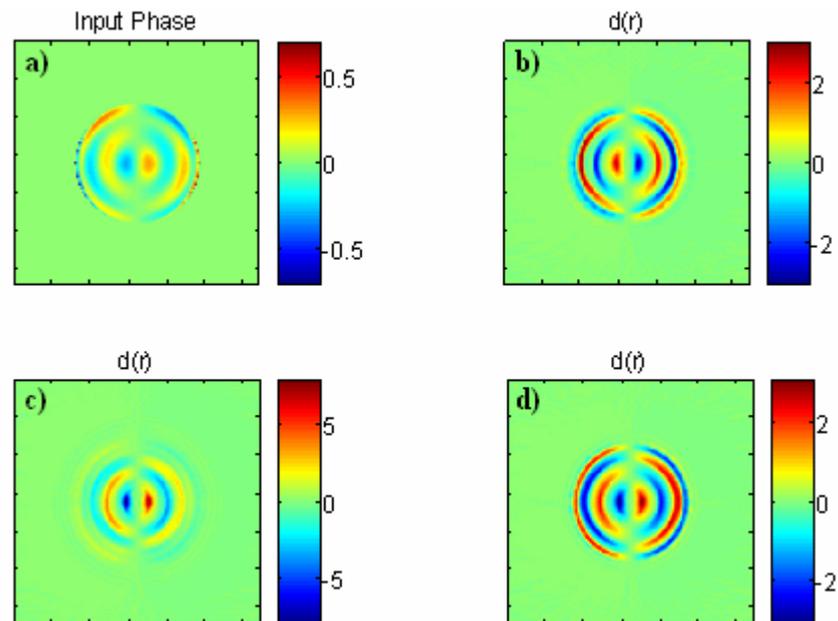


Figure 3.15 The error signals generated in a pupil plane configured GPD wavefront sensor for an aberrated wavefront (a) [scale is in waves] with (b) a defocus filter function with PV 1λ , (c) A spherical aberration filter also with PV 1λ , and (d) a spherical aberration filter with PV of 0.35λ [for figures (b) to (d) the scales are a measure of contrast]

In the simulation using the defocus diversity function (3.15(b)), the sense of the wavefront error appears to be inverted. Areas in the pupil phase which are shown coloured blue (i.e. negative), are shown in red (i.e. positive) in these simulations. This is another example of the contrast reversal problem described earlier.

From these results it may be concluded that, in this case, the spherical aberration filter has shown better performance than defocus. This supports the previous statement that filters with wider blur functions should have greater sensitivity to small deformations in the wavefront than filters with narrow blur functions. These results also suggest that a

spherical aberration based GPD sensor would be able to cope with poorer signal to noise than a defocus based one as the error signal it generates is significantly larger and has better contrast.

3.9 Optimisation

In Section 3.8 simulations were presented that clearly showed that each diversity filter performs differently, and that this is linked to its structure. This will be explored in greater detail in the following chapters where this method is extended to full wavefront sensing. Since the sensitivity of the error signal seems to be linked to the structure of the diversity function it may be possible that certain diversities will prove to be naturally more suited to sensing particular aberrations. This conclusion leads to the obvious question, is it possible to optimise the filter function given *a priori* information about the wavefront errors? In many astronomy and metrology applications a great deal is known about the errors that are most likely to arise, and also which of these have the greatest detrimental effect on the desired outcome. It may prove more useful to provide a sensor which is most sensitive to these modes rather than one which gives adequate performance for a large number of modes.

In order to investigate this further a series of Matlab simulations were conducted, using input wavefronts of known aberration and a selection of diversity functions which satisfy the symmetry conditions outlined in Section 3.6.3. The input wavefronts were chosen to contain an element which matched at least one of the diversity filters used to see whether matching the shape of the diversity function to the wavefront error would produce greater sensitivity to that particular aberration.

In the first set of results (Case 1), shown in Figure 3.17 the input wavefront contains just one aberration, astigmatism (Z_2^2). This is then extended in Figure 3.18 which shows a range of single aberration input wavefronts with matching and non matching diversity functions. Finally, a more complex input wavefront, comprising a mixture of 3 different aberrations (Z_3^{-1} , Z_6^2 , and Z_8^4), was tested. The results of this are shown in Case 2 (see Figure 3.20).

3.9.1 Case 1

A distorted input wavefront was generated, using astigmatism Z_2^2 , with a PV wavefront error of 2λ , this is shown in Figure 3.16. This was used in a pupil plane GPD simulation with diversity applied to generate a pair of intensity images, and from these obtain the difference function $d(r)$. A schematic of this process is shown in Figure 3.5. The diversity functions chosen were defocus (Z_2^0), spherical aberration (Z_4^0), astigmatism (Z_2^2 , same as the input wavefront), and the higher order Zernike polynomial Z_8^0 . In each simulation the strength of the diversity function used was 1λ PV.

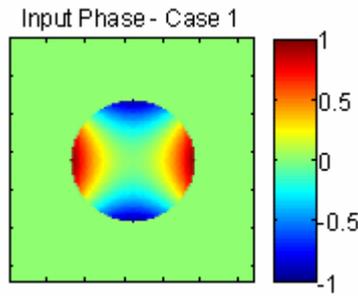


Figure 3.16 Astigmatism (Z_2^2) with a PV of 2λ , used as the input pupil phase [scale is in waves]

Figures 3.17 (a),(b) and (c) show that the radially symmetric diversity functions perform well, and create error signals that indicate the structure of the pupil phase. These results also show that the higher order radially symmetric aberrations create higher PV error signals i.e. Z_8^0 gives a larger error signal than Z_4^0 which in turn generates a large $d(r)$ than Z_2^0 .

In terms of the error signal shape resembling the pupil phase, Figure 3.17 shows the poorest performance is given by the astigmatic diversity function. In 3.17(d) the central structure is dominated by the diffraction effects from the hard edged pupil aperture. There was no structure in the intensity images for this example either, so it is not like the π phase step case (see Section 3.7.3) where taking the difference caused the central

structure to cancel. Therefore, in this example, matching the diversity function to the wavefront shape has been shown to decrease the sensitivity.

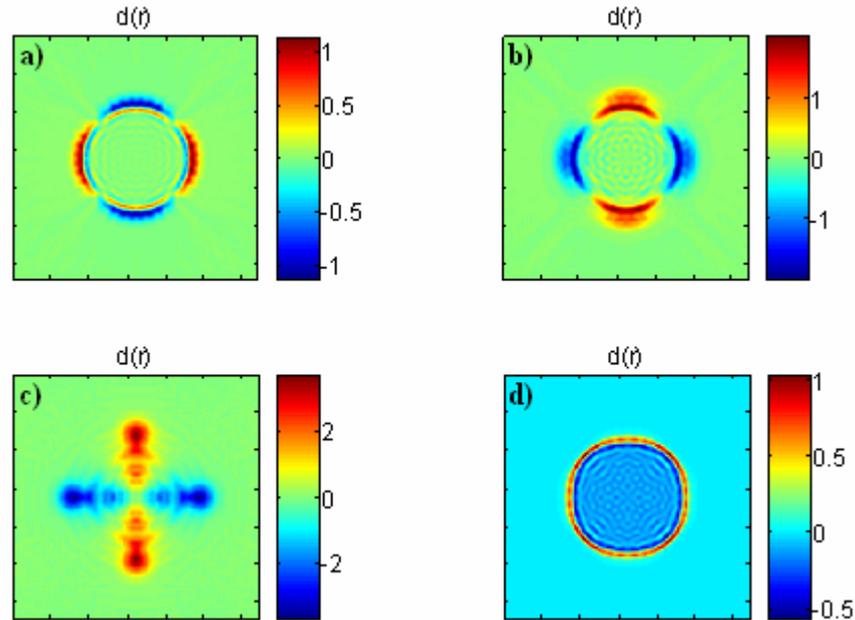


Figure 3.17 The difference functions $d(r)$ generated by applying (a) defocus (b) spherical aberration (c) Z_8^0 and (d) astigmatism, to the input wavefront shown in 3.16 [Scales are a measure of contrast]

To investigate this further a large number of simulations were performed to take single aberration input wavefronts and look at the error signal for a diversity function which matches the wavefront shape, and one non-matching diversity function. The results for some of these are contained in Table 3.1.

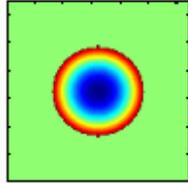
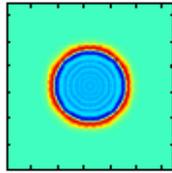
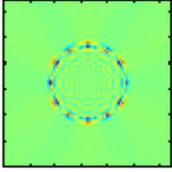
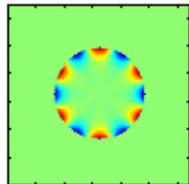
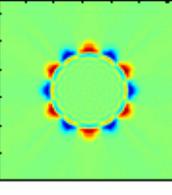
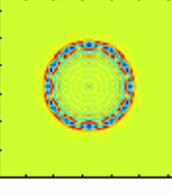
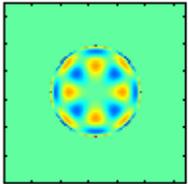
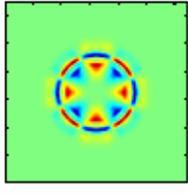
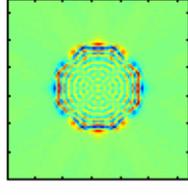
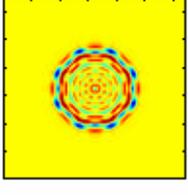
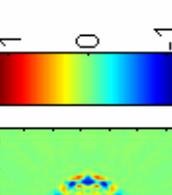
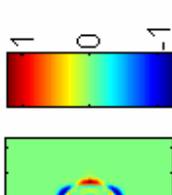
		Difference image $d(r)$ generated by applying 1λ PV of the following diversities:		
		Z_2^0	Z_6^6	Z_8^4
Input Wavefronts created using 1λ PV of these input phase profiles:	Z_2^0	 Input Phase	 $d(r)$ PV of $d(r) = 1.9$	 $d(r)$ PV of $d(r) = 1.2$
	Z_6^6	 Input Phase	 $d(r)$ PV of $d(r) = 3.1$	 $d(r)$ PV of $d(r) = 0.7$
	Z_8^4	 Input Phase	 $d(r)$ PV of $d(r) = 5.2$	 $d(r)$ PV of $d(r) = 1.3$
			 $d(r)$ PV of $d(r) = 3.8$	 $d(r)$ PV of $d(r) = 2.3$
			 $d(r)$ PV of $d(r) = 2.4$	

Table 3.1 A Pictorial Matrix of simulated results demonstrating the effect of matching the diversity function to the phase profile of the input wavefront.

Chapter 3: Generalised Phase Diversity

As Table 3.1 shows, in each case it was found that when a single aberration wavefront and a matching diversity function was used the error signal was poorer than when a non-matching diversity was used. It is easy to see why this should be the case. The intensity images, whose difference generates $d(r)$, are formed by applying equal and opposite amounts of the diversity phase to the input wavefront (i.e. I_+ comes from $\varphi + \varphi_d$ and I_- from $\varphi - \varphi_d$). When the diversity phase and wavefront phase are the same (e.g. $\varphi = +\varphi_d$), in one case the two phases will reinforce resulting in a large aberration in one of the intensity images (I_+). However, in the other image (I_-) adding the diversity phase will act like phase conjugation and the two will cancel to leave something close to a plane wave. Thus when $d(r)$ is calculated it will not be very sensitive or a good representation of the pupil phase shape.

3.9.2 Case 2

In the final set of simulations a more complex pupil plane phase was created from a mixture of Zernike's, Z_6^2, Z_8^4 and Z_3^{-1} :

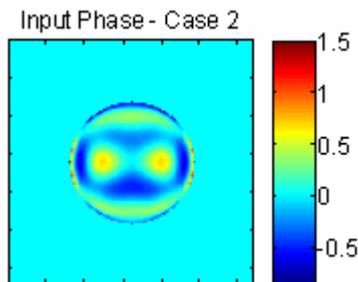


Figure 3.18 Z_6^2, Z_8^4 and Z_3^{-1} with a PV of 2.35λ , used as the input Pupil Phase [scale is in waves]

By using a pupil phase with 3 component aberrations, it was possible to use 2 of these as diversity functions to see if matching the diversity shape to an element of the aberrated wavefront would offer any advantage. Since the wavefront no longer comprises just a single aberration the 'phase conjugation effect' which caused

cancelling in the previous simulations should be less apparent. As before, the PV of each diversity function tested was set to 1λ .

As in previous simulations the radially symmetric diversity functions defocus and spherical aberration (see Figure 3.19 (a-b)) have performed well. The results in 3.20(c) and (d) used diversity functions created from some of the component aberrations in the input wavefront. In these examples the structure in the error signal does bear some resemblance to the pupil plane phase. The Z_6^2 filter (c) has produced the better signal than the Z_8^4 filter (d) which is an interesting result since this was the largest of the 3 aberrations used to create the pupil phase (approximately twice as much of Z_6^2 as Z_8^4 was used).

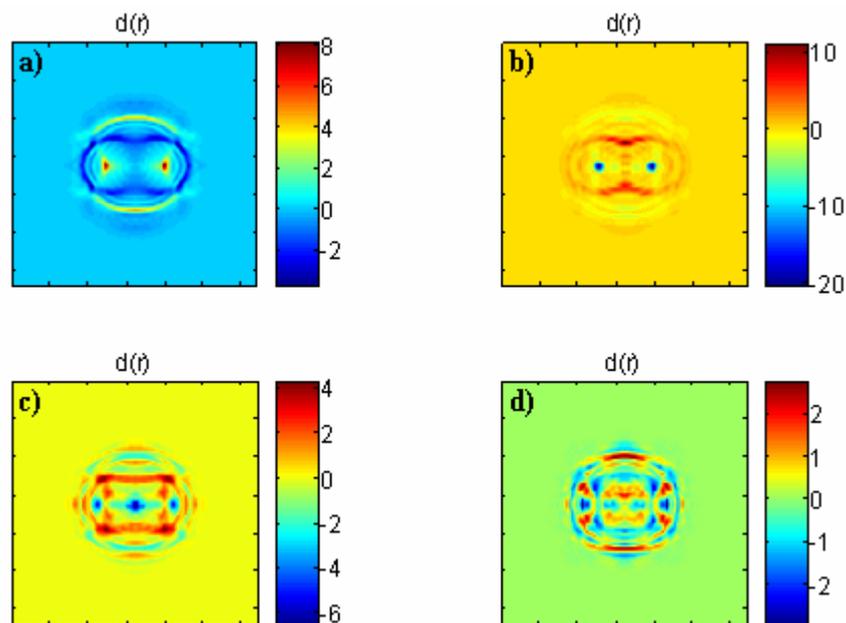


Figure 3.19 The difference functions $d(r)$ generated by applying (a) defocus (Z_2^0) (b) spherical aberration (Z_4^0) (c) Z_6^2 (d) Z_8^4 to the input wavefront shown in Figure 3.18 [Scales are a measure of contrast]

These results are representative of a wide range of input wavefronts and diversity functions tested. In each case, when the input wavefront contained more than one

Chapter 3: Generalised Phase Diversity

aberration, a diversity filter chosen to match one of these component aberrations showed signs of the pupil phase structure with varying levels of accuracy. However, it has also been demonstrated that if only a single aberration is present in the wavefront then matching the diversity function to this will not give the best performance.

3.9.3 Optimisation Conclusions

It is known that the choice of aberration function will greatly affect the performance of a GPD wavefront sensing system [9]. The question is, how then should the user make the best choice of diversity function?

From the results presented here it has been shown that the radially symmetric aberration functions ($Z_2^0, Z_4^0, Z_6^0, Z_8^0$ & Z_{10}^0) perform consistently well, providing error signals with good contrast and with structures that clearly resemble the pupil plane input phase which generated them. In cases where there is no *a priori* information about the aberrations in a given application, one of these diversity functions would make a good choice for all round performance. The higher order aberrations have been shown to give larger error signals, so in cases where the wavefront aberrations are small a diversity function such as Z_6^0 should perform very well. For users concerned with consistently good performance, the radially symmetric functions are a sensible choice.

If the user does have *a priori* information about the wavefront aberrations, which is often the case, these results would suggest that there is some potential for optimisation of the diversity function. If one aberration in particular is known to cause the most problems, choosing a diversity function that matches the shape of this aberration (if it is allowed by the symmetry conditions) should give a slightly greater sensitivity to this aberration over others.

For use in a pupil plane null sensor, certainly visually, the radially symmetric diversity functions appear to be the best choice to preserve the actual wavefront error location on the difference signal. These functions would be the most effective ones to choose if the information is used to directly drive a corrective element to flatten the aberrated

wavefront. However, if the user intends to perform wavefront reconstruction prior to correction then the initial appearance of the error signal is of little importance. In this situation, choosing a diversity function which matches the most problematic aberration(s) expected in the system may prove more advantageous than using a radially symmetric function.

3.10 Conclusions

It has been demonstrated that it is possible to create a diffraction grating based PD wavefront sensor using aberration kernels other than defocus. The analysis presented here derived the necessary and sufficient conditions that an aberration filter function must satisfy in order to produce a useful error signal for a null wavefront sensor.

It was shown that the FT of the filter function ($F_{\pm}(\xi)$) must be complex, and that the real and imaginary parts of this function ($R(\xi)$ and $I(\xi)$) must have the same symmetry. Any function, or combination of functions, which satisfy these conditions may be used in the GPD null wavefront sensor. This is why this method has been named Generalised Phase Diversity, since it no longer relies solely on defocus to apply the diversity phase to the wavefront under test.

The difference between operating the wavefront sensor in the pupil and image planes has been discussed. The pupil plane configuration has the advantage that there is a direct relationship between the position of the wavefront error on the pupil, and its position on the error signal. This 1-1 mapping allows the error signal to be used to directly control a corrective element without the need for reconstruction of the wavefront. When using defocus as the diversity function the image plane sensor can be constructed to have a large enough separation between the sampling planes to image close to the pupil plane and thus also create a 1-1 mapping between error signal and wavefront error. However this is limited to defocus, and therefore for a GPD wavefront sensor the pupil plane arrangement would be more suitable in general.

Chapter 3: Generalised Phase Diversity

The analysis presented describing the form of the error signal is so far completely general and places no limiting assumptions on the input wavefront. It has been shown that the error signal is insensitive to piston errors of exact multiples of π , but this is not considered to be a fundamental problem since the chances of this occurring in a practical situation are small. It could also be solved by creating diversity phase measurements using 2 or more different wavelengths [10, 11].

The simulations presented here have demonstrated that different filter functions have different sensitivities, related to their structure and the width of the blur function they produce. It has been shown that filters with wider blur functions will, in general, be more sensitive to small-amplitude errors on the wavefront. This discussion will be continued in Chapter 4.

Finally, the possibility of optimising the choice of diversity function for a particular application was investigated. It was shown that for good all-round performance the radially symmetric diversity functions, $Z_2^0, Z_4^0, Z_6^0, Z_8^0$ & Z_{10}^0 , are the wisest choice particularly if *a priori* information about the wavefront errors is not available. It was demonstrated that matching the shape of the diversity function to one of the wavefront errors can yield slightly increased sensitivity to that particular aberration in the form of the error signal. However, if only one aberration is present in the wavefront this approach is not the right choice if the user wants an error signal that closely resembles the actual shape of the wavefront in the pupil.

In conclusion, the analysis in this chapter has shown that it is possible to create a GPD wavefront null sensor without placing limiting assumptions on the wavefront under test. The design of this wavefront sensor is very flexible, with the user able to choose between building it to operate in the pupil or image planes, and having a large choice of diversity functions.

3.11 Appendix: Glossary of Commonly Used Functions

Relating to the Diversity/Diffraction Grating:

$f_{\pm}(r)$ The complex filter function, encoded within the diffraction grating.

$\varphi_d(r)$ The diversity phase added by the complex filter function $f_{\pm}(r)$

$F_{\pm}(\xi)$ The Fourier Transform of the complex filter function, this is multiplied by the FT of the wavefront to form the images in the diffraction orders.

$R(\xi)$ The FT of the purely real parts of the complex filter function $\Im\{\text{Re}[f_{\pm}(r)]\}$

$I(\xi)$ The FT of the purely real parts of the complex filter function $\Im\{\text{Im}[f_{\pm}(r)]\}$

Relating to the Input Wavefront:

$\psi(r)$ The input wavefront

$\Psi(\xi)$ The FT of the input wavefront

$H(\xi)$ The Hermitian part of $\Psi(\xi)$, formed by $\Im\{\text{Re}[\psi(r)]\}$

$A(\xi)$ The Anti-Hermitian part of $\Psi(\xi)$, formed by $\Im\{\text{Im}[\psi(r)]\}$

$\varphi(r)$ The phase of the wavefront $\psi(r)$

3.12 References

1. Woods S. C. and G.A. H., *Wave-Front Sensing by Use of a Green's Function Solution to the Intensity Transport Equation*. J.Opt.Soc.Am.A, 2003. **20**(3): p. 508-512.
2. Namias V., *The Fractional Fourier Transform and Its Applications to Quantum Mechanics*. Journal of the Institute of Mathematics, 1980. **25**: p. 241-265.
3. McBride A.C., K.F.H., *On Namia's Fractional Fourier Transform*. Journal of the Institute of Mathematics, 1987. **39**: p. 159-175.
4. Ozaktas H.M., M.D., *Fourier Transforms of Fractional Order and Their Optical Implementation*. Optics Communications, 1993. **101**: p. 163-169.
5. Bernardo L.M., S.D.D., *Fractional Fourier Transforms and Imaging*. J.Opt.Soc.Am.A, 1994. **11**(10): p. 2622-2626.
6. Goodman, J.W., *Introduction to Fourier Optics*. 2nd ed. 1996: McGraw-Hill.
7. Youbin, F., Z. Xuerning, and Y. Jiaxiong, *Research on the Dynamics of a Multidither Adaptive Optical System*. Optical Engineering, 1998. **37**(4): p. 1208-1211.
8. Djidel, S. and A.H. Greenaway. *Nanometric Wavefront Sensing*. in *3rd International Workshop on Adaptive Optics in Industry and Medicine*. 2002: Starline Printing Inc. p. 213-219
9. Zhang, S., H.I. Campbell, and A.H. Greenaway. *Generalised Phase Diversity:Initial Tests*. in *4th International Workshop on Adaptive Optics in Industry and Medicine*. 2003. Munster, Germany: Springer Proceedings in Physics. Vol. 102. p. 187-196
10. Dandliker, R., K. Hug, J. Politch, and E. Zimmermann, *High-Accuracy Distance Measurements with Multiple Wavelength Interferometry*. Optical Engineering, 1995. **34**(8): p. 2407-2412.
11. Zimmermann, E., Y. Salvade, and R. Dandliker, *Stabilized Three-Wavelength Source Calibrated by Electronic Means for High Accuracy Absolute Distance Measurement*. Optics Letters, 1996. **21**(7): p. 531-533.

Chapter 4

Phase Retrieval with Generalised Phase Diversity

Summary of Chapter 4

This chapter briefly reviews a selection of current and historically used methods for phase retrieval from intensity image data. This includes a comparison of the relative merits of iterative and analytic solutions to the data inversion problem. The main body of this chapter concerns the development of a new phase retrieval algorithm, the Small Angle Expansion (SAE), and contains detailed discussion and results from this new method.

4.1 Introduction

The determination of an unknown wavefront given one or more intensity measurements is a well studied problem with importance in many applications, from optical imaging to x-ray diffraction and electron microscopy [1-10]. Retrieval of the phase of an unknown wavefront from intensity images can yield important information about the object which generated the radiation, the quality of the optical system used to form the image and the aberrations the wavefront suffered between emission and detection.

A notable example of the use of Phase Diversity (PD) phase retrieval in the calibration of optical systems can be found in the Hubble Space Telescope (HST). Just weeks after the launch of the HST, in June 1990, scientists discovered that the images received from the telescope were very poor. This led to a flurry of activity in the research community to diagnose the problem, and come up with a suitable solution. Only able to access the optical system remotely, and using the intensity images it captured, it was found that the error could not be removed by realignment of the optics and that it was the shape of the primary mirror which was to blame. Over polishing of the primary mirror by just 2 μm at the edges had resulted in large amounts of spherical aberration in the captured

images [11]. At this time a great deal of work was published on phase retrieval methods that were designed to work with the limited amount of data available [12-15].

A great many phase retrieval algorithms, both iterative and analytic, have been developed by researchers over the years as an attempt to solve this inverse problem for a variety of situations. This is such a wide research topic that this introduction will be limited to just a few of the most well known algorithms, and ones which are related to or helped in the development of the SAE algorithm.

The most widely used iterative algorithms for phase retrieval are compared and reviewed by Fienup [16], these are the error-reduction algorithm, the Gerchberg-Saxton algorithm, the method of steepest descent, conjugate gradient methods and the input-output algorithm. In this paper he concluded that the error reduction algorithm (for single intensity images) and the Gerchberg-Saxton algorithm (for two intensity measurements) will converge, but often at a much slower rate than conjugate gradient or input-output algorithms. There exist a great many algorithms that are variations of these basic iterative structures [17-20] which can prove very effective although in some cases computationally expensive. This is covered in more detail in Section 4.2.1. While developing the SAE algorithm for phase retrieval with Generalised Phase Diversity (GPD) a Gerchberg-Saxton type algorithm was used to obtain early results.

Analytic solutions of the Intensity Transport Equation (ITE) [1, 4, 5] were proven to be fast, accurate reconstruction methods for defocus based PD wavefront sensing (as discussed in Chapter 2). From an early stage it was decided an analytic solution for the GPD phase retrieval algorithm would be pursued, the reasons for this are discussed in Section 4.2. There are many algorithms that rely on the premise that the phase angles to be reconstructed are small, and it is these algorithms that are of particular interest to understand the context of the SAE algorithm, and its relative merits.

Phase recovery techniques are often non-linear processes, but by using the approximation that the phase angles to be reconstructed are small the retrieval process works in a much more linear fashion [21]. In 1983 Ellerbroek and Morrison showed

that linear phase retrieval is possible for wavefront errors of 0 to 0.4 waves rms from focal plane intensity measurements [22], and many analytic algorithms that rely on the Small Angle Approximation (SAA) have been published since then. Wild [21] used the SAA to develop an algorithm that required 3 or more intensity images to reconstruct the wavefront phase and supposed that larger phase angles could be reconstructed if more measurements were made. Gonsalves [6] derived separate formulae for retrieving the odd and even parts of the unknown phase, which required an in focus and one out of focus image. Both of these algorithms have been shown to produce good results for small wavefront deformations. An algorithm that can achieve the accuracies reported by these authors, which uses just two intensity images and calculates the even and odd parts of the phase using the same analytical expression could potentially be a faster solution to the data inversion problem.

In this chapter the requirements for a new algorithm for phase retrieval with the GPD wavefront sensor, based on the SAA, will be outlined. An expression for the retrieved phase will be calculated, for uniform illumination conditions, from the error signal equations introduced in the previous chapter. Several sets of simulation data will be presented to show the accuracy and behaviour of this algorithm under different conditions.

4.2 Requirements and Considerations for a GPD Phase Retrieval Algorithm

In Chapter 3 it was shown that it is possible to construct a null wavefront sensor, using an appropriate filter function which obeys certain symmetry conditions. While it is not always necessary to reconstruct the wavefront nevertheless the problem of inversion of the GPD error signal must be addressed.

An algorithm is required which can take the phase diverse intensity images generated by the GPD sensor and solve for the phase of the test wavefront. This algorithm should in itself be as general as possible, and capable of inverting the data for any (allowable) diversity filter function. In this case the algorithm would then also be an alternative for

phase retrieval amongst the more common practice of defocus based phase diversity systems.

The new phase retrieval method should be designed to be suitable for use in systems with scintillation, and discontinuous wavefronts. There are many reasons for this, which are outlined in Chapters 2 and 3. An algorithm that could meet these requirements would be a very robust and versatile tool in metrology, astronomy and medical applications.

4.2.1 Iterative vs. Analytic

Phase retrieval algorithms tend to fall into one of two categories, iterative or analytic, and some of these are detailed in the introduction to this chapter. An in depth study of the relative merits of the two types of algorithm will not be given here, but in considering the development of a new algorithm it is important to consider which type to pursue. Some issues which should be addressed are the stability of the algorithm, if it will converge, the uniqueness of the solution it computes, and the overall computational ‘expensiveness’ – in run time, power and cost of equipment.

Iterative solutions to the phase problem involve choice of an error metric – a measure of the quality of the computed solution – and will continue to refine the solution until a predetermined minimum in this metric is found. If the solution converges, which is in itself not guaranteed and has to be investigated for each method, there are still many issues to be considered. One such issue is whether the solution found is a global or a local minimum. To ensure it is a global minimum, algorithms of increasing complexity have been produced which, rather than using a simple hill-climbing method, employ a genetic algorithm (GA) approach, or Simulated Annealing (SA) [23-25]. In contrast to other numerical techniques which iteratively refine a single solution, genetic algorithms work in parallel testing many possible solutions at once. GA’s begin with a population of candidate solutions (‘individuals’) and through techniques inspired by evolutionary biology (such as natural selection) evolve towards better solutions [26]. The advantage of this is that it is far more likely that global minima will be found, the disadvantage is

that it often takes a long time to converge and because of the stochastic nature of GA's the global minima found is actually an estimate not an accurate solution [27]. Perhaps the main problem when choosing to use an iterative algorithm is the computational expensiveness of it. The algorithm may converge, and it may be a guaranteed global minimum, but if this requires excessive amounts of computer equipment or if it takes an unreasonable length of time to provide the solution then it will not be suitable.

Analytic solutions can be faster to compute and usually don't require complicated minima search techniques as they essentially go straight to the final answer. The question then becomes whether this is the right answer or not, assumptions may have been made when deriving the analytic solution which then limit its accuracy. Computational expensiveness can also be traded for mathematical complexity in the process of designing the algorithm. The user must consider whether the extra time taken to develop an analytic solution is appropriate for their application, or if a well known iterative solution will be good enough.

There is also the option to develop an algorithm that incorporates both methods, using an analytic approach to get close to the answer and an iterative process to refine the solution. This can help to resolve the local/global minima problems faced by a purely iterative solution and also reduce the overall expensiveness of the process.

All of these issues were considered before work on the new GPD phase retrieval algorithm began. The speed and accuracy of phase reconstruction is paramount for applications which require real-time wavefront sensing. Atmospheric correction, in-line industrial metrology processes, and medical applications like retinal imaging are just a few examples. Therefore, an analytic approach was favoured over an iterative solution, but the possibility of iterative refinement was not ruled out.

4.3 Small Angle Expansion

In Section 4.2 the requirements for a GPD phase retrieval algorithm were outlined and the type of algorithm was chosen. An analytic solution for the phase will be sought,

using the error signal, $d(r)$ (first introduced in Chapter 3), generated by the difference between a pair of intensity images containing equal and opposite applied diversity phase functions.

The starting point for this solution is therefore the expression derived in Chapter 3 (Section 3.6), which describes the error signal created when a diversity filter function of suitable symmetry is applied to an input wavefront in a ‘pupil-plane’ GPD system. For convenience we recall this expression here:

$$\begin{aligned} \frac{d(r)}{2i} = & \left[\int d\xi H(\xi) I(\xi) \exp(-ir\xi) \int d\xi' A^*(\xi') R(\xi') \exp(ir\xi') - \right. \\ & \left. \int d\xi A(\xi) R(\xi) \exp(-ir\xi) \int d\xi' H^*(\xi') I(\xi') \exp(ir\xi') \right] + , \quad (3.16) \\ & \left[\int d\xi A(\xi) I(\xi) \exp(-ir\xi) \int d\xi' H^*(\xi') R(\xi') \exp(ir\xi') - \right. \\ & \left. \int d\xi H(\xi) R(\xi) \exp(-ir\xi) \int d\xi' A^*(\xi') I(\xi') \exp(ir\xi') \right] \end{aligned}$$

where $H(\xi)$ and $A(\xi)$ are the Hermitian and Anti-Hermitian components of the input wavefront $\psi(r)$, formed by the Fourier Transform (FT) of the real and imaginary parts of $\psi(r)$ respectively. $R(\xi)$ and $I(\xi)$ are the real and imaginary parts of $F_{\pm}(\xi)$ (the FT of the diversity filter function), and $d(r)$ is the difference between the intensity images in the ± 1 diffraction orders. Note that while it is assumed here that a diffraction grating is used to apply the diversity, it is not necessary. It is simply convenient that the wavefront is convolved with both the positive and negative diversity functions by passing through the grating, and that the intensity images can be simultaneously recorded on a single CCD plane. The phase retrieval analysis presented in this chapter requires $d(r)$, the method of obtaining the intensity images with equal and opposite diversity used to create $d(r)$ can be chosen to suit the user.

Initially it is assumed that there is no scintillation present, that the input pupil is uniformly illuminated and no light is lost from the system. The effect of this

assumption will be discussed and later extensions of the algorithm to include scintillated wavefronts will be considered in Chapter 5.

4.3.1 Definitions

Consider an input wavefront $\psi(r)$, created by radiation from a distant source, incident on the GPD pupil plane sensor:

$$\psi(r) = P(r) \left[\sqrt{I(r)} \cdot e^{i\varphi(r)} \right] = P(r) \cdot \sqrt{I(r)} \left[\text{Cos}(\varphi(r)) + i \cdot \text{Sin}(\varphi(r)) \right], \quad (4.1)$$

where $\sqrt{I(r)}$ is the amplitude of the input wavefront, $P(r)$ is the pupil function of the system (which may also be called the support of the wavefront), and $\varphi(r)$ is the phase to be retrieved. It can be seen that the wavefront comprises a real $\left(\sqrt{I(r)} \cdot \text{Cos}[\varphi(r)] \right)$ and an imaginary $\left(\sqrt{I(r)} \cdot i \text{Sin}[\varphi(r)] \right)$ part. The FT of the real part will be Hermitian ($H(\xi)$) and the FT of the imaginary part will be Anti-Hermitian ($A(\xi)$). To avoid confusion with the transforms of the filter function terms $I(\xi)$ and $R(\xi)$, let the real part of $\psi(r)$ be given by $h(r)$ and the imaginary part by $ia(r)$. Note that $a(r)$ is a real function. The input wavefront therefore may be described:

$$\psi(r) = h(r) + ia(r) . \quad (4.2)$$

The equation for $F_{\pm}(\xi)$, given in Chapter 3 states:

$$F_{\pm}(\xi) = R(\xi) \pm iI(\xi) . \quad (3.5)$$

The FT of these components is defined here such that $r_e(r)$ is the FT of $R(\xi)$ and $i_m(r)$ is the FT of $I(\xi)$. Finally, we will now consider the error signal when the real

and imaginary parts of the filter function have the same symmetry, and are both even functions of ξ . With these definitions in place the solution of the unknown phase may now begin.

4.3.2 Manipulating the Error Signal Equation

The error signal $d(r)$, under the conditions set out in 4.3.1 is defined to be:

$$\begin{aligned} \frac{d(r)}{2i} = & \left[\int d\xi H(\xi) I(\xi) \exp(-ir\xi) \int d\xi' A^*(\xi') R(\xi') \exp(ir\xi') - \right. \\ & \left. \int d\xi A(\xi) R(\xi) \exp(-ir\xi) \int d\xi' H^*(\xi') I(\xi') \exp(ir\xi') \right] + . \quad (3.16) \\ & \left[\int d\xi A(\xi) I(\xi) \exp(-ir\xi) \int d\xi' H^*(\xi') R(\xi') \exp(ir\xi') - \right. \\ & \left. \int d\xi H(\xi) R(\xi) \exp(-ir\xi) \int d\xi' A^*(\xi') I(\xi') \exp(ir\xi') \right] \end{aligned}$$

In order to find an expression for the unknown phase of the wavefront this equation must be solved, preferably with as few approximations or assumptions as possible. Thus far the only conditions imposed are those outlined in 4.3.1, that $R(\xi)$ and $I(\xi)$ are non-zero and share the same symmetry. Let us now introduce some further assumptions to allow an expression for the phase to be derived. Let us assume that the unknown wavefront phase is close to that of a plane wave (i.e. $\varphi(r) \ll 1$). This assumption, that the phase angles are small, is known as the Small Angle Approximation (SAA). It will form the basis of the algorithm developed to solve for the phase, which is named the Small Angle Expansion (SAE) in reference to this approximation. The amplitude of the wavefront, $\sqrt{I(r)}$ is taken to be the pupil plane intensity and will be represented by $\sqrt{I_0(r)}$. These approximations can be expressed mathematically as:

$$\psi(r) = \sqrt{I_0(r)} \cdot \exp(i\varphi(r)) \approx \sqrt{I_0(r)} [1 + i\varphi(r)] . \quad (4.3)$$

Recalling the earlier definition of $\psi(r)$, given in (4.2), we may write the following:

$$\begin{aligned} h(r) &= \sqrt{I_0(r)} \\ a(r) &= \sqrt{I_0(r)} \cdot \varphi(r) \end{aligned} \quad (4.4)$$

which shows that $a(r)$ is closely related to the phase that we wish to reconstruct, under uniform illumination $h(r)$ is close to being constant and will therefore vary little across the pupil, when scintillation is present this analysis will have to be modified. This will be discussed later, in Chapter 5. An estimate for the pupil plane amplitude, $\sqrt{I_0(r)}$, could be perhaps be obtained experimentally by calculating $\sqrt{(I_+ + I_-)/2}$ where I_+ and I_- are the intensity images in the ± 1 diffraction orders, or perhaps by taking it to be the amplitude in the 0th order, $\sqrt{I_{0^{\text{th}}}}$. It is not immediately obvious from the mathematical description which of these choices, if either, is the correct one to use. This will be discussed in more detail in Chapter 6.

Let us now consider the error signal Equation, (3.16). Let the integrals containing $H(\xi)$ be labelled $t_1 \rightarrow t_4$ such that:

$$\begin{aligned} t_1(r) &= \int d\xi H(\xi) I(\xi) \exp(-ir\xi) \\ t_2(r) &= \int d\xi' H^*(\xi') I(\xi') \exp(ir\xi') \\ t_3(r) &= \int d\xi' H^*(\xi') R(\xi') \exp(ir\xi') \\ t_4(r) &= \int d\xi H(\xi) R(\xi) \exp(-ir\xi) \end{aligned} \quad (4.5)$$

$H(\xi)$ is a compact function in comparison to both $R(\xi)$ and $I(\xi)$, its δ -function like form means that its transform will give a weighted version of $h(r)$. Each integral is in fact a convolution, and can be written in that form. Taking t_1 as an example:

$$t_1(r) = \int dr' h(r') i_m(r' - r) \quad (4.6)$$

$i_m(r)$ is the imaginary part of the filter function and can be carefully chosen to be compact. As discussed in Chapter 3, a compact filter function (in a pupil plane wavefront sensor system) will give good localisation of the error signal to the location of the wavefront deformation which created the signal. When $i_m(r)$ is impulsive compared to $h(r)$, Equation (4.6) can be written:

$$t_1(r) = \tilde{i}_m h(r) ; \tilde{i}_m = \int dr i_m(r) . \quad (4.7)$$

By similar reasoning $t_2 \rightarrow t_4$ can be written:

$$\begin{aligned} t_2(r) &= \tilde{i}_m h^*(r) \\ t_3(r) &= \tilde{r}_e h^*(r) \\ t_4(r) &= \tilde{r}_e h(r) \end{aligned} . \quad (4.8)$$

Using these expressions the error signal, Equation (3.16), may be rewritten:

$$\begin{aligned} \frac{d(r)}{2i} &= \tilde{i}_m h(r) \int d\xi' A^*(\xi') R(\xi') e^{ir\xi'} - \int d\xi A(\xi) R(\xi) e^{-ir\xi} \tilde{i}_m h^*(r) \\ &+ \int d\xi A(\xi) I(\xi) e^{-ir\xi} \cdot \tilde{r}_e h^*(r) - \int d\xi' A^*(\xi') I(\xi') e^{ir\xi'} \cdot \tilde{r}_e h(r) \end{aligned} . \quad (4.9)$$

Since $h(r)$ is real valued, $h(r) = h^*(r)$, we will replace the dummy variable ξ' so that $\xi' = -\xi'$, we also recall the anti-hermitian symmetry of $A(\xi)$ which allows us to substitute $A(\xi) = -A^*(-\xi')$. Therefore (4.9) becomes:

$$\begin{aligned} \frac{d(r)}{2i \cdot h(r)} &= -\tilde{i}_m \left\{ \int d\xi' A(\xi') R(-\xi') e^{-ir\xi'} + \int d\xi A(\xi) R(\xi) e^{-ir\xi} \right\} \\ &+ \tilde{r}_e \left\{ \int d\xi A(\xi) I(\xi) e^{-ir\xi} + \int d\xi' A(\xi') I(-\xi') e^{-ir\xi'} \right\} , \end{aligned} \quad (4.10)$$

Equation (4.10) can be used to derive an expression to solve for either the phase ($\varphi(r)$) or the imaginary part ($a(r)$) of the wavefront. In this section the equation for retrieval

of the phase will be derived, and the benefits of reconstructing the imaginary part will be discussed in later, in this and following chapters. From (4.4) $a(r) \approx \varphi(r)$, and since $A(\xi) = \mathfrak{I}[ia(r)]$, we can make the further substitution that $A(\xi) = \sqrt{I_0(r)} \cdot i\Phi(\xi)$ (where $\Phi(\xi) = \mathfrak{I}[\varphi(r)]$):

$$\frac{d(r)}{2i \cdot h(r) \cdot \sqrt{I_0(r)}} = i \left[\begin{array}{l} -\tilde{i}_m \left\{ \int d\xi' \Phi(\xi') R(-\xi') e^{-ir\xi'} + \int d\xi \Phi(\xi) R(\xi) e^{-ir\xi} \right\} \\ +\tilde{r}_e \left\{ \int d\xi \Phi(\xi) I(\xi) e^{-ir\xi} + \int d\xi' \Phi(\xi') I(-\xi') e^{-ir\xi'} \right\} \end{array} \right]. \quad (4.11)$$

Recalling equation (4.4), $h(r) \approx \sqrt{I_0(r)}$ therefore (4.11) can be rearranged to give:

$$\frac{d(r)}{2I_0(r)} = \int d\xi' \Phi(\xi') [\tilde{i}_m R(-\xi') - \tilde{r}_e I(-\xi')] e^{-ir\xi'} + \int d\xi \Phi(\xi) [\tilde{i}_m R(\xi) - \tilde{r}_e I(\xi)] e^{-ir\xi}. \quad (4.12)$$

Finally, let $R(\xi)$ and $I(\xi)$ be even functions of ξ . Also, using the properties of integrals we can replace ξ and ξ' with ξ :

$$\frac{d(r)}{4I_0(r)} = \int d\xi \Phi(\xi) [\tilde{i}_m R(\xi) - \tilde{r}_e I(\xi)] e^{-ir\xi}, \quad (4.13)$$

Equation (4.13) shows that the error signal is formed by the convolution of the unknown phase with the function $\tilde{i}_m R(\xi) - \tilde{r}_e I(\xi)$. This function shall henceforth be known as the Phase Transfer Function (PTF), $T(\xi)$. The PTF is dependent on the filter function, and will in turn determine the quality of the error signal measured. The importance of the PTF and its properties will be discussed further in Chapter 5.

4.3.3 Checking the Equation by Calculating the Difference Image

Equation (4.13) now presents a second method for obtaining the difference image $d(r)$, the first method being through the use of FT's as described in Chapter 3 (see Figure 3.5

for details). Inevitably, due to the approximations made in the derivation of Equation (4.13), these images will not be identical. By comparing them it is possible to see how much the approximations affect the accuracy of the solution. By performing Matlab simulations, in which all the wavefront information is known, the difference image can be created in 2 different ways using the SAE equations. In the first, $d(r)$ is created using the FT of the input wavefront phase ($\Phi(\xi)$) and the PTF. In the second method the Anti-Hermitian part of the wavefront ($A(\xi)$) and the PTF are used therefore this method makes one less approximation and should, at least in theory, generate a $d(r)$ that is closer to the FT generated $d(r)$. The equations used to generate $d(r)$ for each of these methods are given in (4.14):

$$\begin{aligned} d(r) &= 4 \cdot I_0(r) \cdot \Im\{\Phi(\xi) \cdot PTF(\xi)\} && \text{Method 1} \\ d(r) &= -4 \cdot \Im\{A(\xi) \cdot PTF(\xi)\} && \text{Method 2} \end{aligned} \quad (4.14)$$

$$\text{where } I_0(r) = (I_+(r) + I_-(r))/2$$

The intensity in the pupil ($I_0(r)$) is estimated using the intensity images in the ± 1 diffraction orders ($I_+(r)$ and $I_-(r)$) as shown in (4.14). The difference image was calculated by simulation for SAE methods 1 and 2, and the standard FT method (see Chapter 3 for details). Two sets of results are included here, in Figures 4.1 and 4.2. In both sets of results the difference image calculated for both the SAE methods were very similar and therefore only one 2D plot is included in each figure but all solutions are represented in the 1D section plots. In 4.1 the diversity function used is defocus, and in 4.2 spherical aberration is the diversity phase.

Figures 4.1 and 4.2 clearly show that the difference images calculated using the SAE equations are very close to the difference image calculated by the FT method. This helps to confirm that the arithmetic used thus far is correct, and that the approximations introduced to derive a solution for the phase do not result in the loss of much information. As expected, when using the Anti-Hermitian part of the wavefront instead of the phase the solution is a little better as one less approximation has been made. As these simulations, and the others which were conducted but not included here, show this difference is a small one. It will be discussed later in Chapter 6 that there are other

advantages to reconstructing the imaginary part of the wavefront as opposed to the phase.

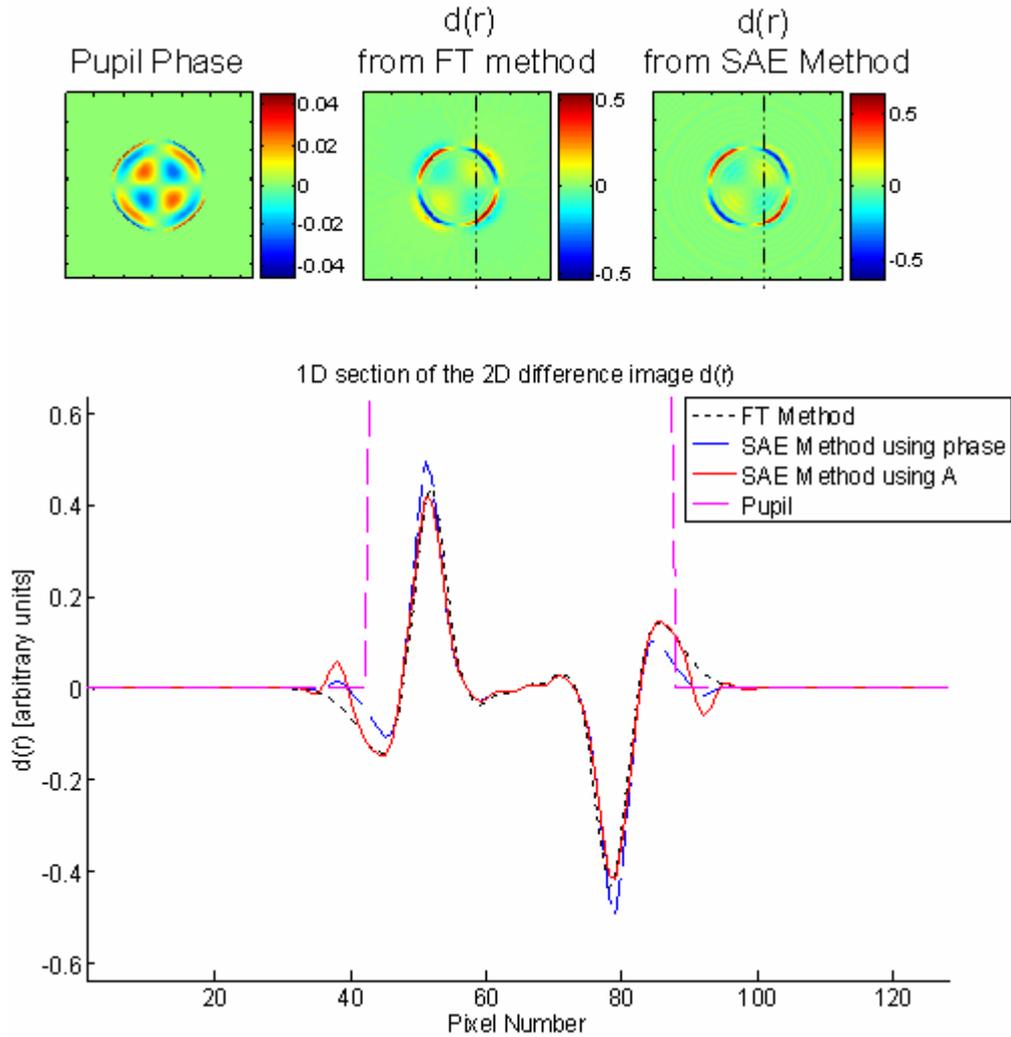


Figure 4.1 The pupil phase of an input wavefront, from which the error signal produced using the FT method and Equation (4.13) have been calculated using a defocus diversity phase [scale of the pupil phase is in waves, and for the difference images the scale is a measure of contrast]. A 1D section of the solutions is also shown.

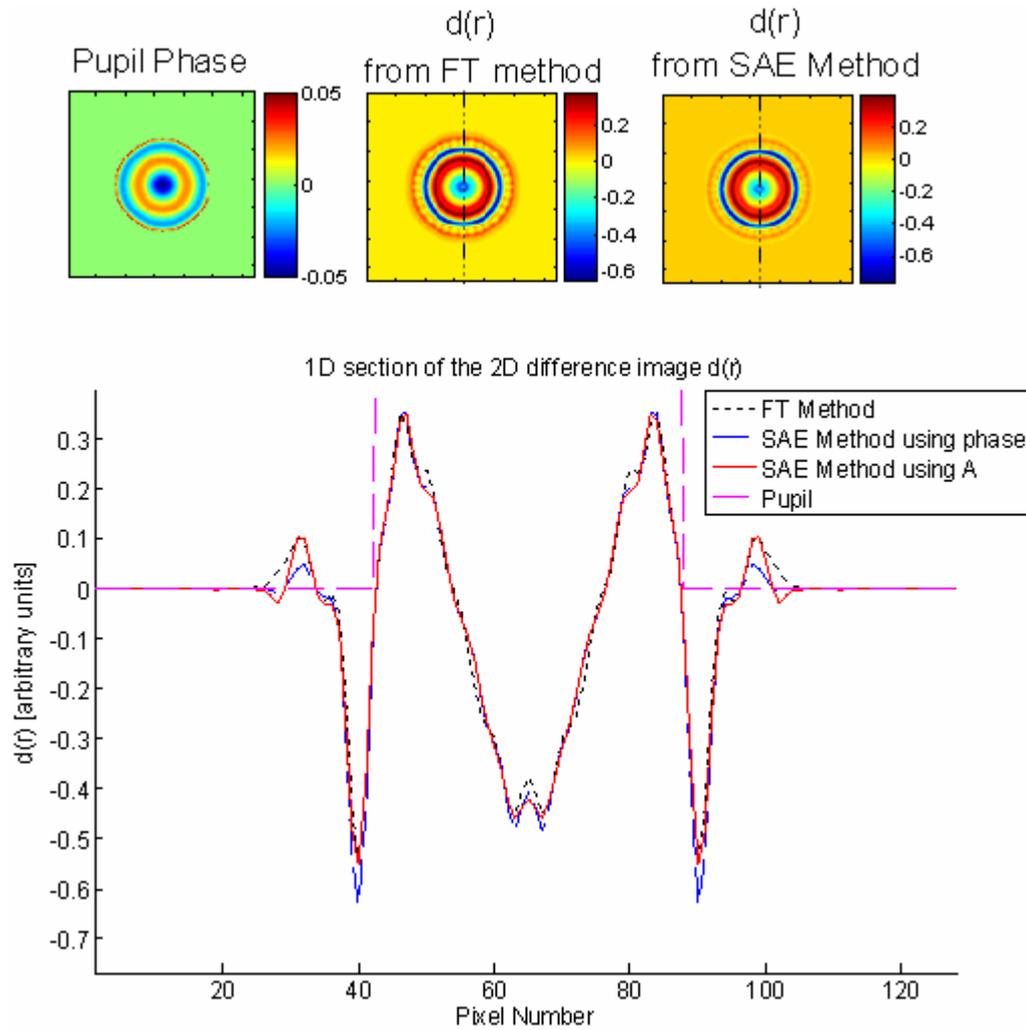


Figure 4.2 The pupil phase of an input wavefront, from which the error signal produced using the FT method and Equation (4.13) have been calculated using a spherical aberration diversity phase [scale of the pupil phase is in waves, and for the difference images the scale is a measure of contrast]. A 1D section of the solutions is also shown.

4.3.4 Solving for the Phase

Starting from Equation (4.13) an expression for the unknown phase may easily be obtained by taking the FT. The resultant expression is given in Equation (4.15):

$$\frac{D_i(\xi)}{T(\xi)} = \Phi(\xi) \quad (4.15)$$

Where $D_i(\xi) = \mathfrak{I} \left[\frac{d(r)}{4I_0(r)} \right]$

Of course, in any real experiment, $d(r)$ will be contaminated by noise and so Equation (4.15) must be altered to take account of this:

$$\frac{D_i(\xi) + N_i(\xi)}{T(\xi)} = \Phi(\xi) \quad (4.16)$$

Where $N_i(\xi) = \mathfrak{I} \left[\frac{n(r)}{4I_0(r)} \right]$

As Equation (4.16) shows, this now presents a classic regularisation problem. Every different (allowable) diversity function will create a different PTF, $T(\xi)$, which will contain some regions of zero, or very close to zero, amplitude. The analysis presented thus far assumes that this wavefront sensor is designed to be a null sensor, which will give no error signal for a plane wave input and is thus insensitive to piston. As such the PTF will be zero at the origin, therefore every allowable function will contain at least one ‘zero crossing’. When $T(\xi) \rightarrow 0$ so too does $D_i(\xi)$, however the noise term $N_i(\xi)$ does not. The noise is amplified due to its division by a number very close to zero. Therefore regularisation is required to minimise the effect of experimental noise on the accuracy of the solution. Initially this regularisation was carried out using a simple threshold for the PTF, where the PTF falls below this threshold the solution for the FT of the phase (as in Equation (4.16)) at that point is rejected. For simulated (and therefore noise free) data this form of ad-hoc regularisation is sufficient to give good results and demonstrate the effectiveness of the SAE algorithm. However, when the algorithm is applied to real experimental data a more rigorous approach is required. This is discussed in Chapter 6 where the SAE is adapted for use with experimental data. In the course of this chapter, where only simulated data are presented, the simpler threshold regularisation is used throughout.

4.3.5 The Nature of the SAE Phase Solution

The SAE solution for the unknown phase, given in Equation (4.16) can be transformed and rewritten to give:

$$\frac{d(r) + n(r)}{4 \cdot I_0(r)} = \varphi(r) \otimes t(r) . \quad (4.17)$$

Further study of Equation (4.17) reveals some interesting features of the SAE phase solution. It shows that the measured difference image $\{[d(r) + n(r)]/4 \cdot I_0(r)\}$ is formed by convolution of the wavefront phase with a transfer function (the PTF) which itself depends entirely on the diversity filter function. A discussion of the form of the PTF and its effect on the GPD wavefront sensor's accuracy will be given in Chapter 5. This section will look at the physical process involved to generate the GPD data and the implications of this on the sensitivity of the GPD sensor.

Figure 4.3 shows schematically how the error signal is formed. The convolution works by mixing the wavefront and diversity phases across an area determined by the width of the blur function. The result of this is the sum of the wavefront phase across the area, weighted by the blur function. Now imagine that the blur function is shifted, and the same operation performed, until the whole wavefront has been covered. The convolution contributions from each area of the wavefront combine by interfering and the result of the interference will depend on the relative phase of the summed contributions. Areas with large phase variations, when weighted by the blur function, will give stronger interference than areas with weaker phase. The grating allows this mixing between the wavefront and the \pm blur functions to occur, the error signal is then generated by taking the difference of the two resulting intensity images.

However, this description of the convolution operation is somewhat inaccurate since the shifting (or shearing) of the blur function to a new position does not occur sequentially as suggested. In fact the convolution of the blur function and wavefront occurs simultaneously across the wavefront. In Wavefront Shearing Interferometry (WSI), as

discussed in Chapter 1, the shearing (between the wavefront and a copy of itself) is performed sequentially and the extent of the shear is controlled by the optical set-up of the interferometer itself. GPD, as this description has shown, works on a similar principle but the variable shear is performed simultaneously and the size of the shear is defined as the width of the blur function. Figure 4.3 is a schematic of this operation.

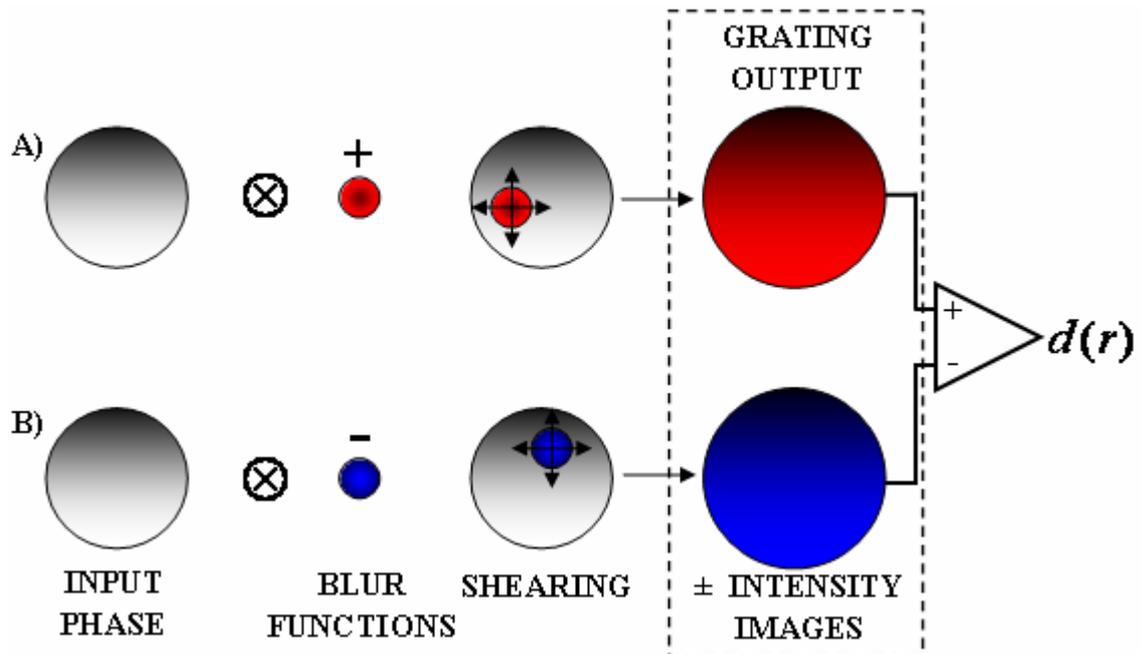


Figure 4.3 Schematic showing the convolution between the +ve (A) and -ve (B) blur functions and the input phase. The intensity images are formed by the variable ‘shearing’ of the blur function across the wavefront and the error signal is created from their difference.

This physical description of the GPD sensor’s operation explains why, when the blur function is compact and peaked at the origin (as all the radially symmetric Zernike diversity functions are), the error signal will be localised to the area of the wavefront error which generated it. Another important conclusion from this description is that it is only the rate of change of the phase across the width of the blur function that will affect the accuracy of the error signal. It is not, as is conventionally the case, the Peak to Valley (PV) aberration of the wavefront which must be limited to small angles. Rather it is the size of the phase step across the blur function which must obey the SAA for the

SAE solution to be accurate. This will be discussed in more detail in Sections 4.4 and 4.5.

4.4 The Relationship between the Error Signal and the Phase Difference

It has been shown that, in the absence of scintillation and within the SAA, there exists a simple analytical expression (4.15) to retrieve the unknown phase from the error signal, $d(r)$. There are still a number of questions which must be addressed. Is this phase solution unique? How large can the PV error on the wavefront be before the SAA no longer holds? What will happen to the solution when this occurs? These questions will be studied in more detail in this, and later, sections.

The relationship between the error signal and the phase difference which generates it has been studied before. Chanan *et al* [28] showed that, when using defocus as the diversity phase, the intensity in each pixel in the difference image varies sinusoidally with the phase difference. Their experiment was constructed to simulate the effects of co-phasing errors in segmented telescopes. They used a split pupil containing a single step in the phase whose height was varied. Chanan *et al* used one in-focus and one out of focus image to create the error signal. Their results showed a clear sinusoidal relationship between the size of the error signal and the phase step which generated it [28]. Simulations were conducted to confirm their results, using the error signal generated by the GPD null sensor. Unlike Chanan *et al* these simulations did not limit the diversity phase to defocus, but instead used a range of allowable GPD null sensor filter functions. In each case the same sinusoidal response was observed, and this is shown schematically in Figure 4.4.

This result has important implications for the SAE solution of the unknown phase. Figure 4.4 shows that the signal is at a maximum at $\pi/2$, which corresponds to a phase step in waves of $\lambda/4$. This is expected, as the signal has been shown to vary sinusoidally and we know that $d(r)=0$ for a phase step of $n\pi$. The relationship between the error signal and the phase is only monotonic for phase steps in the region $\pm \pi/2$. As the phase step approaches the maximum at $\pi/2$, the relationship becomes

increasingly non-linear. When the phase step exceeds this limit the value of the error signal is ambiguous; the same signal will be generated for 2 different phase steps. This is demonstrated by the highlighted pairs of points in Figure 4.4. The result of this is that the SAE algorithm will return the same solution for both phase profiles. This was confirmed by simulation and the results are shown in Figure 4.5. Further implications of this will be discussed in Section 4.6.3.

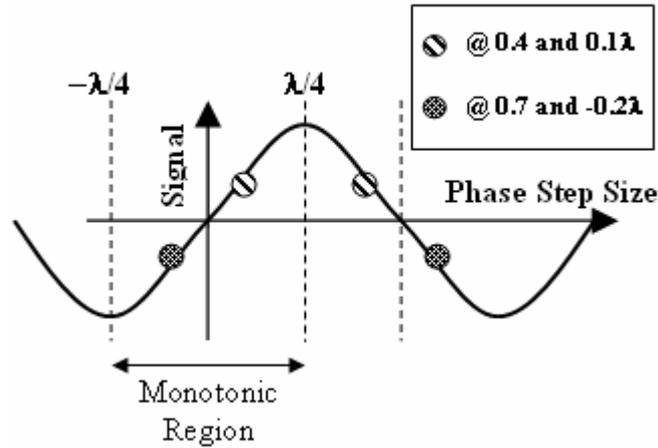


Figure 4.4 Schematic of the relationship between the error signal $d(r)$ and the phase step which generated it.

Matlab simulations were conducted using plane wave input wavefronts with piston phase errors of varying height to test whether the SAE would return the same solution for two different phase step values. In Figure 4.5(a) cross-sections of the retrieved phase profiles for a pair of input wavefronts with PV piston errors of -0.2λ and 0.7λ are compared. Comparison of the retrieved phase profiles for both cases shows that they are indeed the same and that the SAE algorithm has calculated a solution for both input wavefronts in the monotonic region. The same behaviour is shown in Figure 4.5(b), where the SAE algorithm results for wavefronts with piston errors of 0.1λ and 0.4λ are plotted. Therefore to ensure that the error signal represents the correct phase value, care must be taken to limit the size of the phase step to the $\pm \pi/2$ region.

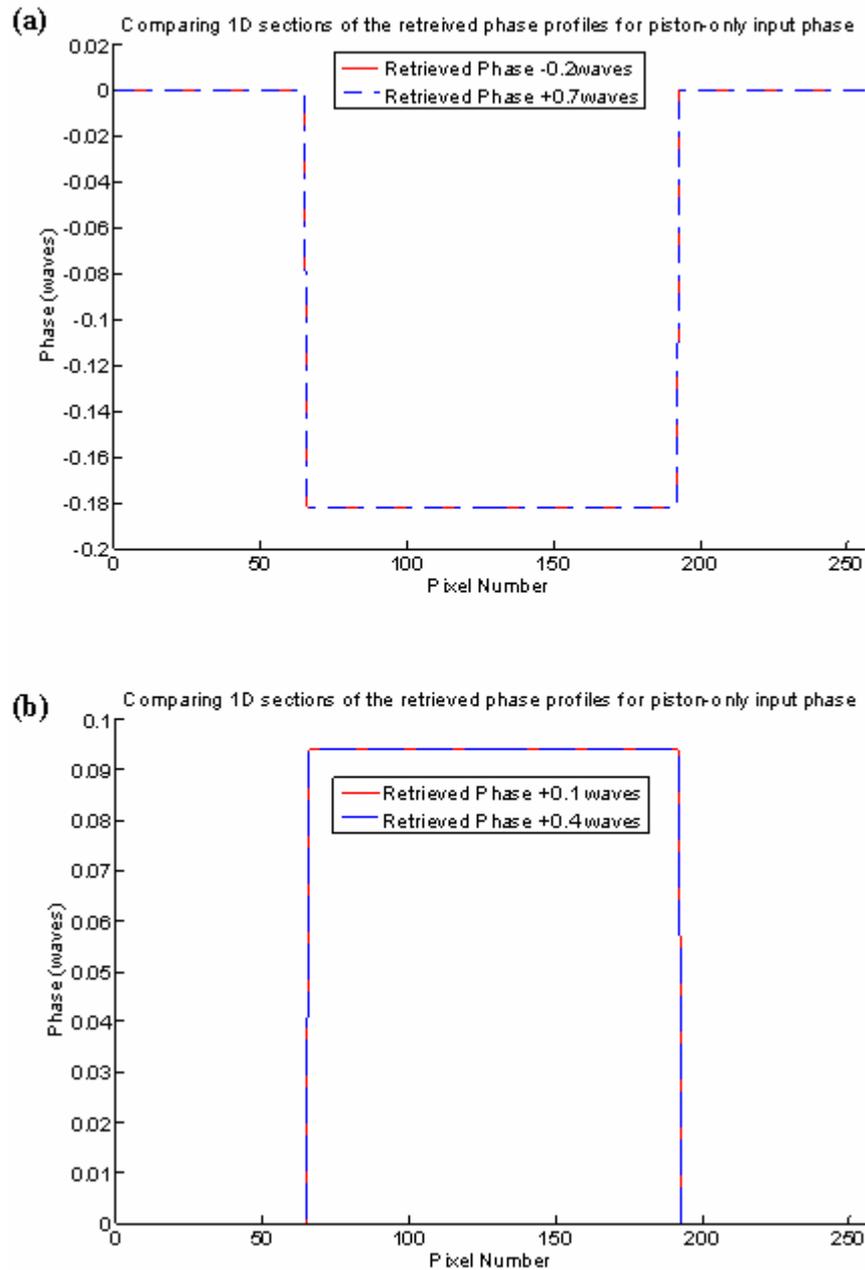


Figure 4.5 1D cross section of the original and retrieved phase profiles of a wavefront with (a) -0.2λ and 0.7λ of piston error and (b) 0.1λ and 0.4λ of piston error

Figure 4.6 is an altered version of Figure 4.4 to summarise these findings. The sine wave response of the error signal to the size of the phase step is shown, and split into sections. In the central black section the SAE solution will be monotonic, in the blue sections the data has become multi-valued. When the size of the phase step falls within the blue regions the algorithm will retrieve a phase from a corresponding point in the monotonic region, as discussed (and shown in Figure 4.5). If the blue region is ‘folded

back' to overlay the black portion of the curve, then this will indicate where on in the black region the SAE will return a phase from. If the phase step is very large, and occurs in the red regions shown in Figure 4.6, then again the SAE will attempt to reconstruct a phase within the monotonic region but with such poor accuracy that no useful result can be obtained ('fold back' the red curve in the same way to see which phase will be reconstructed). Therefore phase steps of this size should always be avoided, phases within the blue regions will be reconstructed but the result will be multi-valued and therefore ambiguous, and phase steps in the black monotonic region are of the optimum size for accurate reconstruction:

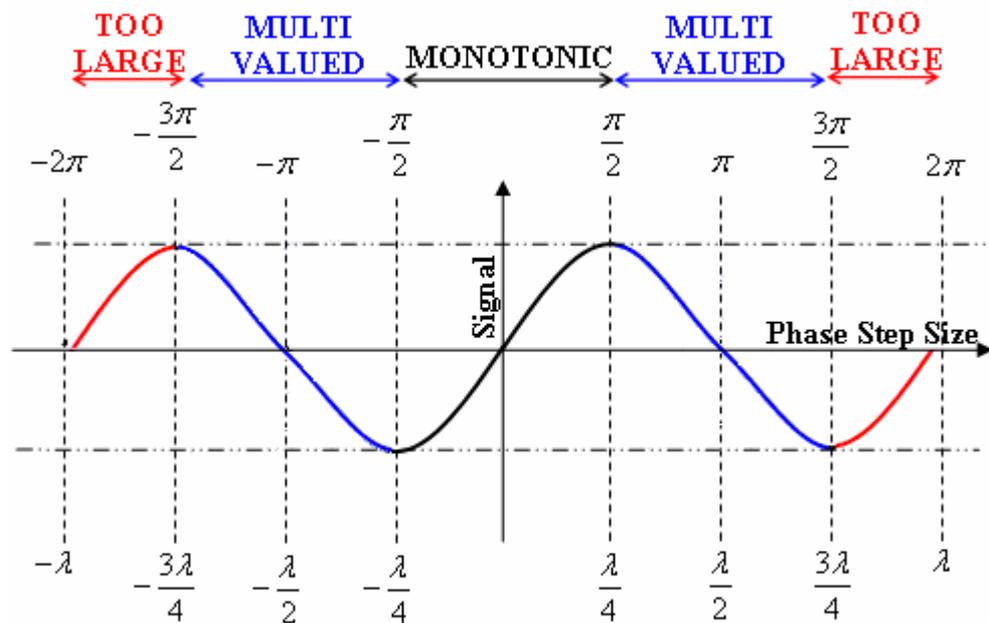


Figure 4.6 Schematic of the relationship between the error signal $d(r)$ and the phase step which generated it, demonstrating which regions will return a single-valued solution (black), an ambiguous result which could have been generated by several different input phase steps (blue), or a result too poor to be useful (red).

4.5 PV Wavefront Error and the SAE

In Chapter 3 (Section 3.8), and then in more detail in Section 4.3.5 it was shown that the error signal is generated by the convolution of the input wavefront with a blur function. If the diversity function used is defined as $e^{i\cdot\phi_d}$ then the blur function is given

by $\mathfrak{I}\{e^{i\phi_d}\}$. The physical width of this blur function determines the area over which the wavefront and diversity signals are mixed by the convolution process (see Figure 4.3 for details). In 3.8.2 the effect of the blur function width on the overall sensitivity of the wavefront sensor was discussed. Its effect was first discussed in the context of the SAE, in Section 4.3.5.

Following the results presented in Section 4.4 it is now possible to quantify what exactly is meant by a ‘small angle’. The size of the phase step as referred to in 4.5.2, is in fact the local change in the phase and its slope across the width of the blur function. If the PV phase difference across the width of the blur function is less than $\pm \pi / 2$ ($= \pm \lambda / 4$) then the error signal generated will be single-valued.

The overall PV distortion of the wavefront, provided this condition is obeyed, does not affect the overall performance of this phase retrieval algorithm. This is an important step forward as previous phase retrieval algorithms have required that the PV of the wavefront itself be limited to small angles for accurate reconstruction. This advantage can be maximised by increasing the sampling (increase the number of pixels/samples across the input wavefront) or reducing the width of the blur function. In this way it would be possible to push this algorithm to deal with extremely aberrated input wavefronts. Gonsalves [6] presented an analytic phase retrieval solution which works only when the phase angle is small. The Green’s function solution (discussed in Chapter 2), is also formulated for small angle use, although in practice it proved capable of handling larger angles. When this was the case the solution is low pass filtered, i.e. the algorithm reconstructed the low-order modes very well but had trouble with higher order modes [1].

One difficulty with the SAE approach is seen when the convolution reaches the edge of the wavefront. Outwith the area of the pupil, where there is no signal or a signal so small it is indistinguishable from the noise, any computed phase value can have no physical meaning. However, as the convolution traverses the boundary, even if the PV wavefront error is relatively small, there may be a large enough discontinuity in the phase to violate the SAA. One possible solution, which has been found to be quite

effective, is to use a soft-edged pupil function and thus decrease the severity of the discontinuity. The use of this type of pupil function is the subject of Section 4.6.2.

The potential to cope with high PV phase wavefronts implies that a GPD wavefront sensor, when coupled with an analytic solution for the phase or an analytic algorithm with iterative refinement of the solution, will be a powerful device whose versatility will make it suitable for a much wider range of applications than previous PD wavefront sensors.

4.6 Simulation Results

Matlab simulations were conducted to assess the performance of the SAE solution (given by (4.13)) for a range of input wavefronts. The effect of the pupil function shape is considered to show the difference in the reconstruction accuracy when using hard-edged or soft-edged pupils. At this point only circular or circularly-symmetric pupil functions are considered. There is of course the possibility that this method could be reformulated for use with more exotically shaped apertures in future work. Note that in all cases the retrieved phase profile has been windowed by multiplication with the hard edged pupil function which will be shown in Figure 4.7. This was done to limit comparison of the original and reconstructed phase profiles to the extent of the original input phase profile. The SAE, where it is unable to reconstruct the input phase accurately across this region, will reconstruct additional ‘wings’ to the retrieved phase profile in order to give a phase profile with zero mean. Therefore the reconstructed phase in regions which are known to contain no signal can have no physical meaning and have been excluded using this windowing technique. It should be noted that the SAE reconstructions, like the original phase profiles, are also of zero mean.

In each case a Strehl-like Error Metric (which will be abbreviated to EM) was calculated to assess the performance of the SAE in terms of the overall fit of the retrieved phase solution to the original input phase profile. Strehl ratio is a measure of overall quality of an image in comparison to the diffraction-limited case. Strictly speaking it is the ratio of the central peak intensities of both the aberrated and perfect image and has a

value between 0 and 1, where 1 is a perfect image. In these simulations the EM will also have a value between 0 and 1 where an EM of 1 would be a perfect reconstruction of the input phase profile. The EM is calculated from the variance (σ^2) defined using Equation (4.18):

$$\sigma^2 = \frac{\sum [\varphi_{\text{original}} - \varphi_{\text{retrieved}}]^2}{N^2}, \quad (4.18)$$

EM = Strehl-like Error Metric $\approx e^{(2\pi)^2 \cdot \sigma^2}$

where $\varphi_{\text{original}}$ and $\varphi_{\text{retrieved}}$ are the phase profiles used, measured in waves and N is the number of pixels in each. Note that the integral in (4.18) is taken over the whole of the phase profiles, not just over the area of the pupil. The calculation of the EM includes a multiplying factor of $(2\pi)^2$ in the exponent to convert the variance (σ^2) from waves to radians. For each simulation included in this section the EM (calculated using (4.18)) will be given to assess the performance of the SAE solution in each situation.

4.6.1 Hard-Edged Pupil

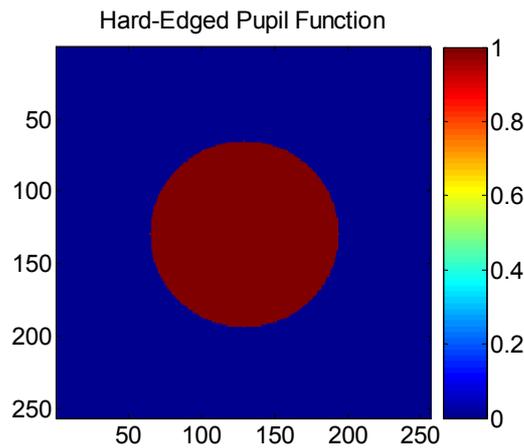


Figure 4.7: The Hard-Edged pupil function used in all the Hard-Edge simulations.

In the first series of Matlab simulations, a hard-edged ‘top-hat’ input pupil function (as shown in Figure 4.7) was used. This is the shape of pupil function that would be created by simply placing a limiting aperture in the pupil plane of the system, or having

an entrance pupil defined by the diameter of a perfect pupil lens. It is also the pupil function shape which would be expected from a telescope with a continuous primary mirror (i.e. non-segmented and with no obscuration from a secondary mirror).

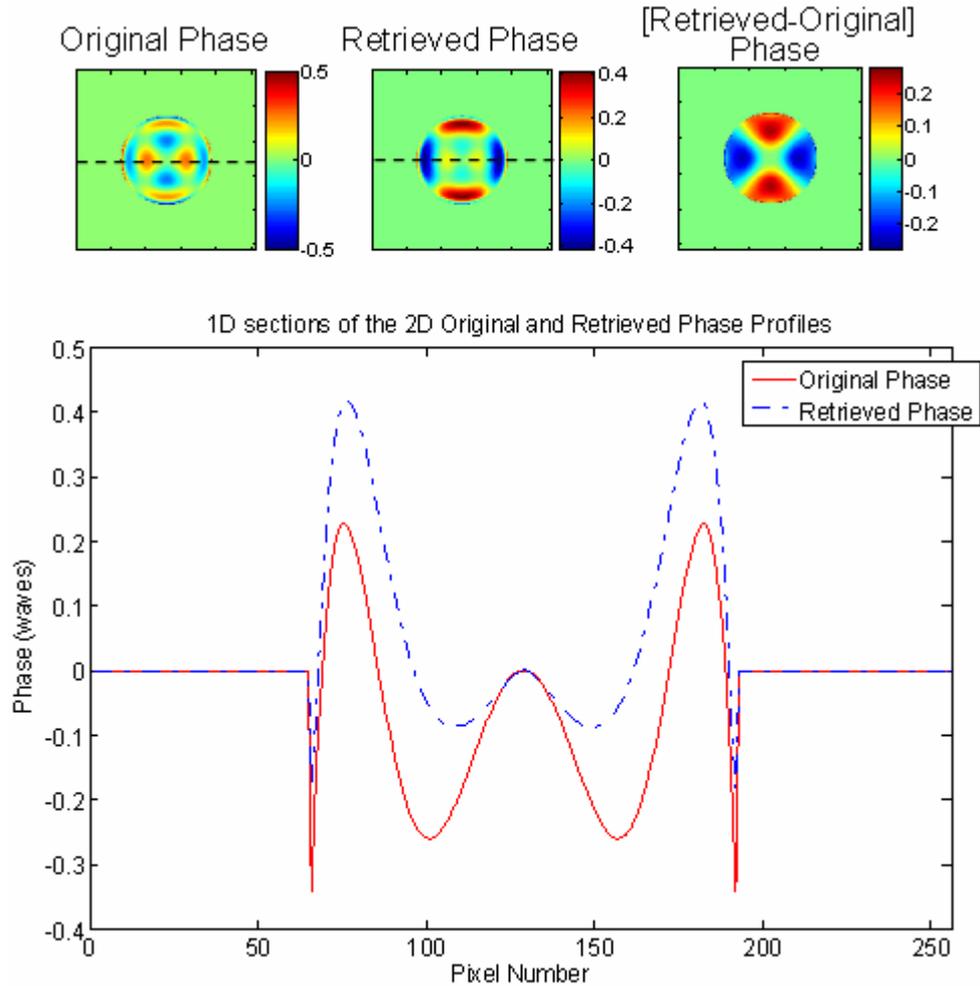


Figure 4.8 Results showing the original and retrieved phase using the SAE algorithm, with a hard-edged pupil function, and Defocus as the diversity phase [all scales are in waves]. The EM for this reconstruction is 0.844.

Figures 4.8 and 4.9 are the results obtained for two different test wavefronts, using defocus as the diversity function. For the first example (Figure 4.8) the input phase profile is a pure Zernike term (Z_6^2). In the second example (Figure 4.9) a phase profile was created using a mixture of 3 different Zernike terms (Z_6^2, Z_2^0, Z_3^3) with random weights. In both cases the PV aberration of the wavefront was kept low to minimise the

size of the phase step at the boundary of the hard-edge pupil. The PV aberrations of the wavefront phase in Figure 4.8 and 4.9 are 1λ and 0.252λ respectively. Each figure contains a plot of the original and retrieved phase profiles and the difference between them, with their scale in waves. Below the phase profiles in each figure is a plot of a 1D slice through the original and retrieved phases to compare the fit of the solution to the original data. The black dashed lines in the phase profiles indicate the position that the 1D slices were taken. The EM's calculated using Equation (4.18) for these reconstructions are 0.844 and 0.994 respectively.

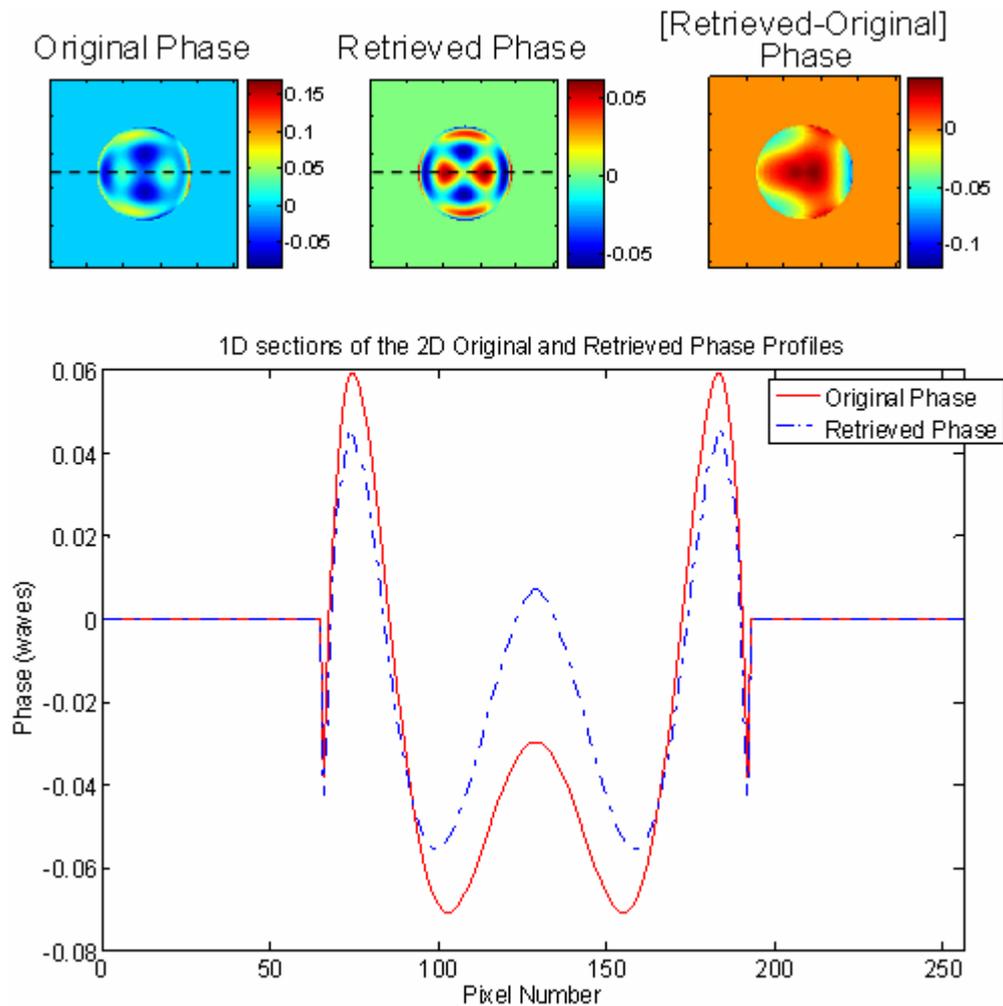


Figure 4.9 Results showing the original and retrieved phase using the SAE algorithm, with a hard-edged pupil function, and Defocus as the diversity phase [all scales are in waves]. The EM for this reconstruction is 0.994.

These results are representative of a much larger range of input wavefronts which were tested. Figure 4.8 and 4.9 show that, in both cases, the SAE algorithm has retrieved the correct overall shape, but not the correct amplitude, of the input wavefront. This could be indicative of a boundary value problem. In order to explore this possibility further a series of simulations were conducted using a soft-edged, Gaussian profile, pupil function.

4.6.2 Soft-Edged Pupil

A soft-edged pupil was considered next to investigate the possibility that the inaccuracies in the SAE were caused, at least in part, by boundary value problems. One possible source of this problem is that when the blur function overlaps the hard edge of the wavefront, convolving the input wavefront with regions which contain no signal, then the slope in the signal at these points could violate the SAA even for relatively small PV aberration (as discussed in Section 4.3.5)

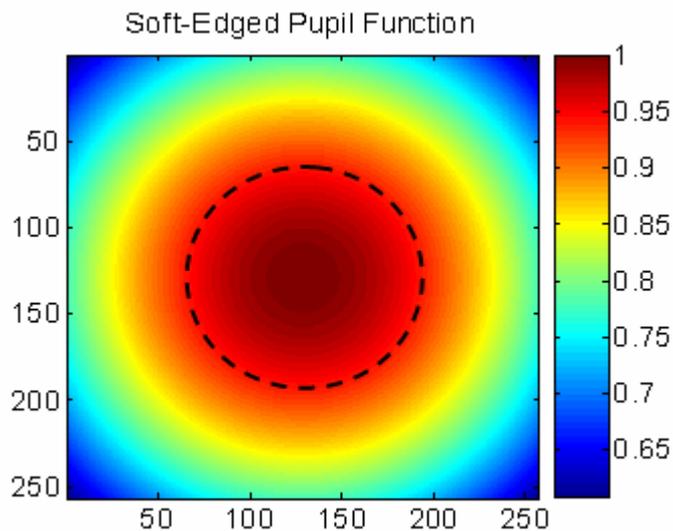


Figure 4.10 The Soft-Edged pupil function used in all the Soft-Edge simulations. The black dashed circle shows the extent of the original hard-edged pupil function.

The black dashed circle shown in Figure 4.10 shows the extent of the original, hard-edged, pupil function. The input phase profiles for wavefront phase ($\varphi(r)$) and diversity phase (φ_d) were defined over this limited pupil area and were set to zero everywhere outwith this region. In addition to this the diversity function ($e^{\pm i\varphi_d}$) is multiplied by the hard edged pupil function (i.e. set to zero outwith the area of the pupil) to simulate the wavefront encountering a finite sized grating and lens combination. Where the input phase $\varphi(r)$ falls to zero (outside the black circle) the value of the wavefront ($e^{i\varphi(r)}$) will be unity. Multiplication of the wavefront by the soft-edge pupil function will soften the illumination across the whole field, not just the area of the pupil. Therefore, in simulation there is information outside the area defined by the hard-edged pupil which comes purely from the soft-edge pupil illumination function and not from the simulated wavefront data. In an experimental situation the softened illumination would still be limited by the finite size of the entrance pupil or optical elements used. For this reason the solutions for the phase presented in this section have been windowed by multiplication with the hard-edged pupil function to present data which is as realistic as possible.

For the soft-edged illumination a Gaussian pupil shape was chosen (shown in 4.10), this is a function that will drop gradually to zero and the ‘softness’ of the pupil can be altered by choosing how much of the Gaussian overlaps the input wavefront. This is shown schematically in Figure 4.11.

Conceptually a pupil function of this kind could be created using a Gaussian profile probe beam to illuminate a phase target of finite size mounted onto a plane sheet of glass. In practical metrology a Gaussian pupil function may be easily obtained by illuminating the surface under test with the output from a single mode optical fibre. Not only does this deliver a smooth illumination function, it is also very practical, fibre delivery of the illumination is a lightweight and robust solution and would allow the wavefront sensor to be mounted onto the end of a robotic arm, or a swing-arm profilometer for example.

In Figure 4.11 the Gaussian pupil function is shown by a dashed line, and the filled oval represents the input wavefront. In 4.11(a) only the central portion of the Gaussian pupil is seen by the wavefront. In 4.11(b) the sloping edges of the Gaussian pupil overlap the wavefront and this creates an overall ‘softer’ edge than in 4.11(a).

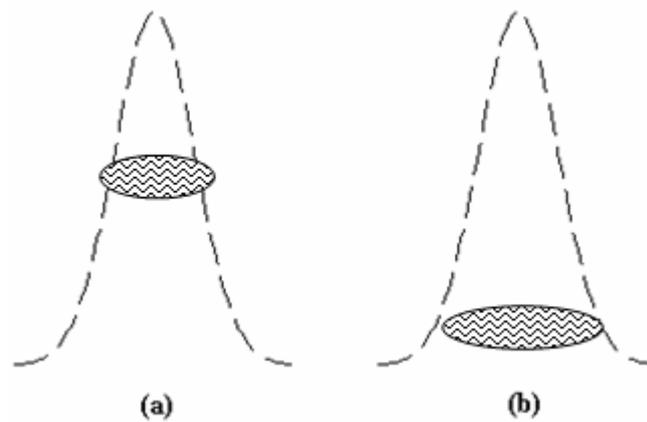


Figure 4.11 Schematic showing the differing degrees of pupil function ‘softness’. The greater overlap in (b) gives a much softer edge than the one shown in (a).

Caution must be exercised when using a soft-edged pupil function, to ensure that the illumination has a gentle slope across the pupil but that the intensity values at the boundary are not significantly changed since this would only create additional inaccuracies. This is demonstrated in Figures 4.15 and 4.16 which show the same wavefront tested with 2 pupil functions of different softness.

For direct comparison with the hard-edged simulations, the same input wavefronts and diversity functions were used in the soft-edged pupil cases. The pupil function in both these simulations was the same, and is shown in Figure 4.10 and also in 4.15(a), where the effect of edge softness is discussed.

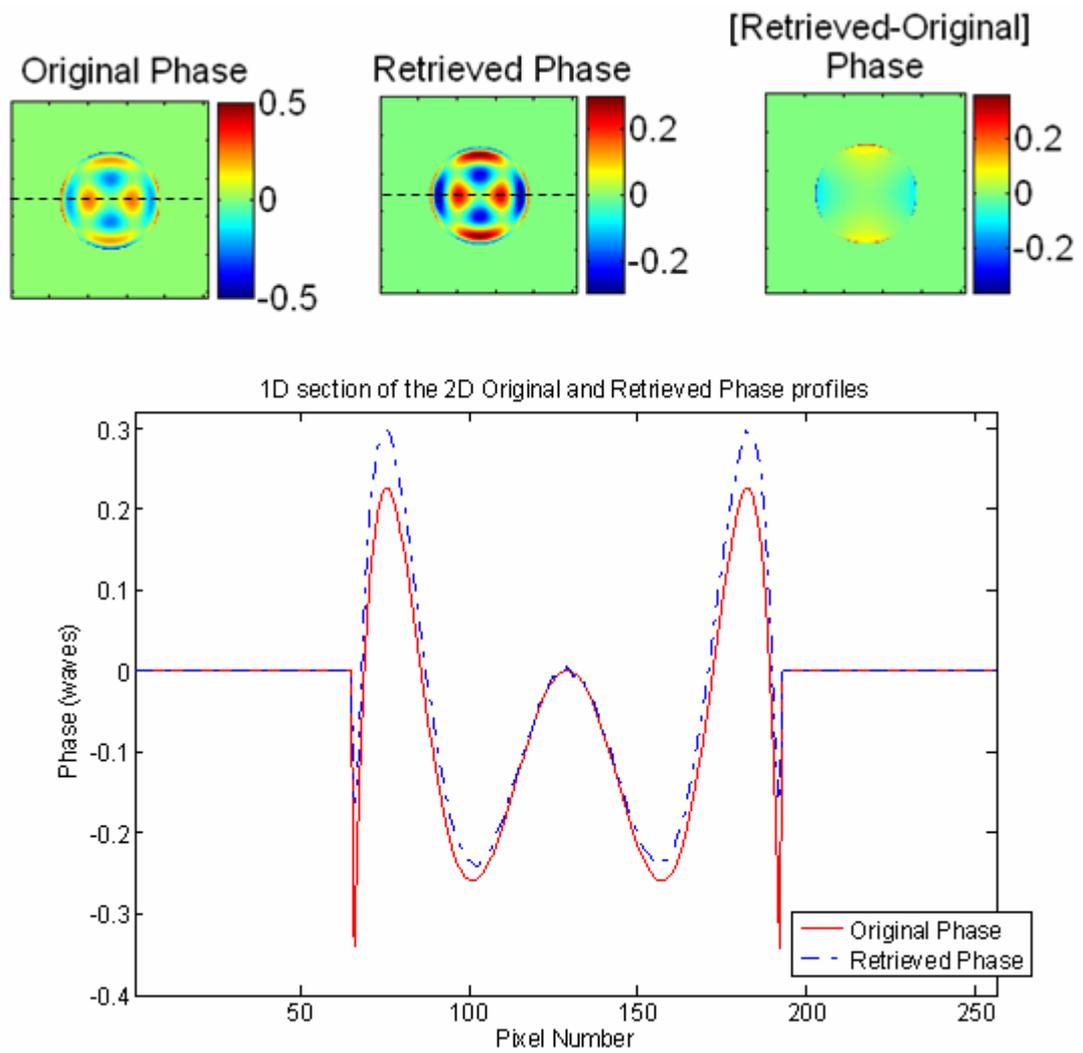


Figure 4.12 Results showing the original and retrieved phase using the SAE algorithm, with a soft-edged pupil function, and Defocus as the diversity phase [all scales are in waves]. The EM calculated for this reconstruction is 0.983.

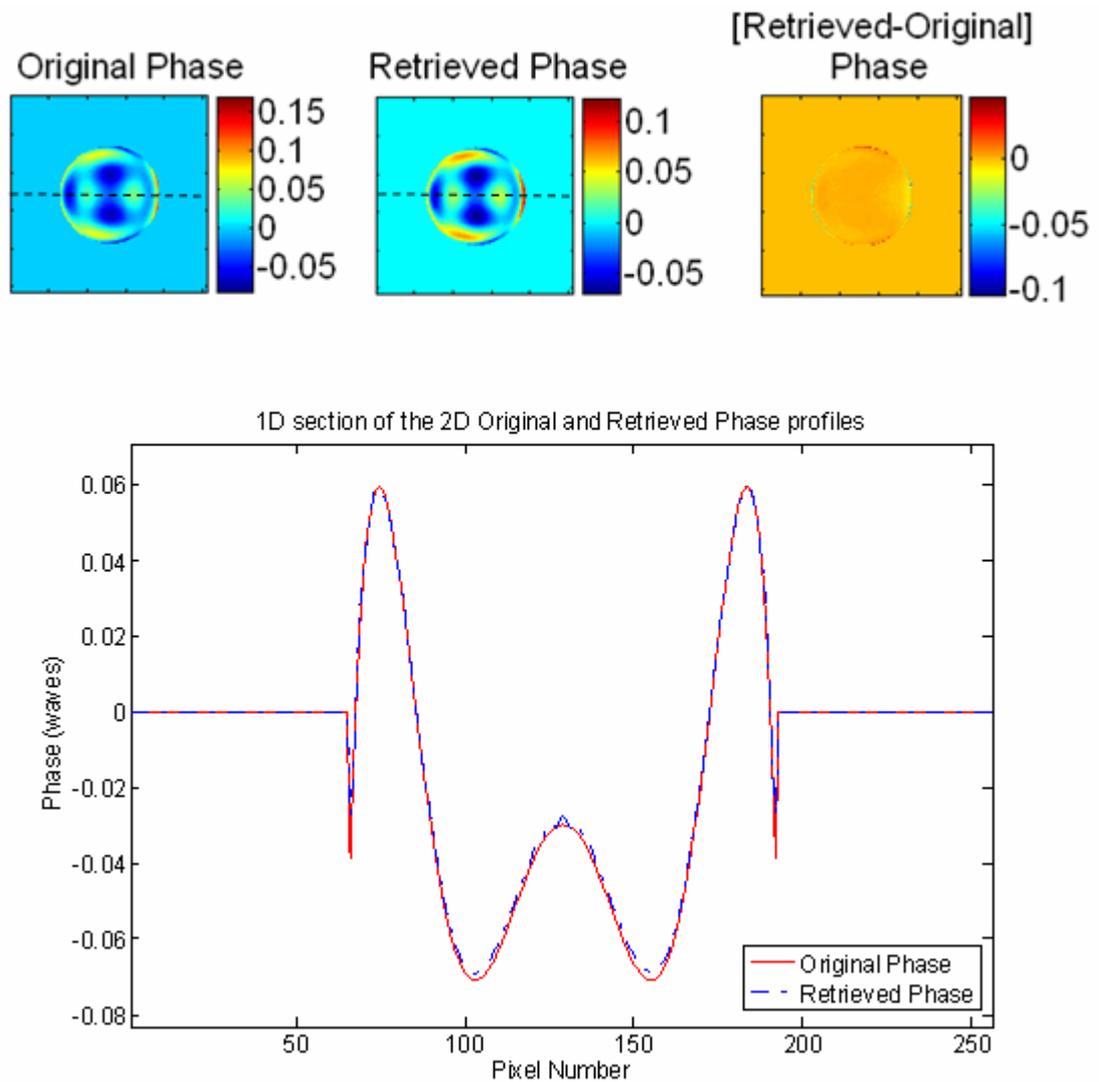


Figure 4.13 Results showing the original and retrieved phase using the SAE algorithm, with a soft-edged pupil function, and Defocus as the diversity phase [all scales are in waves]. The EM calculated for this reconstruction is 0.9998.

The results shown in Figures 4.12 and 4.13 clearly show that the use of a soft-edged pupil function has improved the fit of the solution to the original phase function considerably. Comparing Figure 4.8 and 4.12, it has been shown that the EM was improved from 0.844 to 0.983 when using a soft-edged pupil function. Similarly a comparison of Figures 4.9 and 4.13 shows an increase in the EM from 0.994 to 0.9998. This again would suggest that part of the problem, shown in the hard-edge pupil simulations, was a boundary value issue. In fact for all the cases tested the soft-edged

pupil resulted in an overall better fit to the data. This fit is however subject to several conditions which will be discussed in the following section.

4.6.3 Factors Affecting the Fit of the Solution for the Hard and Soft-Edge Simulations.

In Sections 4.6.1 and 4.6.2 results were presented which demonstrated the use of the SAE algorithm, with both hard and soft-edge pupil functions. The quality of the computed solutions was assessed using the EM defined in Equation (4.18), and it was shown that for identical input parameters the soft-edge pupil function resulted in a better fit. However, this fit is dependent on several factors which will be discussed here.

Firstly, the fit relies highly on the choice of a suitable ‘threshold’ value, which is related to the regularisation issues discussed in Section 4.3.4. A threshold constant is defined and where the PTF falls below this chosen constant the phase solution from those points is not calculated. This, as discussed previously, is a very simplistic approach to regularisation which is not suitable for experimental data but which works sufficiently well on simulated data. The threshold constant provides the means to avoid a ‘divide by zero’ problem when using Equation (4.15) to solve for the unknown phase. This value, for both the hard and soft edged pupil simulations, was chosen to be 10^{-5} . Therefore, use of this threshold should not allow any significant data to be lost but only to allow the simulation to run smoothly. Figure 4.14 is exactly the same simulation as the hard-edge example shown in Figure 4.8 (see 4.8 for a plot of the pupil phase) except in this case the threshold has been increased to 10^{-3} , by doing so the EM increased from 0.844 to 0.988.

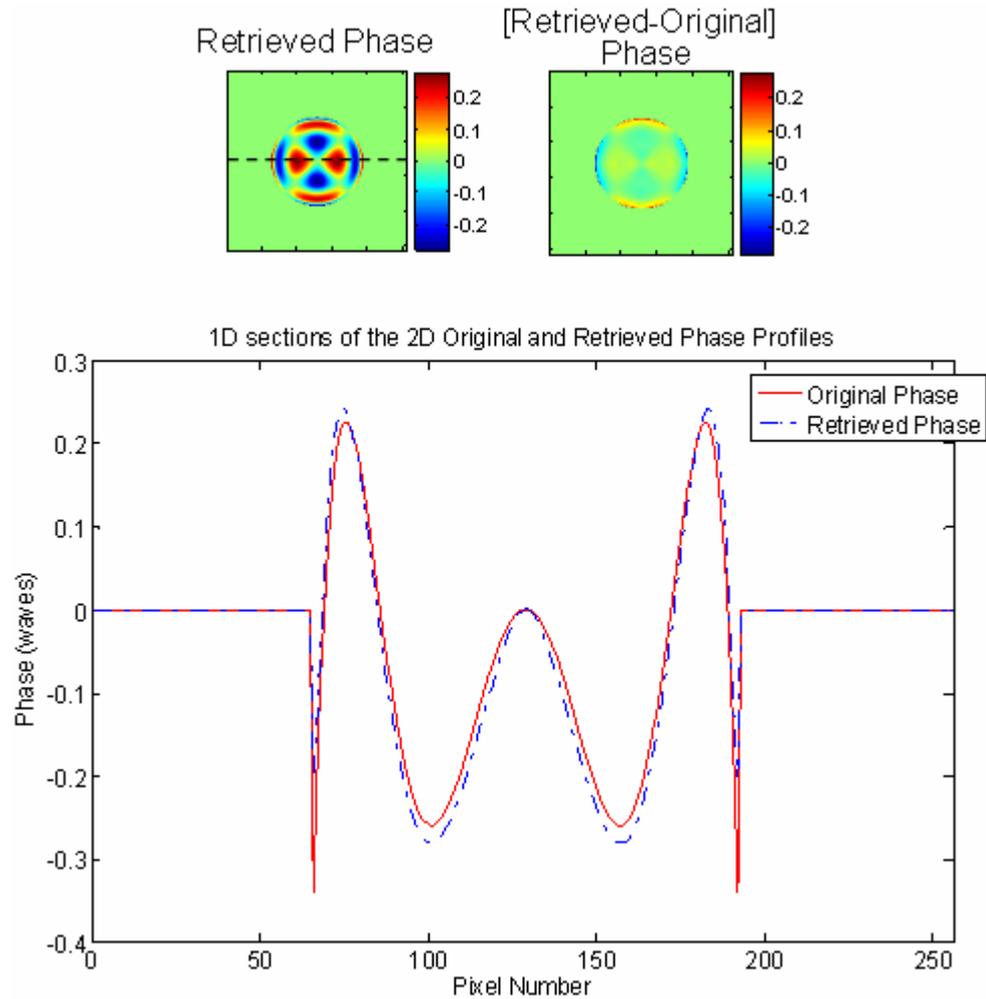


Figure 4.14 Results showing the phase retrieved by the SAE algorithm, with a hard-edged pupil function, and defocus as the diversity phase [all scales are in waves]. This example is the same as the one shown in Figure 4.8, but with a different threshold constant. The EM calculated for this reconstruction is 0.988, an increase in quality solely due to an increase in the threshold constant to 10^{-3} (from 10^{-5}).

Increasing the threshold constant means that more of the solution (about the point that the PTF falls to zero) has been rejected. It would be reasonable to expect that this should therefore decrease the accuracy of the overall solution since more information is lost. In Chapter 5, where the properties of the PTF are studied in more detail, it will be shown that for a defocus diversity function the PTF oscillates about zero at the edge of the pupil. Therefore, the frequencies in the original data which are lost by this crude regularisation procedure are ones which (in this case) the SAE was unable to reconstruct accurately (e.g. the high frequency features at the boundary) which helps explain the

improved fit to the data. In the soft-edged version (results shown in Figure 4.12), the SAE is better able to deal with the boundary and the overall fit is very good to start with (EM of 0.983). In this case increasing the threshold constant to 10^{-3} caused a decrease in image quality to give an EM of 0.974, since useful edge information was lost. These results demonstrate how sensitive the SAE solution is to the threshold constant, which is why the results presented in Figures 4.8 and 4.12 and in 4.9 and 4.13 were deliberately calculated using exactly the same input parameters (except for the shape of the pupil function) to provide a meaningful comparison. The regularisation issue demonstrated here will require a more rigorous approach which will be dealt with in Chapter 6.

The second factor which affects the quality of the SAE is the PV distortion of the input wavefront. In all of the simulations presented in Sections 4.6.1-3 the wavefront PV distortions (both overall, and locally across the width of the blur function) were chosen to ensure that the SAA was not violated. Simulations reconstructing wavefronts with very high local slopes are shown in Section 4.6.5, where iterative refinement is used to improve the solution. Finally, the fit is also dependent on the softness of the pupil function chosen.

Figures 4.15 and 4.16 demonstrate the effect of pupil softness on the retrieved phase. In each figure a 1D scan through the input phase and pupil function is shown, to illustrate the amount of overlap in each case. In the upper plot of each figure the dotted arrows show the position of the edge of input wavefront and where this intercepts the pupil function. In the first example only the upper central portion (11%) of the Gaussian pupil function is combined with the wavefront intensity. In the second example (see Figure 4.16) more of the Gaussian is included within the bounds of the input wavefront (43%), enough that the values of the intensity at its outer edges are altered.

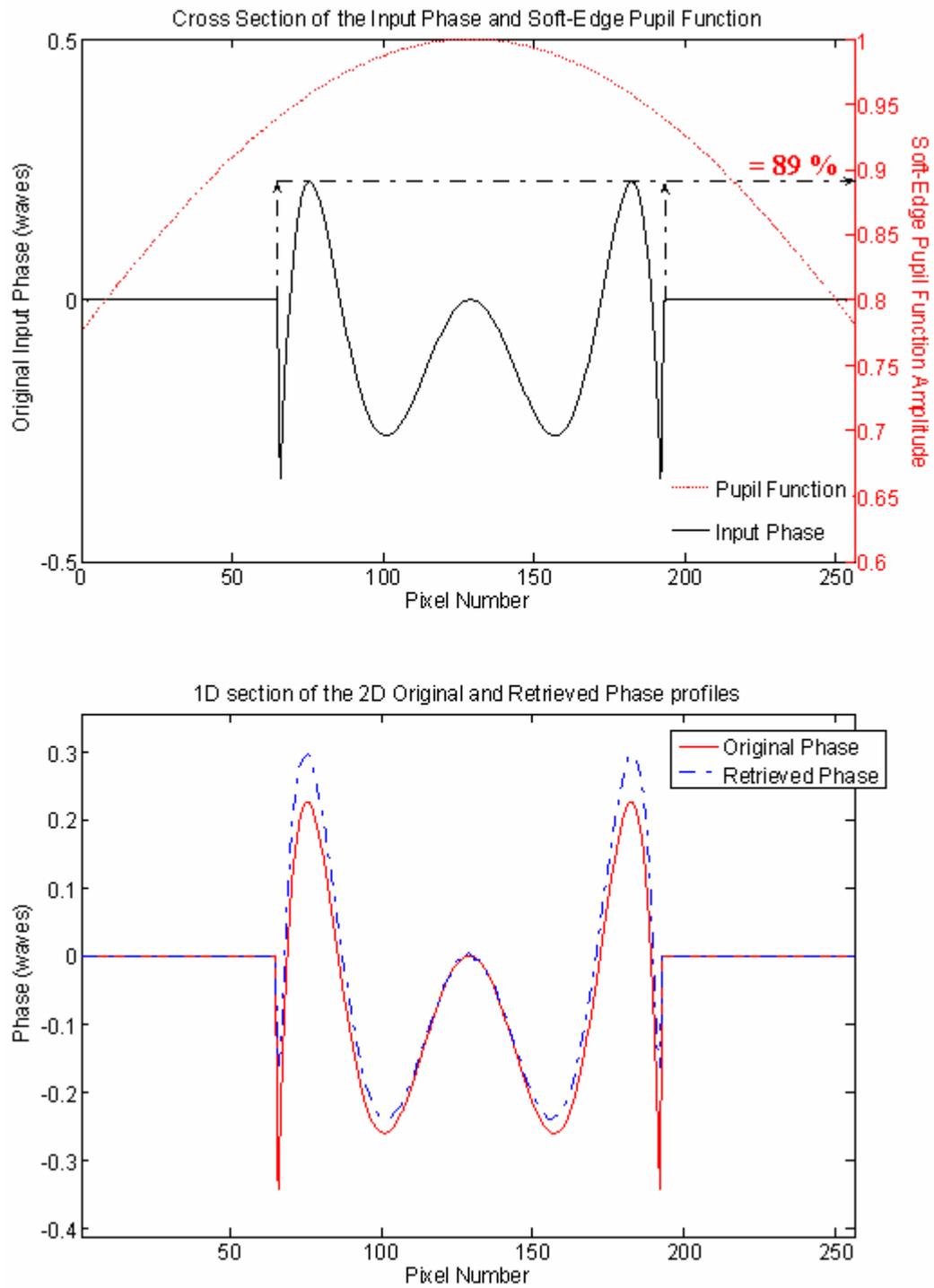


Figure 4.15 (a) a 1D cross section through the input phase function and the Gaussian pupil function. (b) a 1D cross section through the original and retrieved phase profiles for this example. The EM calculated for this reconstruction is 0.983.

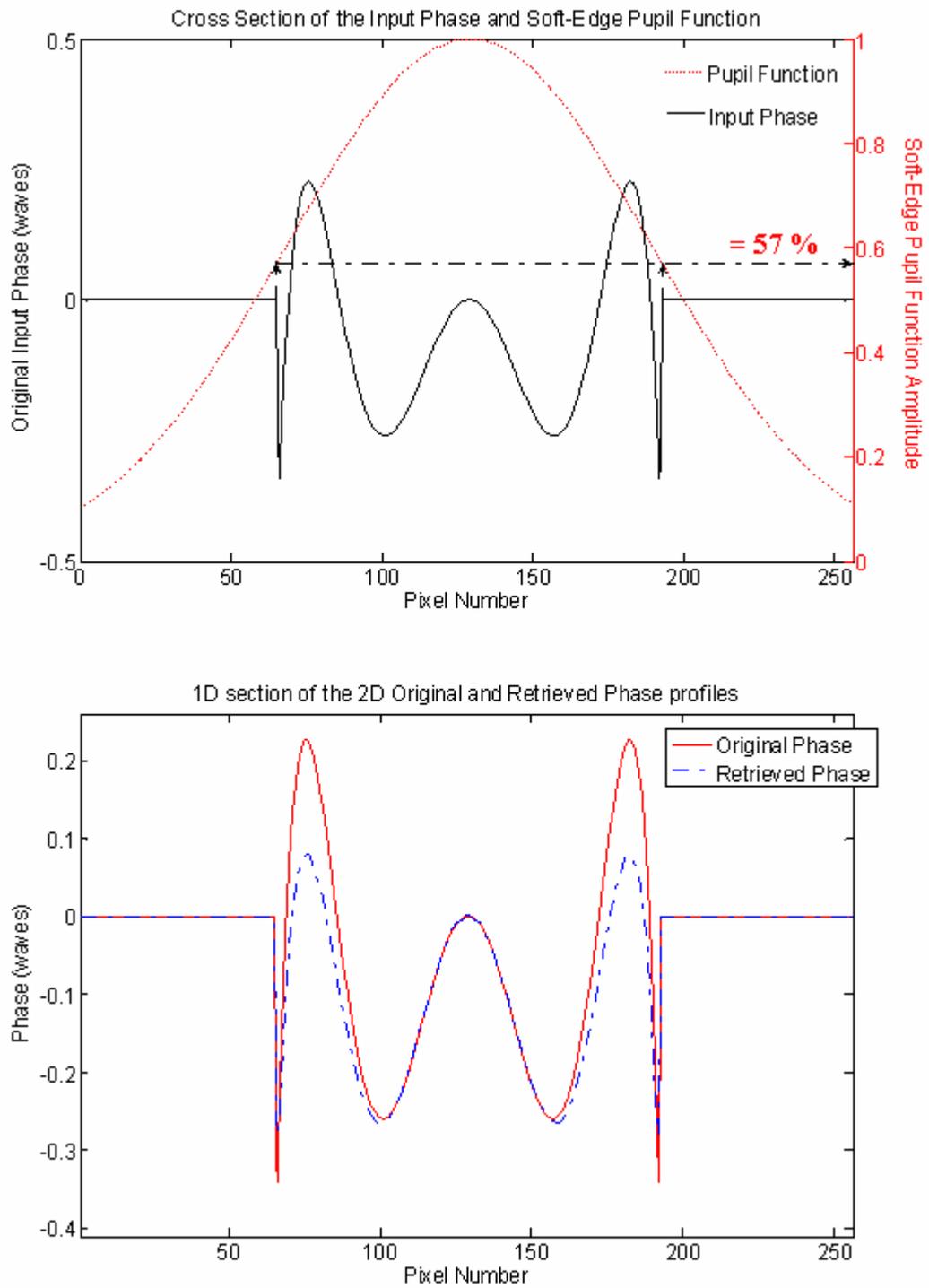


Figure 4.16 (a) a 1D cross section through the input phase function and the Gaussian pupil function. (b) a 1D cross section through the original and retrieved phase profiles for this example. The EM calculated for this reconstruction is 0.969.

Calculation of the EM (using Equation (4.18)) confirms that the fit of the retrieved profile to the original phase profile in Figure 4.15(b) (EM= 0.983) is noticeably better than the fit shown in 4.16(b) (EM=0.969). This indicates that multiplication of the input wavefront with a pupil function of this level of softness has actually reduced the intensity values at its edges enough to cause a new boundary value problem. As the edge softness increases the fit to the original data becomes increasingly poorer. The results in Figure 4.15 when compared to the hard-edged simulations (shown in Figures 4.8 and 4.9) suggest that a slightly soft-edged pupil will give superior phase retrieval results for smooth (continuous) input wavefronts.

The results presented thus far have shown that the SAE algorithm (based on Equation (4.13)), in cases where the local slope across the width of the blur function is not too great, is capable of getting close to the correct solution for the phase. The next step is to consider whether, once this initial retrieved phase profile has been calculated, iterative refinement may be used to improve the solution.

4.6.4 Iterative Refinement

It has been demonstrated that the SAE analytic algorithm is capable of retrieving the correct structure of the input phase, but often with amplitude errors. This may in part be due to a boundary value problem, and in part due to the assumptions made in the derivation of (4.13). The next step is to investigate whether, using the measured data as ‘ground truth’, an iterative stage could be added to check its validity and refine the solution.

In a practical situation the user will have *a priori* knowledge of the diversity function used and, if there is no scintillation, the pupil function. The only other information available is the measured data itself, the intensity images. Therefore it is sensible to build an iterative solution which uses the measured data as the ‘ground truth’ with which to compare the computed solution. When a solution is obtained the data which would create this phase profile can be computed using FT’s. This calculated data and the measured data can be compared. If there is no difference then the solution is correct

(at least as far as it is possible to know). In the case that a difference is found this difference may be used to formulate an iterative refinement of the solution.

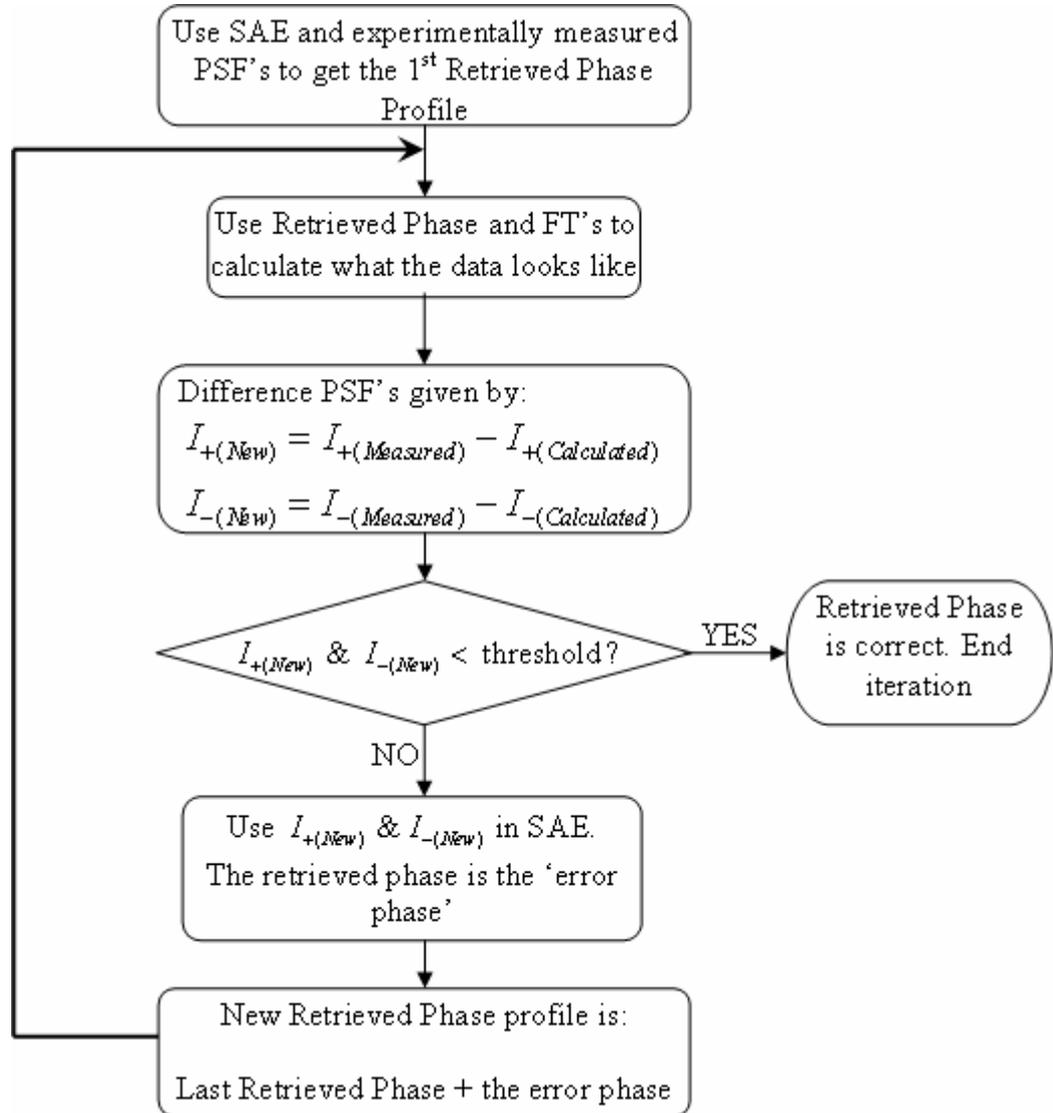


Figure 4.17 Flow chart of the analytic SAE algorithm with iterative refinement.

A data ‘error’ image is obtained by taking the difference between the calculated and measured versions of each intensity image, i.e. $I_{+(New)} = I_{+(Measured)} - I_{+(Calculated)}$ similarly for $I_{-(New)}$. These are called the ‘difference PSF’s’ and are input into the SAE algorithm to compute an ‘error phase’. This process is illustrated by the flow diagram in Figure 4.17.

Due to the linearity of the SAE with the data this error phase can be added to the last obtained retrieved phase profile to create a second, new, retrieved phase profile. This process is repeated until the difference between the measured and calculated data is minimised. The accuracy of the final phase solution will depend on several factors (as discussed in Section 4.6.3).

Figures 4.18 and 4.19 show the results of this iterative approach for two different test wavefronts, both using defocus as the diversity phase. In Figure 4.20 a simulation using Spherical Aberration diversity is presented. All the scales are in waves unless otherwise indicated. The EM's for the 1st and final solutions were calculated for each example and are summarised in Table 4.1. In these examples the SAE algorithm achieved a good initial fit to the original data analytically and iterative refinement has resulted in a slight improvement in each case

Figure Number	EM of 1 st Retrieved Solution	EM of Final Retrieved Solution
4.18	0.9997	0.9999
4.19	0.9995	0.9997
4.20	0.99992	0.99996

Table 4.1 Comparing the EM's calculated for the 1st and final retrieved phase profiles from the SAE algorithm with iterative refinement.

In these examples the iterative refinement process has converged very quickly to the correct solution. This was found to be the case for all smooth input wavefronts that did not violate the SAA conditions outlined in Section 4.4. In cases where the PV across the width of the blur function violated this condition the phase solution becomes ambiguous and this process was unable to converge to the correct solution (discussed further in Section 4.6.5). As in the soft-edged pupil simulations (see Figs 4.12 and 4.13) the quality of the solution was reliant on a suitable choice of the 'threshold' constant (described in Section 4.6.3). Again this was taken to be a very small value (10^{-5}) in each case.

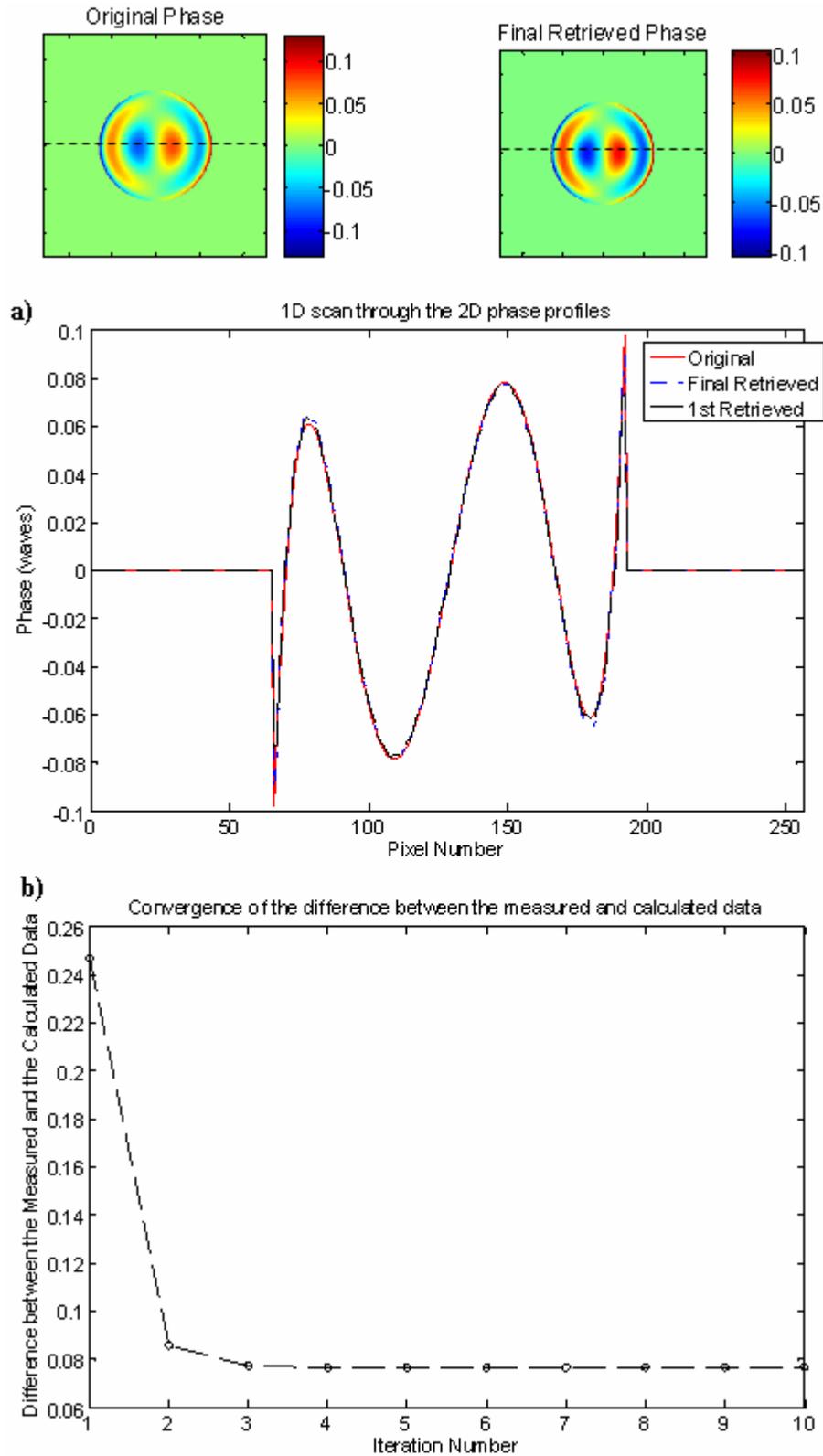


Figure 4.18 Original and Final Retrieved phase profiles (after 10 iterations) (a) a 1D cross section through each of them. Defocus diversity and a soft-edged pupil function were used. The convergence of the iterative solution is shown (b), the EM of the 1st retrieved profile was 0.9997, and the EM of the final solution was 0.9999.

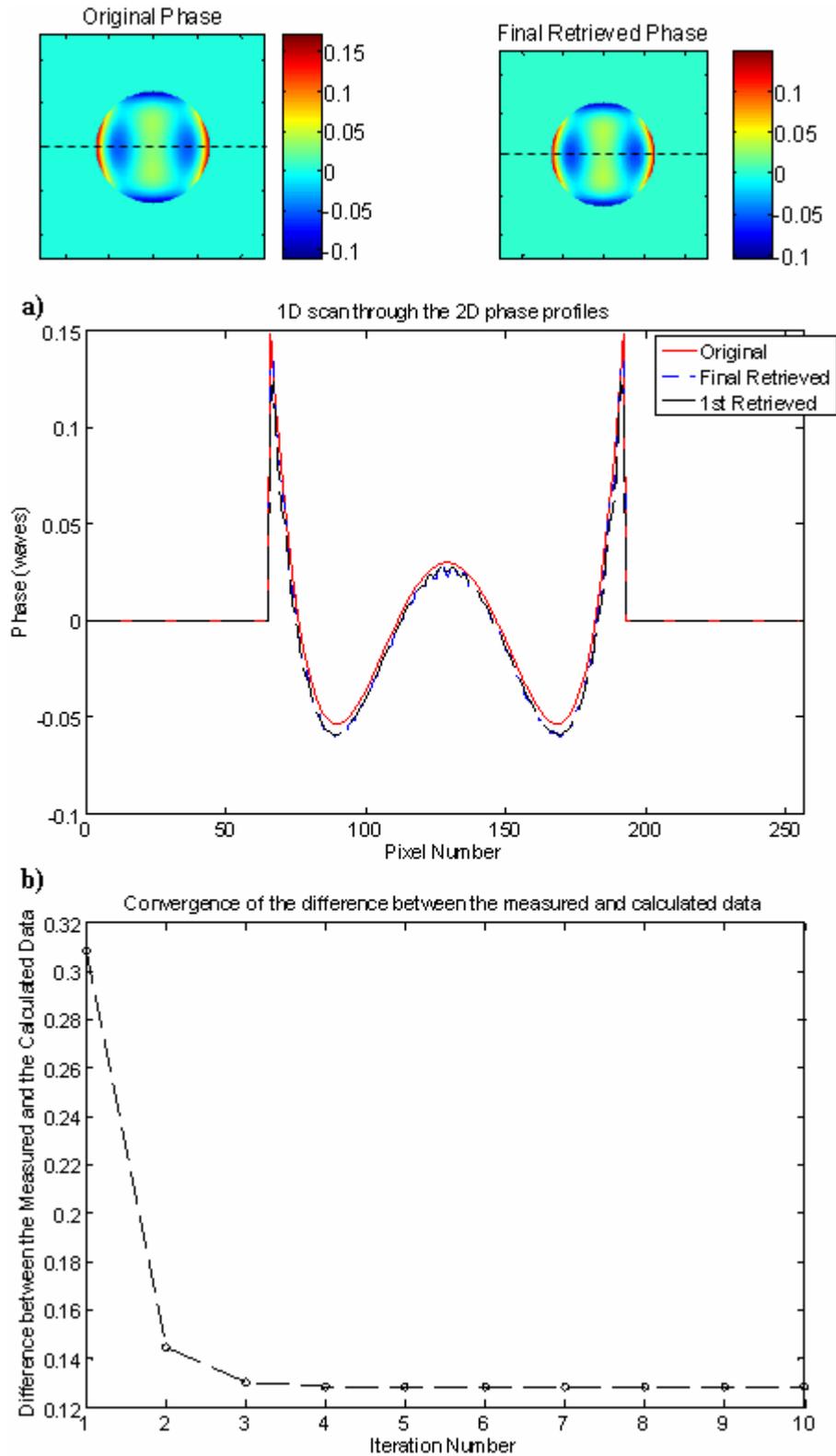


Figure 4.19 Original and Final Retrieved phase profiles (after 10 iterations) (a) a 1D cross section through each of them. Defocus diversity and a soft-edged pupil function were used. The convergence of the iterative solution is shown (b), the EM of the 1st retrieved profile was 0.9995, and the EM of the final solution was 0.9997.

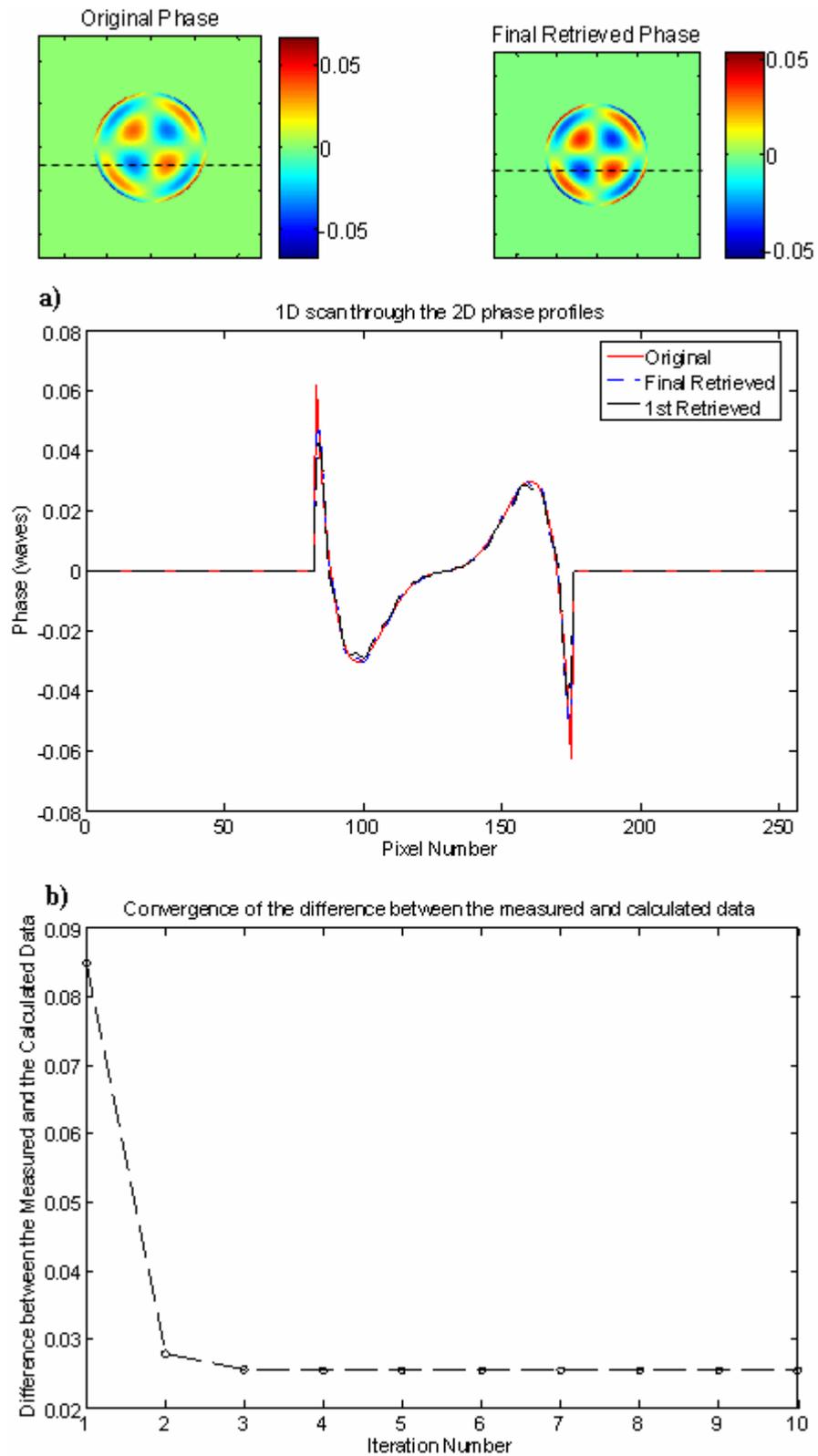


Figure 4.20 Original and Final Retrieved phase profiles (after 10 iterations) (a) a 1D cross section through each of them. Spherical Aberration diversity and a soft-edged pupil function were used. The convergence of the iterative solution is shown (b), the EM of the 1st retrieved profile was 0.99992, and the EM of the final solution was 0.99996.

The results presented here have demonstrated that it is possible to use the experimentally acquired data as a quality check for the solution of the input phase. Also, that with this simple process it is possible to iteratively refine the solution if the SAA has not been violated.

4.6.5 Iterations and Wavefronts with High Local Slope

When the local PV phase step across the width of the blur function is outwith the monotonic region ($\pm\pi/2$) described in Section 4.4 the SAE algorithm will instead reconstruct a corresponding phase in the linear region (as shown in Figures 4.4 and 4.6). As the size of the phase step increases (beyond the monotonic region) the quality of the SAE solution decreases. To demonstrate this a series of Matlab simulations, using functions with very high slopes, was conducted. Iterative refinement was used to investigate whether the SAE solution would be improved using this method.

In Figure 4.21 the results of one of these simulations are shown. The phase profiles in plots (a) and (b) show that the iterative algorithm has reconstructed the structure of the original function very well. Plot (c) contains the difference between the retrieved and original phase profiles. It shows that the difference is a constant piston error, of 0.5λ . The results shown in Section 4.4 can be summarised mathematically to say for phase steps $-0.25 \leq \varphi \leq 0.25$ the SAE will reconstruct φ correctly. For phase steps of $0.25 \leq \varphi \leq 0.75$ the SAE will reconstruct $(0.5\lambda) - \varphi$. Therefore in the example shown in Figure 4.21 the phase step across the blur function has fallen into the latter category and the difference between the original and retrieved phase profiles is 0.5λ . Plot (d) shows a 1D cross section through the original, first, and final retrieved phase profiles. This shows that the structure of the reconstructed phase is correct, and that the retrieved solution is merely shifted with respect to the original phase by the piston error of 0.5λ (shown in Figure 4.21(c)). This is consistent with the arithmetic presented in Chapter 3, which states that this method is blind to phase steps of $n\pi$ (0.5λ). It is also evident, by comparing the cross sections of the first and final retrieved solutions, that the iterative process has improved the solution slightly. In the example shown 10 iterations were

used because it was found that, for this case, the error between the measured and calculated data plateaus after 10 iterations.

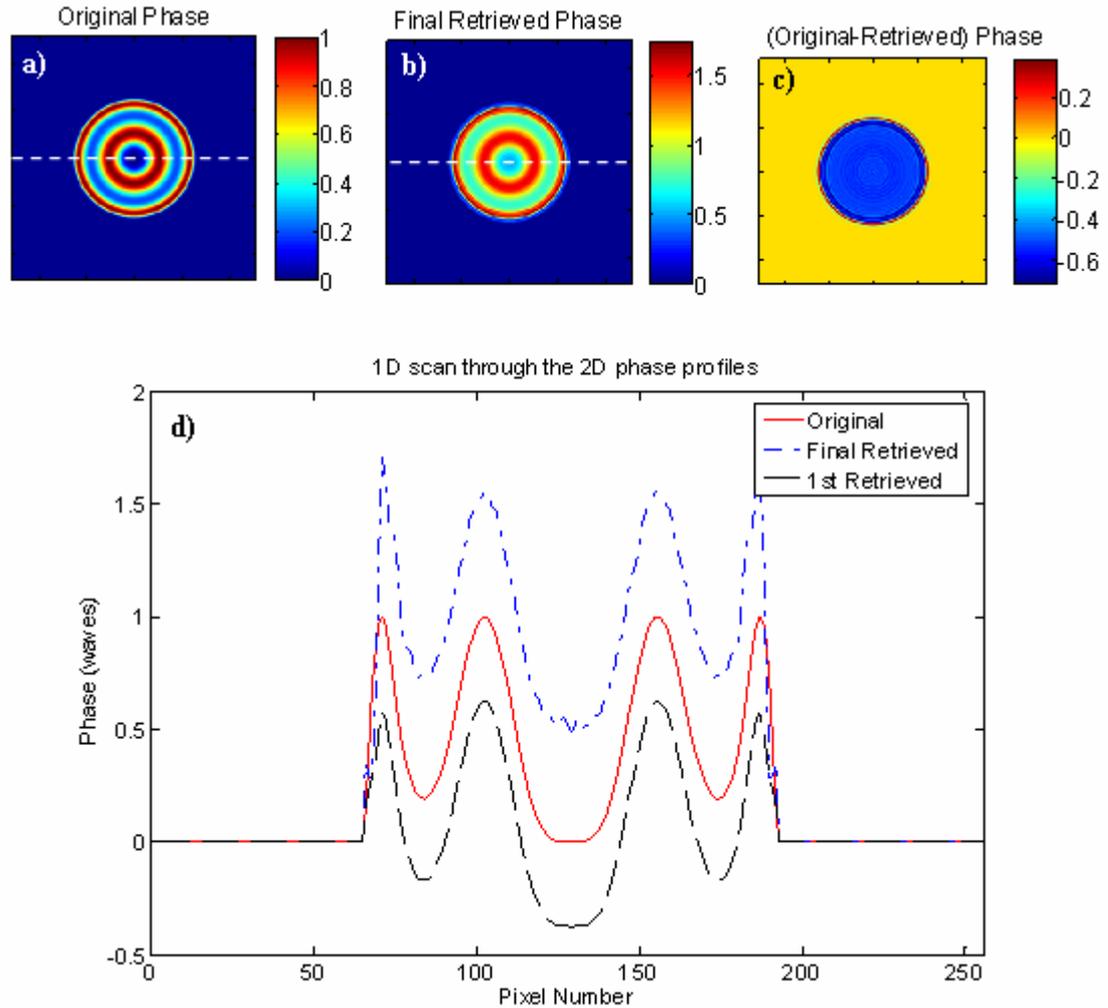


Figure 4.21 (a) the original high slope wavefront phase (b) the reconstructed phase achieved after 10 iterations (c) the difference between the retrieved and original phase (d) a 1D cross section of the original, first and final retrieved phase profiles [all scales are in waves]

Figure 4.22 is an example of the algorithm's failure to converge when the wavefront slope is too high. In this case the structure does bear some resemblance to the original phase profile, but with much less precision than any of the previous simulations. In this case, after 10 iterations, the solution begins to diverge which means that the algorithm is unable to reduce the errors between the measured and calculated data.

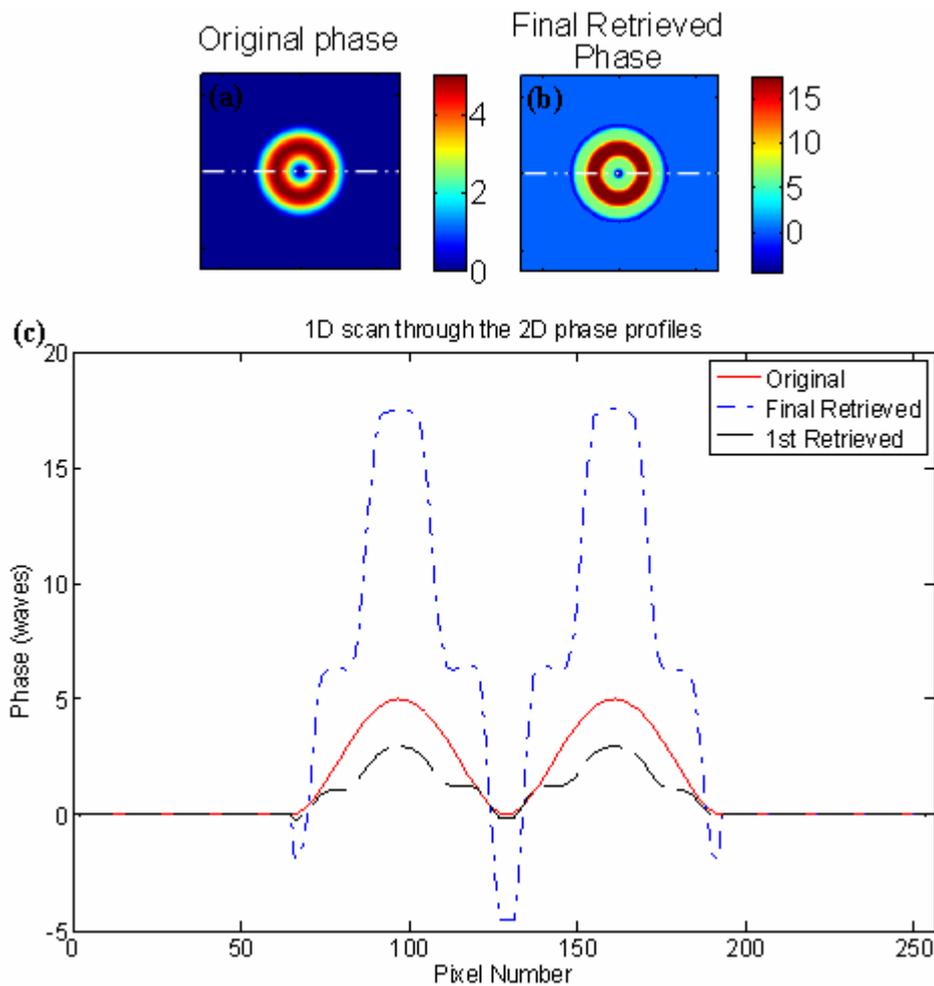


Figure 4.22 (a) the original phase profile, with local slope values that far exceed the 0.75λ limit (b) the retrieved phase profile (c) a 1D cross section of the original, first and final retrieved phase profiles [all scales are in waves].

From these results it may be concluded that the SAE algorithm works very well for phase steps of $\pm\pi/2$, will reconstruct to within a piston error for phase steps from $\pm\pi/2$ to $3\pi/2$, and will fail to provide a satisfactory solutions for phase errors beyond this. Since this phase step is the local phase change across the blur function and not the PV of the wavefront itself, this represents a significantly large range capability and this algorithm should prove a very versatile phase reconstruction method for continuous phase distributions.

4.6.6 Alternate Approaches

It would also be possible to implement other, well established techniques to refine the SAE solution, such as least squares fitting or a Steepest-Descent algorithm. Least squares is a mathematical optimization technique that attempts to find a "best fit" to a set of data by trying to minimize the sum of the squares of the differences (called *residuals*) between the fitted function and the data [29]. This is a well used method and image analysis software packages (such as Matlab) often have a built-in function which can perform least squares fitting.

Steepest Descent (or Gradient descent as it is also known) is another optimisation algorithm that is commonly used. This algorithm approaches a local maximum of a function by taking steps proportional to the gradient (or the approximate gradient) of the function at the current point. If instead one takes steps proportional to the negative of the gradient, one approaches a local minimum of that function [29]. In this case the user would seek to minimise the difference between the actual phase profile and the retrieved phase profile calculated by the SAE algorithm. The disadvantage of using this approach is that it can take a large number of iterations to converge if the slopes at adjacent points are very different. Also, finding the optimal step size to use can be time consuming, and in this case a Conjugate-Gradient method can be better. Conjugate-Gradient is a similar, but sometimes more effective method, which works on the same principle as Steepest Descent but which applies different rules in the selection of the direction that the function should be investigated in. This first direction chosen is the gradient of the value of the function at x_0 , and subsequent search directions are chosen to be the conjugate of this gradient, hence the name. Both of these are well know and trusted methods by which to optimise the iterative refinement procedure.

The method presented in Section 4.3.4 is simple, effective, and has been shown to work well for a range of situations. It is therefore the user's decision as to whether the added complexity of a Steepest-Descent or similar method would present greater benefits for their particular application.

4.7 Conclusions

In Chapter 3 it was shown that a null wavefront sensor, which produces an error signal only for aberrated wavefronts, could be constructed using the principles of GPD. While it is not always necessary for AO applications to reconstruct the wavefront, in metrology and other applications the ability to retrieve the unknown phase is desirable. By no longer being limited to the use of defocus as the diversity phase this called for the development of a new, non-ITE based, algorithm for phase recovery.

It has been shown, using the equation derived for the error signal in Chapter 3, that an analytic expression for the unknown phase of a test wavefront can be derived which is valid for small phase angles and uniform illumination. The extension to scintillated and discontinuous wavefronts will be discussed in the next chapter. For phase retrieval the algorithm requires a pair of intensity images to which equal and opposite diversity phase has been optically applied, and knowledge of the diversity and pupil functions used.

Regularisation, in the form of a simple threshold limit, was introduced to avoid ‘divide by zero’ errors at points where the PTF went to zero. The solution had no physical meaning at these points and was therefore rejected. This method is sufficient for simulated, noise-free, data but a more rigorous approach will be required for use with real experimental data.

It was shown that, in the context of this algorithm, ‘small angles’ do not refer to the PV deformation of the wavefront, but rather to the small local areas of the wavefront defined by the width of the blur function. This is a significant step forward as it means that this method should be capable of dealing with quite severely aberrated wavefronts provided that the phase change across the width of the blur function is not too steep. This advantage can be maximised by selecting a diversity whose blur function will have a very narrow width, and by increasing the sampling (number of pixels) of the wavefront. For unambiguous data to be generated the phase step must fall within the $\pm\lambda/4$ ($\pi/2$) region since the error signal was shown to vary sinusoidally with the step

size. It was also demonstrated that for phase steps between $+\lambda/4$ and $+3\lambda/4$ or $-\lambda/4$ and $-3\lambda/4$ the SAE will calculate the phase to be within the linear region at the point $(0.5\lambda) - \varphi$.

When this algorithm was first tested it was discovered that there were often amplitude errors in the reconstructed phase, but that the basic structure was well retrieved. From this it was supposed that the errors may be caused by a boundary value problem caused when the convolution of the wavefront with the blur function reached the outer edge of the wavefront and the sudden change in phase slope violated the SAA. In an attempt to solve this problem a Gaussian profile, soft-edged, pupil function was introduced. This was found to work very well, provided only the very central area of the Gaussian pupil function was used. As the degree of edge ‘softness’ increased the fit of the solution to the original phase function decreased due to the intensity values at the edges being significantly altered by multiplication with the Gaussian.

Finally an iterative refinement stage was introduced to improve the fit of the solution to the original phase function by comparison of the ‘measured’ data with the data calculated from the solution. The error between the measured and calculated data typically converged in around 10 iterations or less, making this a very fast procedure. Convergence was dependent on a sensible choice of the threshold value used for regularisation, the degree of pupil function softness, and whether the local phase change across the blur function was within the linear $\pm\lambda/4$ region. The possible application of this iterative form of the SAE algorithm to the shaping of femtosecond laser pulses will be discussed in Chapter 6.

In conclusion, a fast and robust algorithm has been developed to use in combination with the GPD wavefront sensor for phase retrieval of non-scintillated wavefronts. This algorithm combines analytic and iterative elements to maximise the accuracy of the solution without sacrificing computational speed. Experimental testing and further applications of this system to segmented optics and scintillated wavefronts will be the subject of the following chapters.

4.8 References

1. Woods, S.C. and A.H. Greenaway, *Wave-Front Sensing by Use of a Green's Function Solution to the Intensity Transport Equation*. J.Opt.Soc.Am.A, 2003. **20**(3): p. 508-512.
2. Teague, M.R., *Deterministic Phase Retrieval: A Green's Function Solution*. J.Opt.Soc.Am.A, 1983. **73**(11): p. 1434-1441.
3. Misell, D.L., *An Examination of an Iterative Method for the Solution of the Phase Problem in Optics and Electron Optics*. Journal of Physics D, 1973. **6**: p. 2200-2216.
4. Gureyev, T.E., A. Roberts, and K.A. Nugent, *Phase Retrieval with the Transport-of-Intensity Equation: Matrix Solution with Use of Zernike Polynomials*. J.Opt.Soc.Am.A, 1995. **12**(9): p. 1932-1941.
5. Gureyev, T.E. and K.A. Nugent, *Phase Retrieval with the Transport-of-Intensity Equation. Ii. Orthogonal Series Solution for Nonuniform Illumination*. J.Opt.Soc.Am.A, 1996. **13**(8): p. 1670-1682.
6. Gonsalves, R.A., *Small-Phase Solution to the Phase-Retrieval Problem*. Optics Letters, 2001. **26**(10): p. 684-685.
7. Gerchberg, R.W. and W.O. Saxton, *A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures*. Optik (Stuttgart), 1972. **35**: p. 237-246.
8. Fienup, J.R., *Reconstruction of a Complex-Valued Object from the Modulus of Its Fourier-Transform Using a Support Constraint*. J.Opt.Soc.Am.A, 1987. **4**(1): p. 118-123.
9. Greenaway, A.H., *Proposal for Phase Recovery from a Single Intensity Distribution*. Optics Letters, 1977. **1**(1): p. 10-12.
10. Burge, R.E., M.A. Fiddy, A.H. Greenaway, and G. Ross, *The Phase Problem*. Proc. Roy. Soc.A, 1976. **350**(1661): p. 191-212.
11. Saisse, M., K. Rousset, and E. Lazarides, *Modelling Technique for the Hubble Space Telescope Wave-Front Deformation*. Applied Optics, 1995. **34**(13): p. 2278-2283.
12. Barrett, T.K. and D.G. Saundler, *Artificial Neural Network for the Determination of Hst Aberration from Stellar Images*. Applied Optics, 1993. **32**: p. 1720-1727.
13. Burrows, C.J., A. Holtzman, S.M. Faber, and e. al, *The Imaging Performance of the Hubble Space Telescope*. Astrophys.J.Lett., 1991. **369**: p. L21-L25.
14. Roddier, C. and F. Roddier, *Combined Approach to the Hst Wave-Front Distortion Analysis*. Applied Optics, 1993. **32**: p. 2992-3008.
15. Roddier, C. and F. Roddier. *Reconstruction of the Hst Mirror Figure from out-of-Focus Stellar Images*. in *Space Astronomical Telescopes and Instruments*. 1991. Vol. 1494. p. 78-84
16. Fienup, J.R., *Phase Retrieval Algorithms: A Comparison*. Applied Optics, 1982. **21**(15): p. 2758-2769.
17. Baba, N., H. Tomita, and M. Noriaki, *Iterative Reconstruction Method in Phase Diversity Imaging*. Applied Optics, 1994. **33**(20): p. 4428-4433.
18. Baba, N. and K. Mutoh, *Measurement of Telescope Aberrations through Atmospheric Turbulence by Use of Phase Diversity*. Applied Optics, 2001. **40**(4): p. 544-552.

19. Fienup, J.R. and A.M. Kowalczyk, *Phase Retrieval for a Complex-Valued Object by Using a Low-Resolution Image*. Journal of the Optical Society of America a-Optics Image Science and Vision, 1990. **7**(3): p. 450-458.
20. Ivanov, V.Y., V.P. Sivokon, and M.A. Vorontsov, *Phase Retrieval from a Set of Intensity Measurements: Theory and Experiments*. Journal of the Optical Society of America, 1992. **9**(9): p. 1515-1524.
21. Wild, W.J., *Linear Phase Retrieval for Wave-Front Sensing*. Optics Letters, 1998. **23**(8): p. 573-575.
22. Ellerbroek, B. and D. Morrison. *Linear Methods in Phase Retrieval*. in *Wavefront Sensing*. 1983. San Diego, CA: Proc. SPIE. Vol. 0351. p. 90-95
23. Give'on, A., N.J. Kasdin, R.J. Vanderbei, D.N. Spergel, M.G. Littman, and P. Gurfil. *Stochastic Optimal Phase Retrieval Algorithm for High-Contrast Imaging*. in *Astronomical Adaptive Optics Systems and Applications*. 2003: Proc. SPIE. Vol. 5169. p. 276-287
24. Carmody, M., L.D. Marks, and K.L. Merkle, *Current Transport Along Grain Boundaries in D-Wave Superconductors*. Physica C, 2002. **370**: p. 228-238.
25. Thust, A., M. Lentzen, and K. Urban, *The Use of Stochastic Algorithms for Phase Retrieval in High Resolution Transmission Electron Microscopy*. Scanning Microscopy, 1997. **11**: p. 437-454.
26. Mitchell, M., *An Introduction to Genetic Algorithms*. 1st ed. Complex Adaptive Systems Series. 1998: The MIT Press.
27. Baker, R., *Genetic Algorithms in Search and Optimisation*. Financial Engineering News, 1998(5).
28. Chanan G, T.M., Sirko E, *Phase Discontinuity Sensing: A Method for Phasing Segmented Mirrors in the Infrared*. Applied Optics, 1999. **38**(4): p. 704-713.
29. Atkinson, K.E., *An Introduction to Numerical Analysis*. 2nd ed. 1989: John Wiley and Sons.

Chapter 5

Further Study of the Small Angle Expansion Phase Retrieval Algorithm

Summary of Chapter 5

In Chapter 4 the Small Angle Expansion (SAE) was proposed as a new algorithm for the retrieval of phase information from data obtained by a Generalised Phase Diversity (GPD) wavefront sensor. In this chapter we will study the fundamental properties of the Phase Transfer Function (PTF), whose form depends on the diversity function chosen, and which is used in the SAE algorithm to perform the deconvolution with the measured data. Optimisation of the GPD wavefront sensor through proper choice or manipulation of the PTF will be considered. In this chapter several practical examples of the SAE will be given. These include its performance with non-uniform illumination and discontinuous wavefronts and its use within the familiar application of Phase Contrast Microscopy (PCM).

5.1 Introduction

In Chapter 3 the concept of Generalised Phase Diversity (GPD) was described and it was shown how a null wavefront sensor could be constructed using this principle. In Chapter 4 the extension of GPD to full wavefront sensing was discussed and equations were presented which can be used to retrieve the unknown wavefront phase using the error signal provided by the GPD wavefront sensor. The results shown in Chapter 4 demonstrate how effective this solution is when used on simulated data.

The SAE simulations presented thus far have all used continuous example wavefronts with uniform illumination. As discussed in Chapter 2, one of the key driving factors for the development of a new wavefront sensor is the need to be able to cope with scintillated and discontinuous wavefronts. In this chapter the extension of the SAE to these types of wavefront will be presented and its performance assessed. In light of

these results, further discussion of the possible applications of this new sensor system will be given in Chapter 6.

In Chapter 4 it was shown that the SAE is simply a deconvolution process between the measured phase diverse data and the Phase Transfer Function (PTF). The PTF is created by the properties of the phase diversity filter function programmed into the diffraction grating. Therefore, the user has full knowledge of the PTF and can choose a function whose properties will suit their particular application. In this Chapter the PTF will be studied in more detail. The question of whether the PTF may be optimised for particular applications, which will be called ‘PTF Engineering’, will be discussed. Practical implications of the form of the PTF will be considered here and extended in Chapter 6 where the GPD sensor and SAE algorithm will be tested experimentally.

Finally, in this chapter a famous example of phase diversity will be studied. The Phase Contrast Microscope (PCM) is a widely used technique, particularly in the field of Biology. In this method small phase changes created by the sample are observed by adding a phase shift to the measured images. In this chapter it will be shown that the PCM phase shift can be made to satisfy the symmetry conditions for GPD. Simulations will be presented which show the performance of a GPD sensor with PCM diversity with the SAE used to analyse the data.

5.2 The Phase Transfer Function

In Chapter 4 it was shown that the phase of a wavefront may be reconstructed using the GPD null sensor error signal (given by Equation 3.16) and the Phase Transfer Function (PTF) created by the diversity filter function. Equation 4.13, reiterated below, shows that the error signal ($d(r)$) is the result of the convolution of the input phase ($\varphi(r)$) with the phase blur function ($t(r) = \Im\{\tilde{i}_m R(\xi) - \tilde{r}_e I(\xi)\}$):

$$\frac{d(r)}{4I_0(r)} = \int d\xi \Phi(\xi) [\tilde{i}_m R(\xi) - \tilde{r}_e I(\xi)] e^{-ir\xi} . \quad (4.13)$$

The form of this equation will differ if, instead of the wavefront phase, the imaginary part of the wavefront ($a(r) = \Im\{A(\xi)\}$) is reconstructed. However, the decision of which wavefront property to reconstruct will have no affect on the PTF itself, and this function is the only part of the inverse problem which the user can choose. It is therefore vital that the diversity function is chosen to provide the most useful PTF for any given application. The form of the PTF will determine which spatial frequencies will be sampled best and, in regions where it falls to zero, which frequencies the reconstruction will be blind to. As mentioned in Chapter 4 there is also a classic regularisation problem which arises whenever the PTF falls to zero, so care must be taken to minimise this effect where possible and to provide good regularisation when zero-points are unavoidable. This will be discussed further in Chapter 6.

In the following sections the form of the PTF for several allowable diversity filter functions will be studied and compared. In Section 5.2.2 the question of optimisation and whether the PTF can be specifically chosen to have minimal numbers of zero crossings will be addressed.

5.2.1 Comparisons of the PTF's Generated by Different Diversity Functions.

The PTF is formed by a combination of the real and imaginary parts of the diversity filter function. The overall structure and form of allowable diversity filters (see the conditions presented in Chapter 3) can vary greatly from function to function, and so too can their associated PTF's. In this section some examples of PTF's will be given and their properties discussed.

In Equation 4.13 the PTF, in the Fourier plane, is given by $T(\xi) = \tilde{i}_m R(\xi) - \tilde{r}_e J(\xi)$. This may also be written in the form given in Equation 5.1, and it is this equation which is used to generate the PTF's included in this section:

$$T(\xi) = \Im\{[h(r) \otimes i_m(r)] \cdot r_e(r) - [h(r) \otimes r_e(r)] \cdot i_m(r)\} . \quad (5.1)$$

Matlab simulations were conducted to simulate the PTF which would be generated by several different allowable diversity functions. The results obtained from 4 of these are shown in Figures 5.1 to 5.4. These include defocus (Z_2^0) and spherical aberration (Z_4^0), since much comparison has been made between these functions in previous chapters. Also chosen for inclusion here were the results from astigmatism (Z_2^2) and the higher order aberration Z_6^6 , although both of these are allowable functions by their symmetry properties, they shall prove to be unsuitable for reasons which shall be discussed. In each case the strength of the diversity filter function was set to 1λ and $h(r)$ (see Equation (5.1)), the real part of the input wavefront, was assumed to be a plane wave of unit amplitude limited by a hard edged pupil function. All scales are in waves unless otherwise stated.

In each cross section plot of the PTF's (plot (c) in Figures 5.1-5.4) the lens aperture representing the transparent region of the camera lens is shown by the red dashed line. Points where the PTF is low valued (crosses the zero position on the y-axis) will lead to regularisation problems (as introduced in Chapter 4). These are highlighted by the dotted circles in each of these plots.

Figures 5.1 and 5.2 show the results obtained from the two radially symmetric diversity filters; defocus and spherical aberration. In both cases the 1D scan of the PTF shows it to be zero at the origin (the centre of the plot at $n/2$ where n is 128 pixels in these examples); this is not unexpected. The real part of the diversity function, $r_e(r)$, is related to the Cosine of the diversity phase which will be 1 at the origin. However, the imaginary part, $i_m(r)$, depends on the Sine of the diversity phase and will therefore go to zero at the origin. From the form of Equation (5.1) it may be expected that wherever either the real or imaginary part of the diversity phase goes to zero, so too will the PTF. This will therefore always be the case at the origin. All the arithmetic presented thus far is based on the assumption that the difference image $d(r)$ should go to zero for plane wavefronts. For this reason this analysis is blind to piston error provided that the wavefront is flat, and therefore any d.c term present will be neglected. This d.c term would be located at the origin in Fourier space and thus is cancelled by the null in the PTF at this position. If the sum of the intensity images, instead of their difference, were

considered then even for suitable GPD functions the signal will not be zero for plane wavefronts. This fact may allow a second analysis to be performed using the sum image to reconstruct spatial frequencies near the origin which are lost in the present case wherever the PTF goes to zero. Further study of the potential and usage of the sum image will be confined to future work.

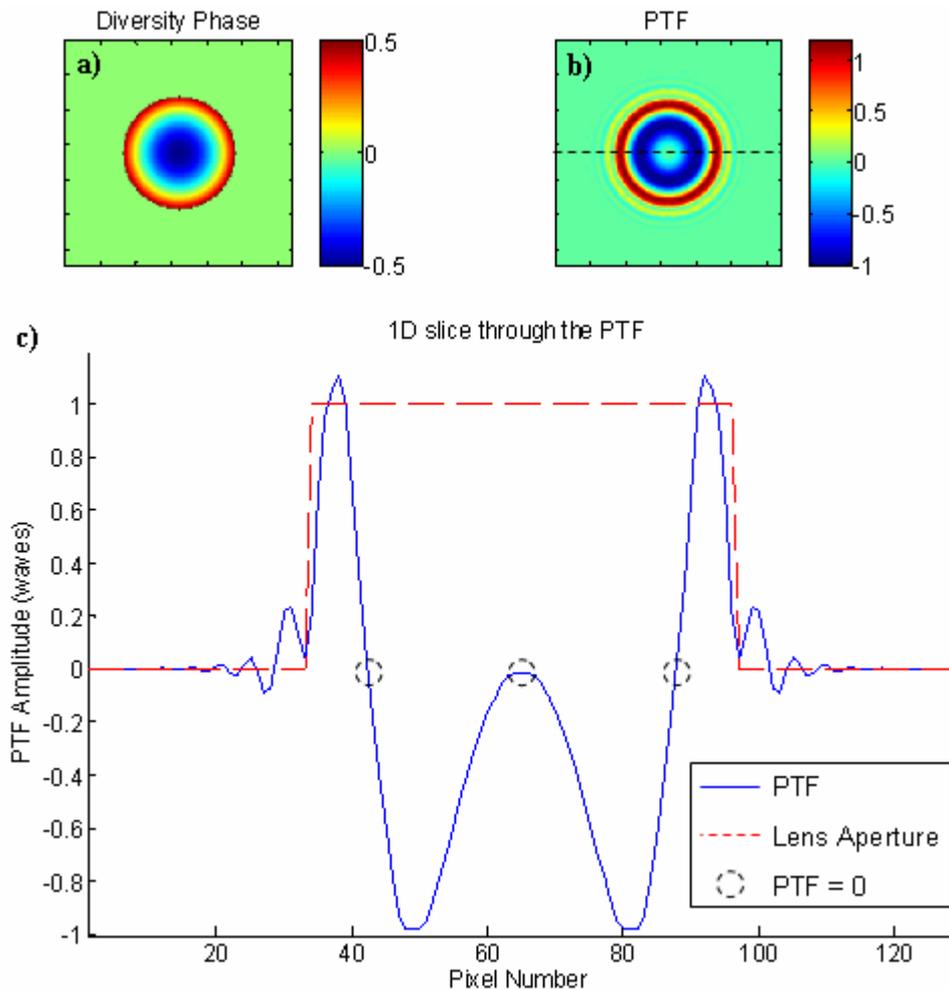


Figure 5.1 a) The phase of the diversity filter function (defocus) b) The PTF this generated (in Fourier space) c) a 1D scan of the PTF, through the line shown in plot b.

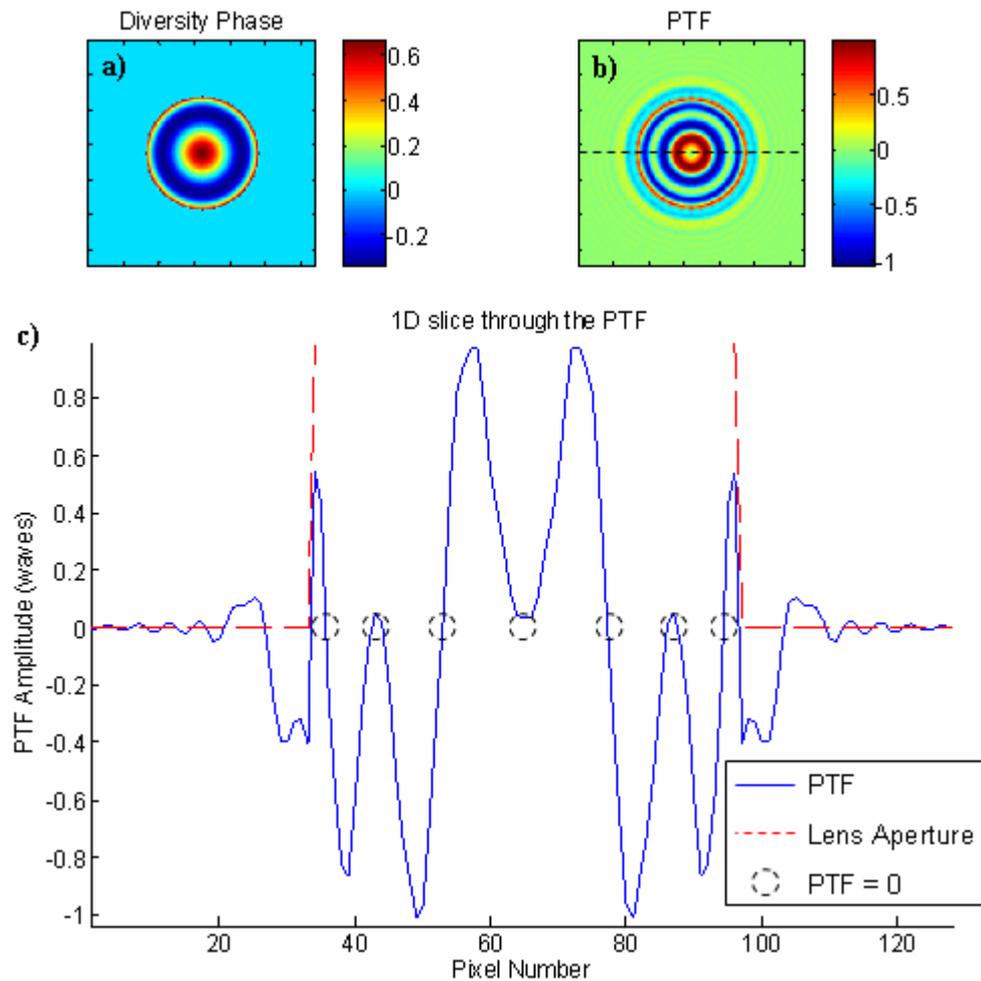


Figure 5.2 a) The phase of the diversity filter function (spherical aberration) b) The PTF this generated (in Fourier space) c) a 1D scan of the PTF, through the line shown in plot b.

As it has now been ascertained that the PTF should tend to zero at the origin for all suitable null sensor diversity filters, the behaviour of the PTF across the width of the pupil must be compared to discover the most suitable functions to use. To allow regularisation to be implemented more easily, and also to minimise the loss of information, a desirable diversity function should have a PTF with as few ‘zero crossings’ as possible (highlighted by the dotted circles in Figures 5.1-4). Wherever these crossings occur the frequencies contained in that part of the wavefront will be irretrievably lost. Since it is unavoidable that all functions should have at least one such crossing at the origin it would be preferable to use a function whose PTF recovers quickly, rising sharply about the origin, to sample as many of the low frequencies as

possible. With these qualities in mind it is now possible to assess the relative suitability of the two functions presented in Figures 5.1 and 5.2.

In Figure 5.1(c) the defocus filter PTF is shown to have 2 zero crossings (in addition to the origin) compared to 6 for the spherical aberration filter PTF shown in 5.2(c). However, the slope of the spherical aberration PTF is much higher than that of the defocus, and will therefore recover more quickly to sample the frequencies about the zero points. Therefore regularisation will be more of an issue when using spherical aberration diversity but the loss of information should not be too detrimental.

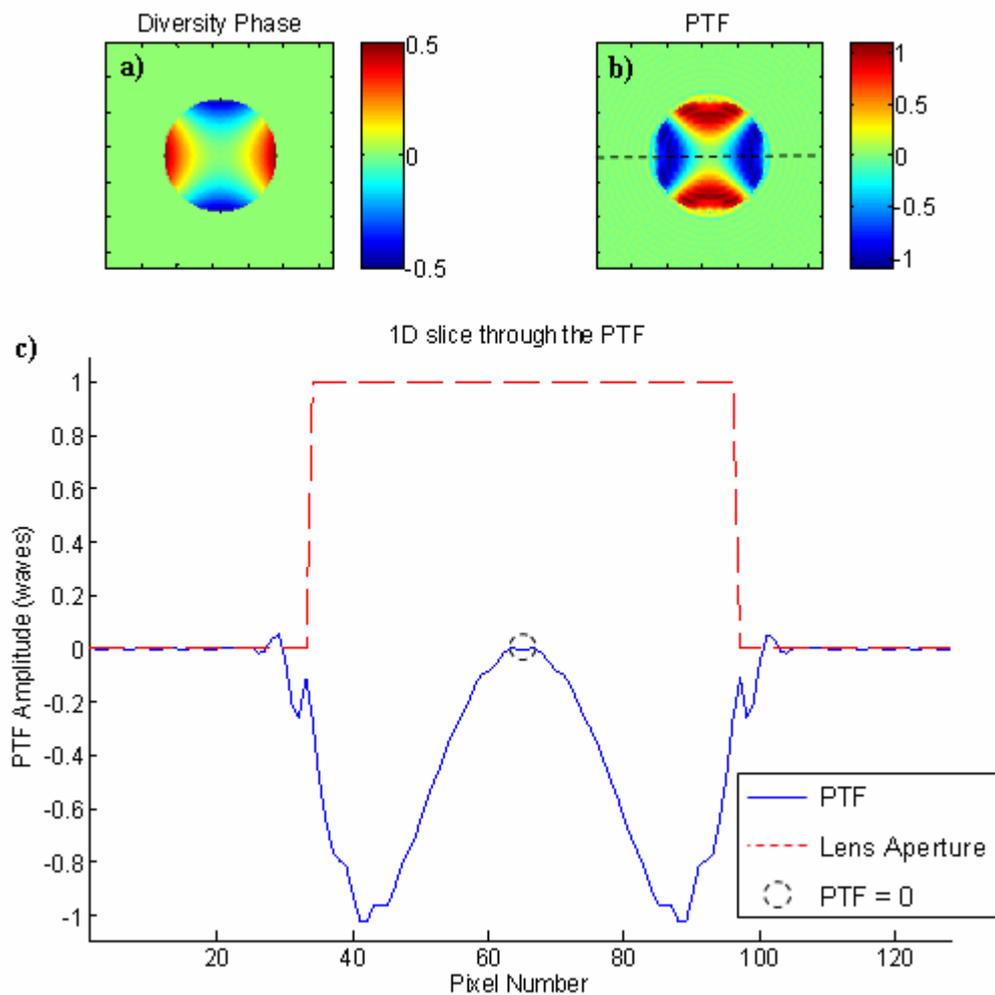


Figure 5.3 a) The phase of the diversity filter function (Astigmatism) b) The PTF this generated (in Fourier space) c) a 1D scan of the PTF, through the line shown in plot b.

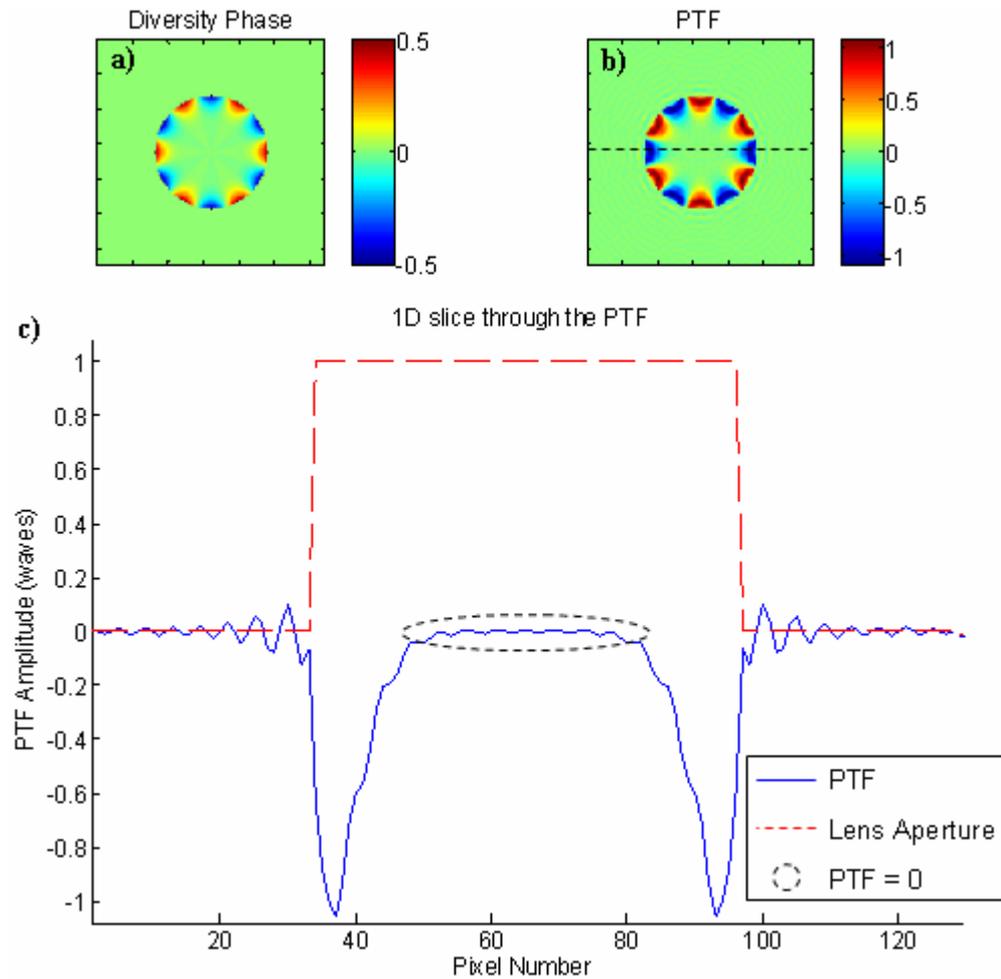


Figure 5.4 a) The phase of the diversity filter function (Z_6^6) b) The PTF this generated (in Fourier space) c) a 1D scan of the PTF, through the line shown in plot b.

Figures 5.3 and 5.4 contain the same results, as calculated for astigmatism and one higher order aberration Z_6^6 . Considering only 5.3(c), the 1D section of the astigmatism PTF, it would appear that it is as favourable as the defocus PTF (see 5.1(c)). However, looking at 5.3(b) this is clearly shown to be untrue. In the astigmatism PTF there is a X like structure of zeros within it which is not seen by the 1D section because it is not trivial to take the slice at 45° to either axis. The X-like structure is particularly detrimental as it causes a loss of both low and high frequency information, unlike a ring of zeros which would only affect one or the other. This extended area of zeros within the PTF makes this an unsuitable diversity function to use as the loss of information would be too great. Figure 5.4 shows a similar problem with the higher order aberration Z_6^6 . The structure of this function (shown in 5.4(a)) is a disc with a large flat region for

much of its area and a high frequency ripple around the edges. Subsequently the PTF formed from this function has a large extended region of zeros, and only selected high frequency structures would be reconstructed at all. For most applications this would be completely unsuitable, therefore Zernike functions of this form should not be considered for use in a GPD wavefront sensor.

In this section results have been presented which have allowed, at least in the broadest sense, the desirable properties of a PTF to be described and some examples of suitable and unsuitable functions shown. The next question to ask is whether there is any way to manipulate the PTF, using combinations of filter functions, to optimise the performance of the wavefront sensor.

5.2.2 Possible Optimisation of the PTF

This section will address the issue of optimisation of the PTF through use of more than one diversity function. In combining functions to create new, and hopefully improved, PTF's the method of combination must also be considered. The most practical solution is to add the diversity phases linearly. In a real experiment this could perhaps be implemented using a crossed grating, where the rulings in one direction apply one diversity and the rulings in the perpendicular direction add a second diversity phase. The question here is whether using linear combinations of different diversity phases the zero points can be cancelled out completely, or bunched at the edges of the pupil, to simplify the regularisation problem and maximise the sensitivity of the wavefront sensor.

5.2.2.1 Crossed Gratings: A Method of Applying Multiple Diversities

A practical means of applying more than one diversity phase to the input wavefront would be to use 'crossed gratings'. A pair of GPD phase gratings are physically placed together such that their rulings are positioned perpendicular with respect to each other (i.e. crossed). In Figure 5.5 an example of such a pair is given. The orientation of the diffraction orders produced by each grating is shown in this diagram. To identify which

diffraction spot belongs to which grating the notation used is ${}_nG^m$ where m is the number of the grating (1 or 2 in this example) and n is the order number (i.e. ± 1 or 0).

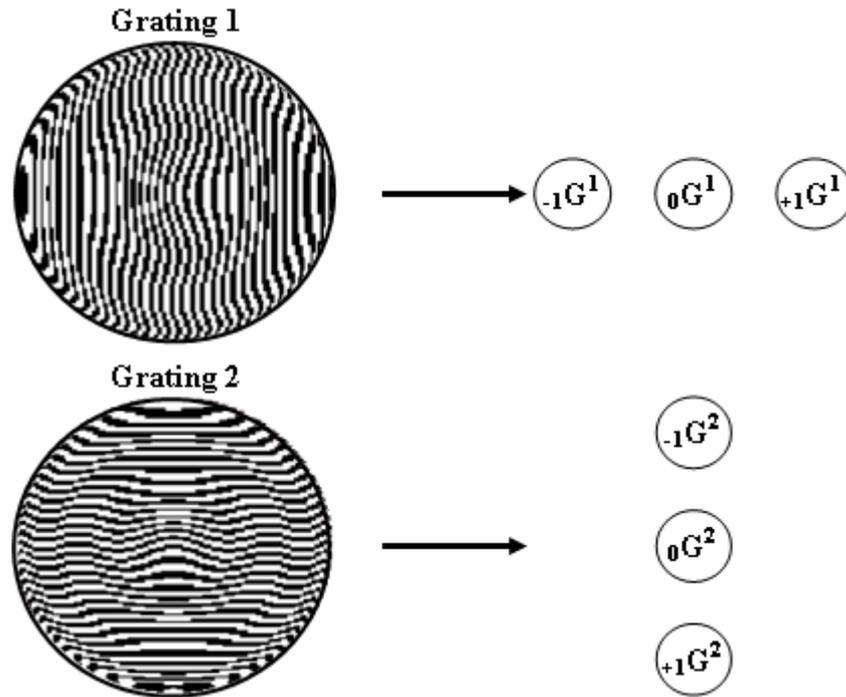


Figure 5.5 A pair of GPD gratings which can be ‘crossed’ by overlaying them. Their associated diffraction orders and their relative orientation is also shown.

There are two possible patterns (see Figure 5.6) which could be produced when Gratings 1 and 2 are overlaid. Which pattern is observed will depend on the grating etch depths. If Gratings 1 and 2 are considered to be simple two-level phase gratings then light of wavelength λ will be diffracted by the grating provided the phase delay, defined by the etch depth, is not modulo 2π . In this description a wavelength will be said to ‘see’ a grating if the etch depth for this wavelength is not modulo 2π . Figure 5.6 contains a plot of the two possible outputs of crossed gratings. If both gratings have the same etch depth, then a single wavelength λ_1 will ‘see’ both gratings and the pattern in Figure 5.6 (a) will be produced. This same pattern will also be observed if the gratings are etched such that two input wavelengths, λ_1 & λ_2 , will each ‘see’ both gratings. The second option is for Gratings 1 and 2 to be each etched to give a phase delay of an integral number of 2π for one of the input wavelengths (e.g. Grating 1 has an etch depth

of modulo 2π for λ_1 , but will be ‘seen’ by λ_2). When this is the case each wavelength will only ‘see’ one grating, the other will appear to be simply a sheet of glass, and the output will not be combined in the same manner as in Figure 5.6.(a). The diagram in Figure 5.6 (b) shows what this pattern would look like. This principle was first described by Dammann [1] whose Colour Separation Gratings used stepped phase structures in a single grating to separate incoming white light into its component wavelengths (each being diffracted into a different order). Dammann’s colour separation technique is currently in use today in a wide field of applications and therefore fabrication of gratings of this type is well understood and relatively simple [2].

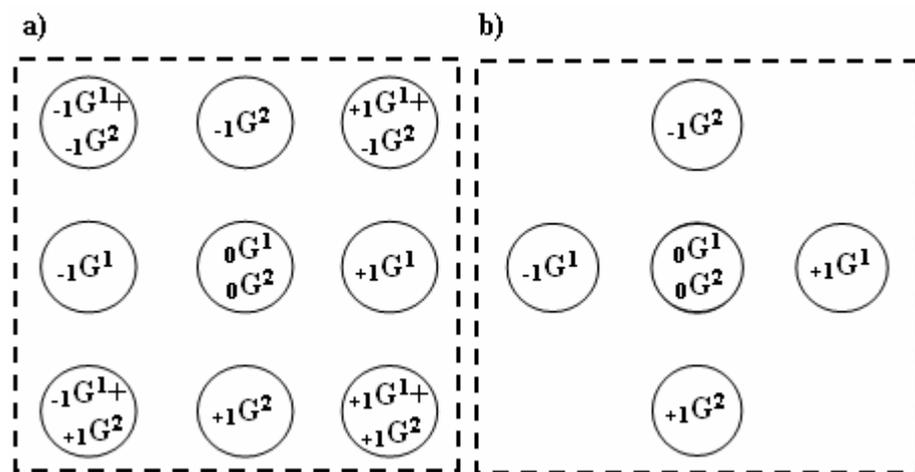


Figure 5.6 The diffraction pattern created when (a) Grating 1 and 2 are etched to the same depth (both designed for λ_1) or etched such that λ_1 and λ_2 ‘see’ the same grating (b) grating 1 and 2 are etched such that λ_1 and λ_2 each only ‘sees’ one grating.

The same effects as described in Figure 5.6 would be generated if the gratings were designed to respond to different polarisations instead of different wavelengths. Use of crossed gratings conveniently allows the linear combination of two different diversity phase filter functions. In the next section some simulated examples of linearly combining diversity phase terms will be presented. This will be done to determine whether the new PTF’s created will give better performance than either diversity term would achieve individually.

5.2.2.2 PTF Engineering By Linear Combination Of Diversity Phase Terms.

In Section 5.2.1 it was shown that a defocus filter function produces a PTF with very few zero crossings (3 including the origin) compared to spherical aberration (7 in total) but with lower slope around the zero regions thus giving a slower recovery in sensitivity compared to the spherical aberration filter PTF. Here it shall be considered whether a combination of these two diversities can lead to a more ideal PTF with few zero crossings and high slopes wherever zeros are unavoidable.

In Figure 5.7(a), 4 sets of results are plotted which demonstrate the effect of adding defocus to the spherical aberration diversity function. The amount of spherical aberration is gradually decreased, while the amount of defocus added is set at a constant 1 wave of PV error thus the relative strength of the defocus term in the combination filter function increases. The results show that the addition of the defocus term spreads the PTF out towards the edge of the pupil (shown by the pair of solid vertical lines). The overall number of zero crossings, compared to the spherical aberration PTF on its own, decreases as the relative strength of the defocus term increases. Simultaneously the PV amplitude of the combination PTF is reduced. Figure 5.7(a) shows that by the time the spherical aberration term is reduced to 1/3 its original value (0.3λ) the combination PTF contains an extended region of almost zero amplitude. This has therefore become unsuitable, for the same reason as the example in Figure 5.4; too much information would be lost due to the low amplitude region. From these results it was concluded that this combination of defocus and spherical aberration does not produce any significant benefit. While it was possible to reduce the number of zero crossings within the pupil the reduction in overall PTF PV amplitude negated this benefit through loss of sensitivity.

Figure 5.7(b), in which the spherical aberration contribution to the diversity function is subtracted from the defocus term, yields a much better result. This plot shows that, for each of the 4 cases tested, the combination PTF has a sharper peak at the origin than the defocus-alone PTF and less zero-crossings within the area of the pupil than the spherical-aberration-alone PTF. Performing the combination in this manner has

successfully managed to produce PTF's which possess the best qualities of each of their component parts.

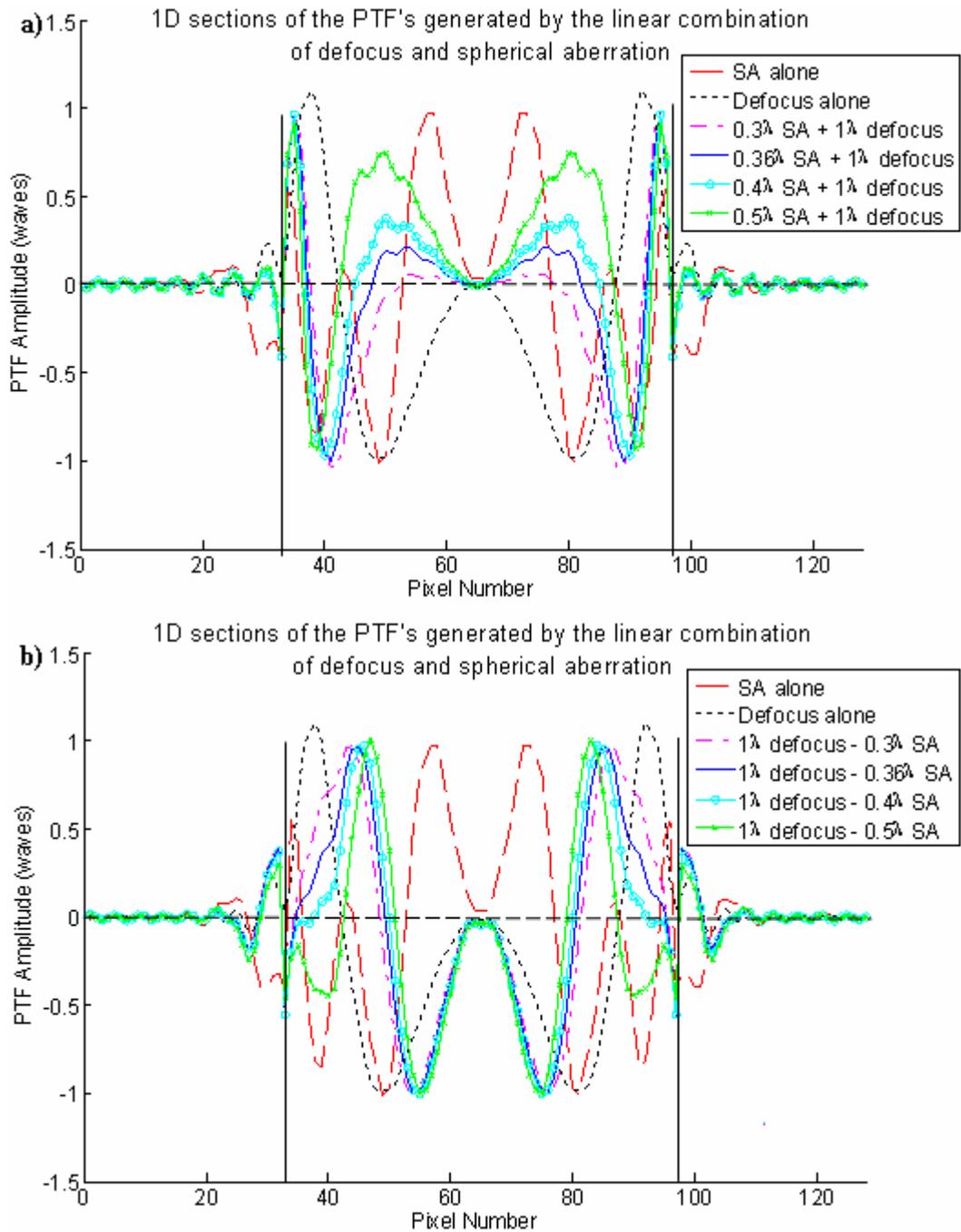


Figure 5.7 Cross sections of the simulated PTF's for linear combinations of defocus and spherical aberration, combined by (a) addition and (b) subtraction.

The results shown in Figure 5.7 would suggest that, due to the narrower peak at the origin, the subtraction-combination of defocus and spherical aberration ($\varphi_d = Z_2^0 - Z_4^0$) should be more sensitive to lower order aberration modes than the addition-combination equivalent ($\varphi_d = Z_2^0 + Z_4^0$). To test this hypothesis a series of Matlab simulations were conducted to compare their relative sensitivities when reconstructing low order modes. Figure 5.8 shows one example of this, giving cross sections of the original and retrieved phase profiles for an astigmatic wavefront with the $1\lambda \cdot Z_2^0 \pm 0.4\lambda \cdot Z_4^0$ diversity functions. This plot shows that, despite being noisy, the subtraction-combination has given a slightly better fit to the original phase than the addition-combination.

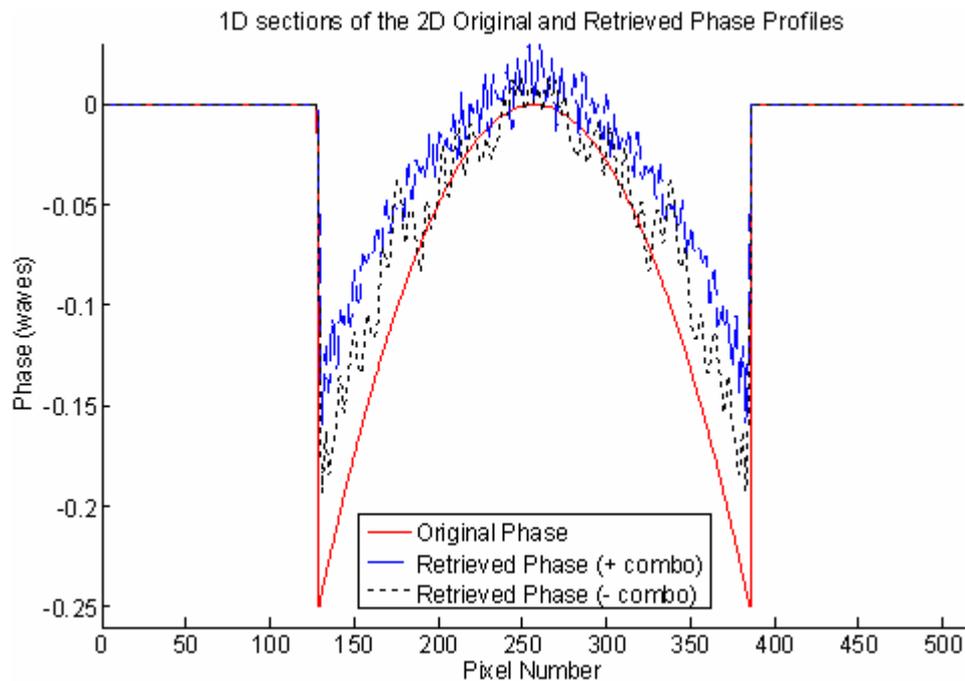


Figure 5.8 Cross sections of the original and retrieved phase profiles for an astigmatic input wavefront with two different diversity functions created by adding ('+ combo') or subtracting ('- combo') a defocus and a spherical aberration diversity phase term.

Other combinations of diversity terms were tested and in many cases it was found that a new PTF could be generated to combine the best properties of its constituent diversity phases. Whether the best result is obtained by the addition or subtraction-combination depends on the form of the original PTF functions and the relative strengths of the

diversity phase terms. In some cases both methods of combination produced equally good PTF's (in terms of slope and number of zero-crossings). The decision as to which is 'best' would have to be assessed by testing the sensitivity of each combination to particular examples of aberrated wavefronts.

Interestingly it was also found that polynomials which individually are completely unsuitable for GPD due to extended zero regions (see for example Figure 5.4) could be combined with a second diversity to produce a useful function. One example of this is given in Figure 5.9 which shows the PTF's of defocus (Z_2^0 - as in Figure 5.1), the higher order Zernike Z_6^6 (as in Figure 5.4) , and the resulting PTF of their combination in plots 5.9(a)-(c) respectively.

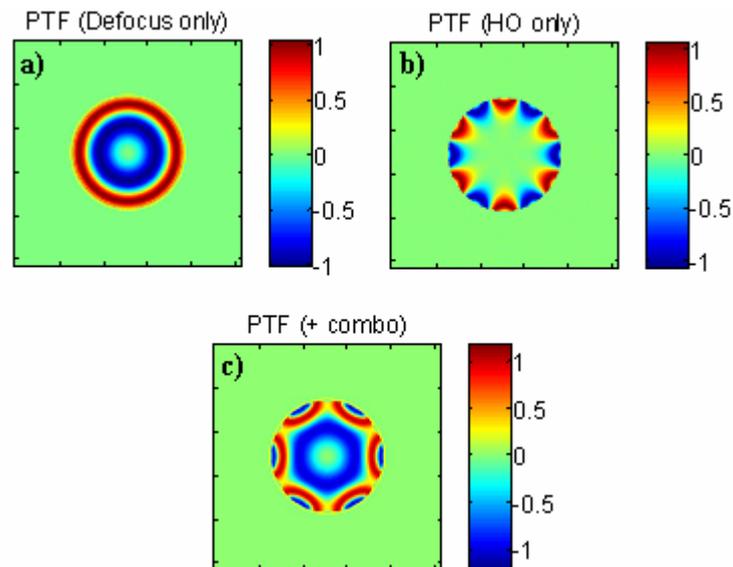


Figure 5.9 2D plots of the generated PTF's for (a) defocus only, (b) higher order aberration Z_6^6 only and (c) the linear combination of the diversity phases used to create the PTF's in plots (a) and (b) [scales are in waves]

In conclusion, it was found that optimising the PTF through linear combination of diversity terms (by addition or subtraction) is a viable option. In Chapter 3 it was concluded that there exists some potential for optimisation of the GPD sensor for particular applications through careful selection of the diversity function. PTF

engineering presents a method for performing this optimisation. By investigating the effects on the PTF of combining different diversity phases, varying their relative strengths and the method of combination the user will be able to select a diversity function which provides the best sensitivity for their given application. The range of possibilities is huge, by considering combination as well as single phase diversity functions the likelihood of finding an optimum function is greatly increased.

In all the PTF plots presented thus far it has been shown that, in accordance with the theory, the PTF will go to zero at the origin. The result of this is poorer sampling of low spatial frequencies than that of the higher spatial frequencies, this is accentuated by the curved shape of the PTF around this point. A PTF with a sharp 'notch' shaped region at the origin would perhaps help to minimise the problems encountered at low spatial frequencies. In search of a function which would produce such a PTF the Phase Contrast Microscope (PCM) was studied. The results of this investigation are the subject of the next section. Whether or not PCM may be used to help to optimise the GPD sensor at low spatial frequencies will also be covered in later sections.

5.2.3. The Phase Contrast Microscope

The Phase Contrast Microscope (PCM) was first proposed by Frits Zernike in 1934, this discovery earned him a Nobel prize which was awarded in 1953. Since then it has become a valuable tool in a wide variety of fields; principally medical and biological. PCM optically converts minute phase changes induced by transparent specimens into amplitude variations in the resulting image. This allows many delicate biological specimens to be imaged in their natural state, whereas previously these specimens were killed, fixed and stained to be studied. As a result the dynamics of ongoing biological processes and cultures can be observed using PCM. Larger specimens, due to scattering and diffraction around their edges, are also easily visualized using this technique. The sensitivity of PCM to small changes in refractive index of the sample also makes this useful in industrial fields such as crystallography and studies of polymers [3].

Figure 5.10 shows images of living human glial brain tissue (grown in culture) captured using PCM and Bright Field Microscopy (BFM). In the BFM image (Figure 5.10(a)) only areas with high refractive index (the membrane and nuclei for example) have been imaged clearly whereas in the PCM image (Figure 5.10(b)) the internal features and cellular attachments are imaged with good contrast. This example demonstrates the usefulness of PCM and was taken from the Nikon website [4].

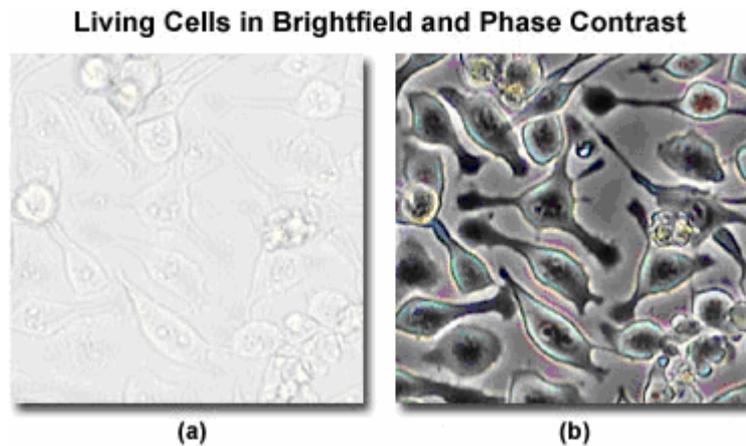


Figure 5.10 Images of living human brain tissue captured using (a) Bright field microscopy and (b) Phase contrast microscopy. Taken from the Nikon website [4].

PCM is employed as an imaging method for transparent specimens which, although they do not create amplitude modulation on the transmitted wave, do alter the phase of the input wavefront. Even samples which are physically very small or optically thin will have some effect on the phase of the input wave due to variations in thickness or refractive index (or both). A wavefront encountering such a sample is separated into 2 parts; a surround wave (which is unaffected or passes around the sample) and a diffracted wave (which has been phase modulated by the transmission). The Optical Path Difference (OPD) which is imposed on the diffracted wave is, for most biological specimens, very small. For example, single cells grown in tissue culture will yield an OPD of approximately $\lambda/4$ (for green light) [3]. With a phase shift between the surround (also known as the zeroth-order wave) and the diffracted wave of only $\lambda/4$ when they recombine in the image plane these waves will be $\pi/2$ out-of-phase. An additional phase shift of $\pi/2$, added before these waves are recombined, would allow

them to interfere at the image plane. This interference would create a change in contrast and effectively turn the resulting image into an amplitude object that can be captured on camera.

In his original description of the PCM Zernike proposed that the phase shift be applied to only the very centre of the objective's rear focal plane. In practice PCM is implemented in microscopes built for BFM. To convert between BFM operation and PCM a phase plate is placed at (or close to) the back focal plane of the objective. These phase plates are designed for use with particular condenser annuli, which are used to produce conical illumination of the specimen in BFM. The condenser annulus is focussed onto the phase plate and the region of the phase plate this illumination covers is called the conjugate area. The conjugate area is either raised or sunk with respect to the rest of the phase plate such that the etch depth will apply the required phase delay to the beam for PCM imaging. In positive contrast PCM the phase delay is $+\pi/2$ and negative contrast PCM applies a delay of $-\pi/2$. Figure 5.11 shows the difference between positive and negative phase contrast filters, and shows their annular design which has been chosen to match the ring illumination of the annular condenser.

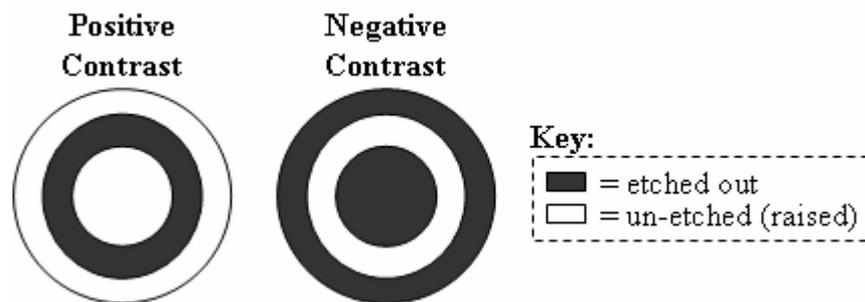


Figure 5.11 Schematic showing the physical difference between positive and negative contrast phase plates for PCM.

In PCM the input wavefront is passed through a distorting medium (the sample) and a solution is obtained through the addition of an extra phase term. In a sense therefore this can be thought of as a technique akin to phase diversity. With this in mind the symmetry of a PCM phase shift function was studied to determine whether it would meet the requirements for GPD wavefront sensing [5]. The phase shift can be written as

$\{2\pi/\lambda\} \cdot \varphi_{PCM}$, where φ_{PCM} is the shift applied. The total phase shift required is $\pi/2$ (or $\lambda/4$). If, as in general PD, one aberrated and one aberrated image is to be compared then this phase shift would be applied to just one of the images. In this case we find that the phase shift function would be unsuitable for GPD since the real part of the phase function ($\exp[-i.(\pi/2)]$) is zero ($\text{Cos}[\pi/2]=0$, $\text{Sin}[\pi/2]=1$). This violates the necessary condition that any function suitable for GPD must be complex. However, if as in Curvature Sensing (CS) the diversity is applied in equal and opposite amounts to generate a pair of phase diverse images then the phase function becomes $\exp[\pm i.(\pi/4)]$. In this case the filter function is complex and is suitable for GPD wavefront sensing. It is this form of the filter function which has been used in the following simulations.

5.2.3.1 PCM – Its PTF and Performance as a Diversity Function

The next step was to test the accuracy of PCM as a phase diversity function for GPD. A pupil plane system was created in simulation and the PCM diversity ($\exp[\pm i.(\pi/4)]$) was applied to the central pixel of the FT of the wavefront ($\Psi(\xi)$). It was found, through simulation, that increasing the size of the phase shifted region decreased the accuracy of the phase reconstruction. The one pixel shift used was chosen to demonstrate the ideal case.

Figure 5.12 shows the PTF calculated for the PCM diversity function. As this plot shows, since the phase shifted area is so small (1 pixel), the PTF drops very sharply to zero at this point, and this point only. Therefore the amount of low spatial frequency information lost should be minimised in this case, compared to those studied previously in Section 5.2.1. Figure 5.12 also shows that, apart from the region very close to the origin, the PTF has a ‘top-hat’ structure that should mean that it samples evenly across the range of spatial frequencies. By the definitions given earlier this PTF should give a good performance when used in the SAE algorithm. A number of simulations were conducted to test this, and the results of some of them are given in Figure 5.13 to 5.15.

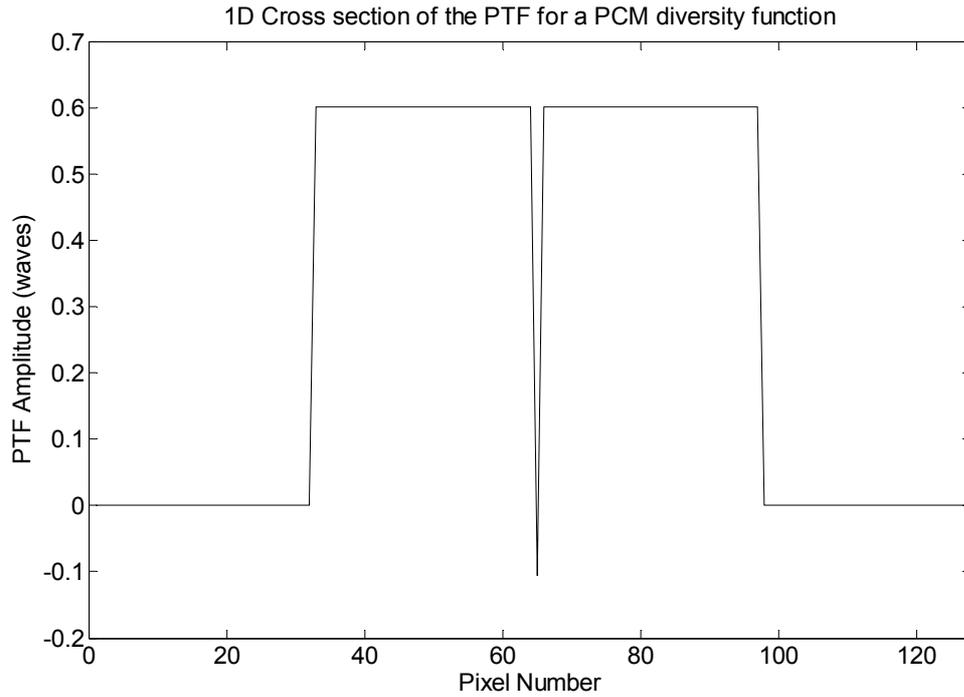


Figure 5.12 A 1D cross section of the PTF when the diversity function used is a simple constant phase shift at the origin, simulating the operation of the PCM.

Three different input wavefronts were tested, and each used the same PCM diversity function (whose PTF is shown in Figure 5.12). The three examples chosen show the performance of the PCM GPD function with defocus (Z_2^0 - an example of low frequency structure), a higher order aberration (Z_7^{-1} - a high frequency example), and the same random wavefront used in Chapter 4 in Figures 4.9 and 4.13 (created using a mixture of Z_6^2, Z_2^0, Z_3^3). Note that in all of these simulations the same soft edged pupil and threshold (regularising constant) was used in the SAE algorithms (see Chapter 4 for details).

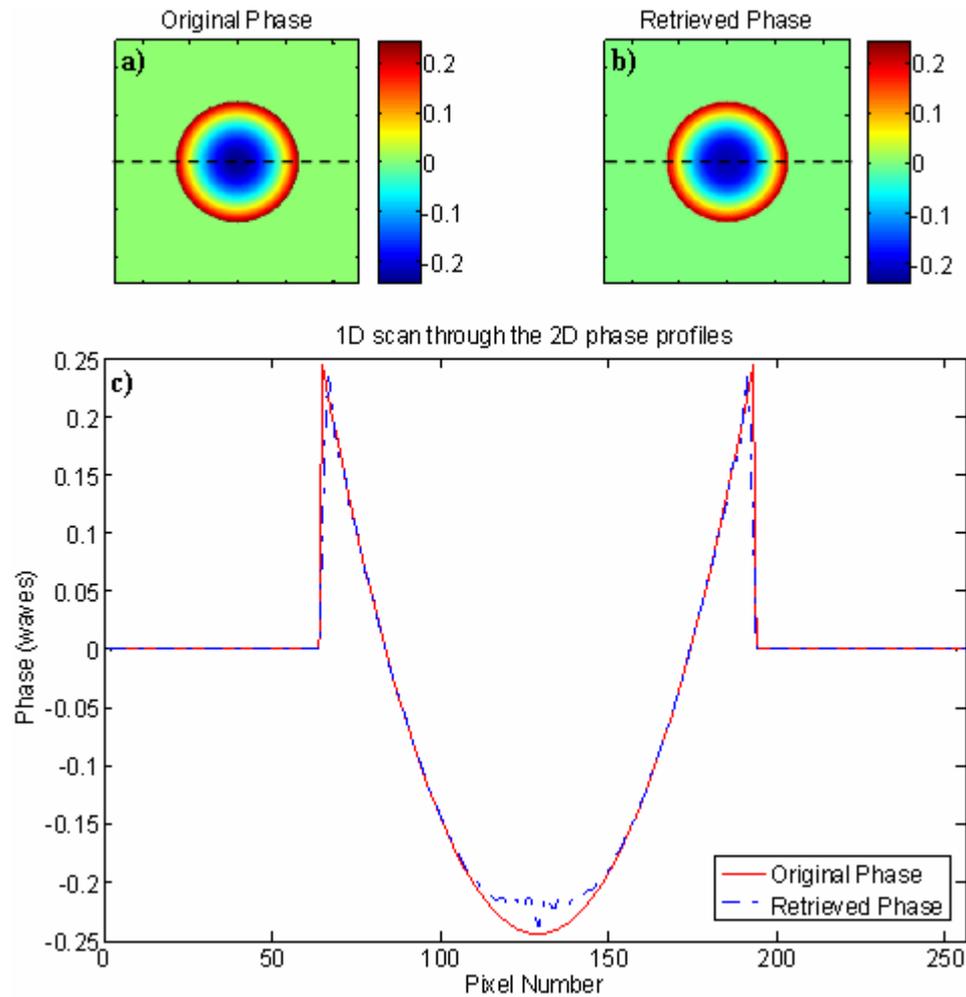


Figure 5.13 The original (a) and retrieved (b) phase profiles generated by the SAE algorithm for a defocused wavefront with the PCM diversity filter function [scales in waves]. A 1D section of these profiles (taken along the black dotted line in plots (a) and (b)) is shown in (c).

Figure 5.13 shows the SAE retrieved results for a defocused wavefront with PCM diversity. The Error Metric (EM) for this reconstruction (calculated as described in Chapter 4 Section 4.6) is 0.998. As this plot clearly shows the fit of the retrieved phase to the original phase profile is very good, with the only difference being the fit at the central peak. By the nature of the PTF for this diversity function (see Figure 5.11) problems at the origin (in Fourier space, therefore corresponding to low spatial frequencies) are anticipated. A low frequency structure such as defocus will be affected by this and therefore the deterioration of the fit to the peak of this function is not unexpected. By the same argument this diversity function should perform very well

with high spatial frequency information. An example of this is studied in the simulation shown in Figure 5.14.

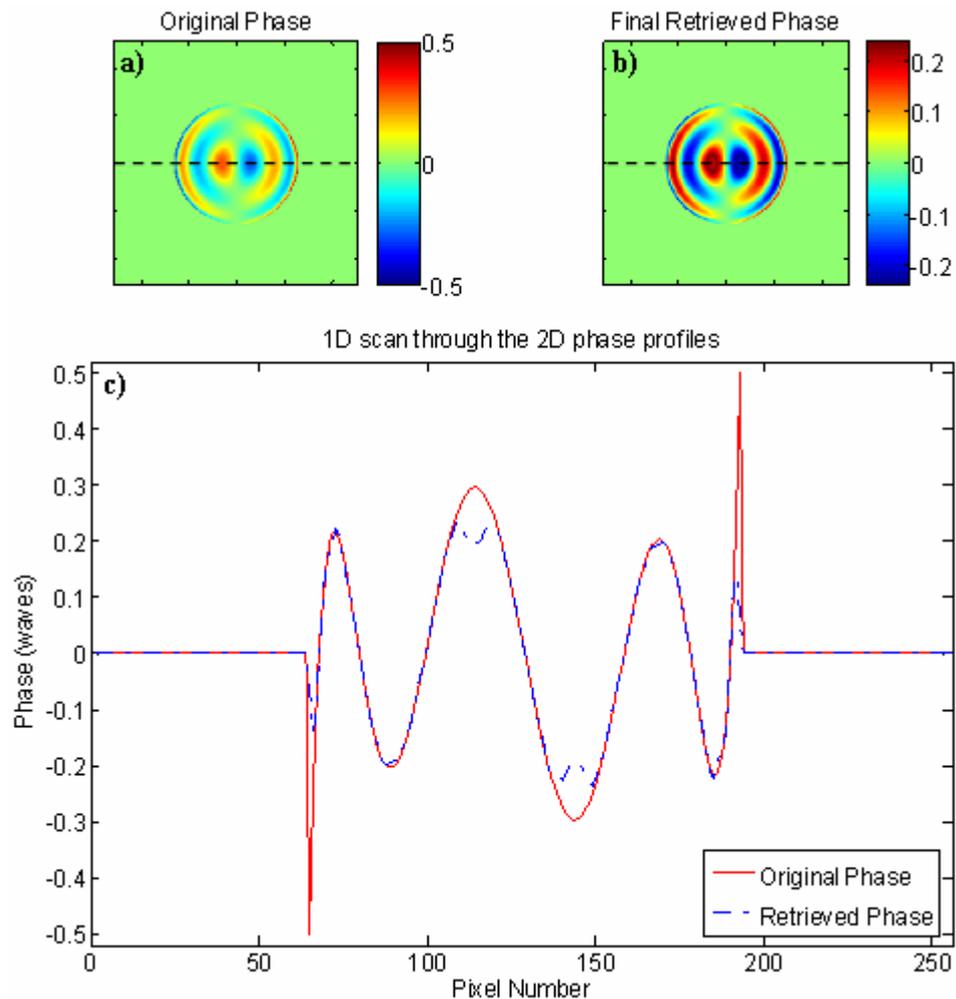


Figure 5.14 The original (a) and retrieved (b) phase profiles generated by the SAE algorithm for a Z_7^{-1} input wavefront with the PCM diversity filter function [scales in waves]. A 1D section of these profiles (taken along the black dotted line in plots (a) and (b)) is shown in (c).

Figure 5.14 demonstrates that the PCM diversity has retrieved the high frequency test wavefront with high accuracy. The EM for this reconstruction is 0.992. This example shows that the PCM function deals well with high frequency information with a small amount of error on the two largest peaks. These simulations are representative of a large number of similar tests which all followed the same trend; that low frequency and

high frequency structures were both reconstructed to high accuracies with small errors at the peaks.

Finally, since in any real situation it is more likely that the input wavefront is a made of a combination of error terms a wavefront, created using a linear combination of Z_6^2, Z_2^0, Z_3^3 with random weights, was used. This example is the same wavefront used in the hard edge and soft edge pupil simulations from Chapter 4.

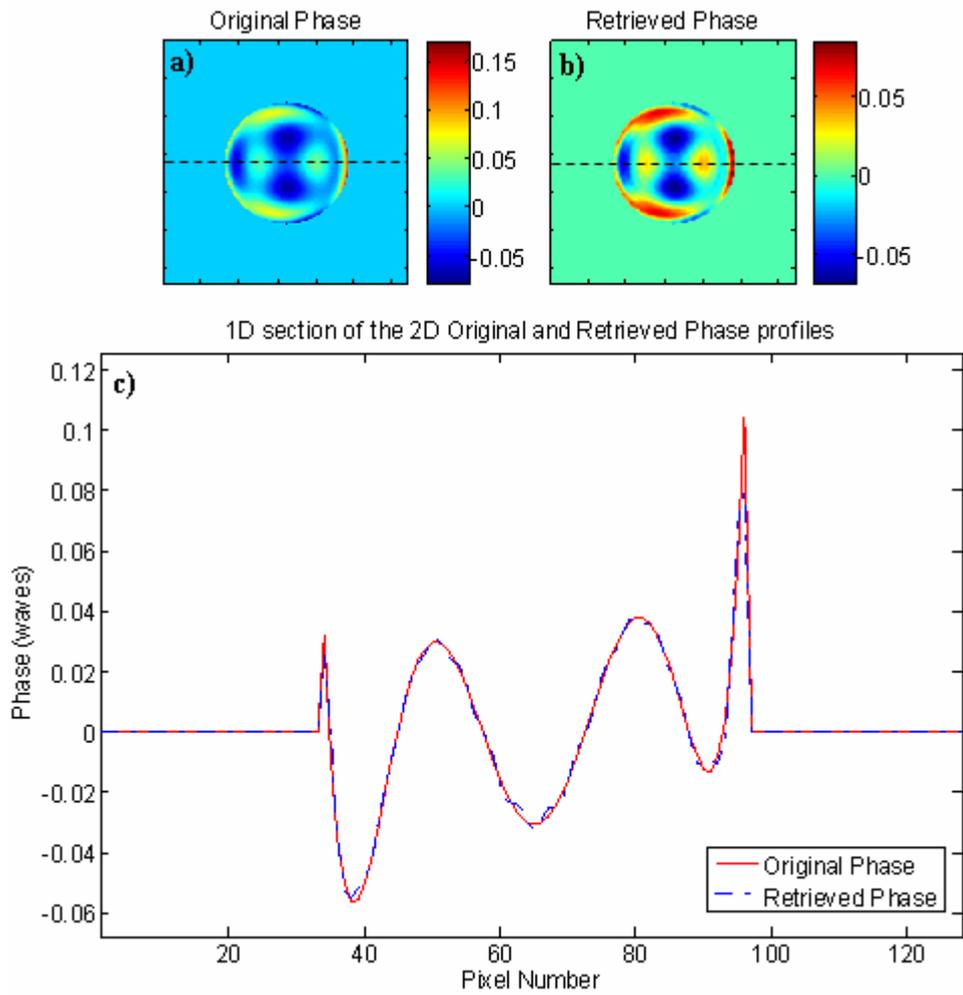


Figure 5.15 The original (a) and retrieved (b) phase profiles generated by the SAE algorithm for a random input wavefront with the PCM diversity filter function [scales in waves]. A 1D section of these profiles (taken along the black dotted line in plots (a) and (b)) is shown in (c).

Figure 5.15 has shown that the PCM diversity function has produced an excellent fit to the original phase profile. In this case the EM is 0.9996 (as compared to 0.994 for the hard edge defocus diversity example and 0.9998 for the soft edge case – see Chapter 4 for details). For comparison purposes the PV amplitude of this random example was kept the same as the Chapter 4 simulations ($PV = 0.167\lambda$). This is a much smaller phase angle than the other examples in this section (Figure 5.13 = 0.5λ PV and Figure 5.14 = 1λ PV) which explains why there are no errors in the peak amplitudes in this case.

The results presented in Figures 5.13-15 have demonstrated that a diversity function which creates a phase shift of the central pixel by $\pi/2$, thus simulating the operation of a PCM, is highly suitable for use in GPD wavefront sensing. The results show a good fit to the original data with small errors occurring when the FT of the input phase is concentrated at the origin. The form of the PTF (shown in Figure 5.12) would suggest that this function would perform better (i.e. lose less low frequency information) than say defocus. The results in Figure 5.15 show that this is the case when comparing it to the hard edge example (from Chapter 4) and the EM is only slightly less than the soft edge example (also from Chapter 4).

In Section 5.2.2.2 it was determined that there was potentially significant advantages to be gained by using a linear combination of Zernike filter functions instead of using these singly. Finally the combination of the PCM diversity function with a Zernike filter will be studied to see if the sharp notch-like PCM PTF will improve the sensitivity of the combination filter to low spatial frequency aberration modes.

5.2.4 PTF Engineering using PCM in Linear Combination with a Zernike Filter

In the previous section it was shown that a diversity function based on the concept of the PCM performed very well in simulation, providing retrieved results with good fit to the original phase profiles. A final attempt at ‘PTF Engineering’ was attempted by creating a combination of the defocus and PCM filter function to see if this would give better performance than either function individually.

The combination diversity phase φ_d was created by adding a $\lambda/8$ phase shift on the central pixel of a defocus phase, a cross section of this is shown in Figure 5.16. The intensity images in the ± 1 orders are generated by adding $\pm\varphi_d$, therefore there will be the full $\lambda/4$ phase shift between these images required for PCM.

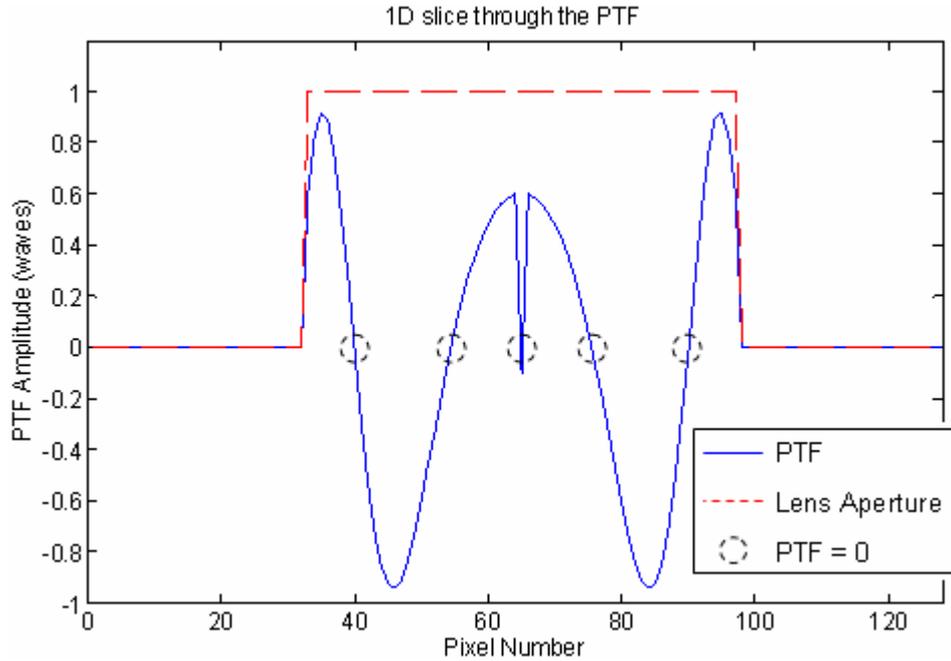


Figure 5.16 A 1D cross section of the 2D PTF for the combination of defocus and PCM diversity phase.

The PTF shown in Figure 5.16, due to the PCM phase shift component, has a sharp drop to zero at the origin compared to the defocus only PTF (see Figure 5.1). It is unclear from this plot alone whether this will lead to an increase in accuracy compared to the retrieved phase solutions from defocus and PCM alone. This PTF achieves the aim of having the minimum number of pixels affected by the zero point (at the origin) and a sharp recovery about this point, but in effect only a few pixels have been altered by the addition of the PCM term which may not prove to be enough to make a positive impact on the results. To study this further a series of simulations were conducted to calculate the EM's of the retrieved phases calculated for 2 different input wavefronts using defocus and PCM singly and in combination. The two input phase profiles chosen were defocus (Z_2^0) and the higher order aberration Z_{10}^4 to compare the performance of the

diversity functions with a low and a high frequency example. Table 5.1 shows the EM's of the SAE retrieved phase profiles for each example:

Diversity Function	EM's	
	Low Freq. example (Z_2^0)	High Freq. example (Z_{10}^4)
Defocus	0.99945	0.99964
PCM	0.99939	0.99966
Defocus + PCM	0.99918	0.99957

Table 5.1 EM's for the SAE retrieved phase profiles for several simulations to compare the performance of defocus and PCM used both singly and in combination.

The results in Table 5.1 clearly show that, in both examples, the combination of PCM and defocus did not give greater accuracy than either of these diversity functions when used separately. There is not a great deal of difference between any of the results, they are all of an acceptably good standard, however these results suggest that there is no benefit to be gained from their combination. Several sets of simulations were conducted to combine PCM with another Zernike diversity function and in each case the accuracy of the combination solution was lower.

5.2.5 Studying the PTF and PTF Engineering: Conclusions

In this section it has been shown that merely satisfying the conditions for use in a GPD null sensor does not guarantee good performance as a diversity function. The question of what constitutes a good diversity function has been approached through study of the PTF that it generates and an attempt to identify the qualities in a PTF that will lead to more accurate reconstructions has been presented. It has also been seen that the well known PCM technique performs particularly well as a GPD filter function.

In a practical situation the user will have complete control of the diversity function chosen to create their data. Analysing the PTF of the diversity function is just one tool to aid the user in making an informed choice about which diversity function is most

suited to their application. An investigation was conducted to find whether combinations of diversity functions could be used to combine the useful traits of each of the functions to maximise the SAE performance both in terms of accuracy and stability. Linear combination of the component filter functions was chosen due to the relative simplicity of applying this in a real situation using crossed gratings. In each case tested it was found that at least one of the combination methods (addition or subtraction) resulted in a new PTF which combined the best qualities of each of its component diversity phase PTF's.

There may be other ways in which several diversity functions can be combined to optimise the performance of the GPD sensor. However, the user will have to decide whether the improvement gained warrants the added complexity of design and implementation.

5.3 Extending the SAE

Thus far the development of GPD and the SAE algorithm has been covered in detail and shown to work well with a variety of diversity functions and input wavefronts. Up to this point this study has been confined to continuous wavefronts and wavefronts with uniform illumination. One of the main science drivers behind this project was the need for a wavefront sensor which would provide accurate wavefront sensing of scintillated and discontinuous wavefronts. The reasons for this were discussed in detail in Chapter 2 and will be revisited briefly in Chapter 6 when several possible applications of the GPD SAE sensor system will be considered. The next step is to extend the SAE analysis to investigate its behaviour in the presence of non-uniform illumination. This will be covered in the next section and followed by an example of the SAE's performance with a discontinuous wavefront in which a series of simulations of a tessellated segmented telescope mirror will be presented.

5.3.1 Use of the SAE with Non-Uniform Illumination

Scintillation modelling is a complex topic in itself and one which could easily fill a thesis on its own, such a study is outwith the scope of this project. To begin to study the effects of non-uniform illumination on the accuracy of the SAE solution this problem will be reduced to its simplest form. The non-uniform illumination will be represented as comprising a d.c component modulated by an a.c ripple. If the SAE proves incapable of dealing with such a simplified version, then the chances of it handling full scintillation are small where, in addition to modulation of the intensity, information has been lost.

Firstly, note that the analysis presented thus far has not made any assumptions that preclude the use of this technique with non-uniform illumination. A scintillation pattern is still a real function, and as such should be recognised by the GPD sensor. In the past the Defocus-only Phase Diversity (DPD) sensor has been shown to work with scintillated wavefronts despite the assumptions made by the Green's function solution (see Chapter 2 for details) [6]. There are two reasons why this is sometimes the case. Firstly, if the scintillation is modest enough then only a small amount of information is lost and the algorithm may be able to interpolate to fill in the gaps in the solution. Secondly, if the distance between the planes in which the defocused images are sampled is small enough then to a rough approximation the dark spots in the intensity images will match and the difference of these images will create a useable signal. This is because over very short distances the scintillation pattern will not have sufficient propagation distance to evolve. The latter description only applies to the DPD sensor since it relies on the fact that one intensity image is effectively propagated with respect to the other. The first exception, that the DPD sensor can handle moderate scintillation, would conceivably be true regardless of the diversity function used. It would be more desirable to have a reconstruction method which is capable of dealing with significant levels of scintillation and which does not rely on defocus being used as the diversity. Further analysis of the SAE is required to test the accuracy of the solution as the non-uniformity of the input illumination is increased.

5.3.1.1 Extending the SAE Equations for Non-Uniform Illumination

Firstly, recall Equation (4.10) from the analysis in Chapter 4; this equation is reproduced below:

$$\frac{d(r)}{2i \cdot h(r)} = -\tilde{i}_m \left\{ \int d\xi' A(\xi') R(-\xi') e^{-ir\xi'} + \int d\xi A(\xi) R(\xi) e^{-ir\xi} \right\} + \tilde{r}_e \left\{ \int d\xi A(\xi) I(\xi) e^{-ir\xi} + \int d\xi' A(\xi') I(-\xi') e^{-ir\xi'} \right\} , \quad (4.10)$$

where $d(r)$ is the difference of the intensity images in the ± 1 diffraction orders, $h(r)$ is the real part of the input wavefront and $A(\xi)$ (the Anti-Hermitian part) is the FT of the imaginary part of the input wavefront $i.a(r)$. The real and imaginary parts of the complex filter function are given by $i_m(r)$ and $r_e(r)$ respectively and their FT's are $I(\xi)$ and $R(\xi)$. The tildes in $\tilde{i}_m(r)$ and $\tilde{r}_e(r)$ represent that these are the integrals over r of $i_m(r)$ and $r_e(r)$ (e.g. $\tilde{i}_m(r) = \int i_m(r) dr$).

In Chapter 4, using the Small Angle Approximation (SAA), the following approximations were given:

$$h(r) = \sqrt{I_0(r)} \text{ and } a(r) = \sqrt{I_0(r)} \cdot \varphi(r) . \quad (4.4)$$

It is the $\sqrt{I_0(r)}$ term which will be allowed to become 'non-uniform' by representing it as an a.c. modulated d.c. function such that:

$$\sqrt{I_0} = dc + ac . \quad (5.2)$$

Substituting this for $h(r)$ in Equation (4.10) and recalling that $A(\xi) = i\sqrt{I_0}\Phi(\xi)$ (where $\Phi(\xi)$ is the FT of the input phase) the following is obtained:

$$\frac{d(r)}{4(dc+ac)^2} = \int \Phi(\xi) [\tilde{i}_m(r).R(\xi) - \tilde{r}_e(r).I(\xi)] e^{-ir\xi} d\xi \quad . \quad (5.3)$$

Substituting $T(\xi)$ for the PTF term (given in the square brackets in Equation (5.3)) and neglecting the ac^2 term as being very small, Equation (5.3) may then be rearranged to give:

$$\frac{d(r)}{4} = \underbrace{(dc)^2 \mathfrak{I}[\Phi(\xi).T(\xi)]}_{\text{ORIGINAL TERM}} + \underbrace{2(ac.dc) \mathfrak{I}[\Phi(\xi).T(\xi)]}_{\text{TERM ADDED BY THE 'SCINTILLATION'}} \quad . \quad (5.4)$$

Equation (5.4) shows that, when the input illumination is non-uniform, the error signal $d(r)$ comprises 2 expressions. The first (labelled ‘original term’) is the same expression derived in Chapter 4 (see Equation (4.14)) for the difference image (where $I_0 = (dc)^2$). The second term is caused by the ac modulation of the input wavefront and (5.4) shows that the strength of this term will increase at twice the rate the a.c. modulation increases.

This analysis shows that the addition of non-uniform illumination will result in alteration of the difference image, as we would expect. The question this raises is whether the SAE is able to reconstruct an accurate input phase profile from the altered form of $d(r)$, and how does this relate to the strength of a.c?

A series of Matlab simulations were conducted to investigate this further. A wavefront was defined as shown in Equation (5.5); $P(r)$ is the hard edged pupil function, $(dc+ac)$ is the illumination function (previously written $\sqrt{I_0}$) and $\varphi(r)$ is the input phase function:

$$\psi(r) = P(r) \cdot (dc + ac) e^{-i.2\pi.\varphi(r)} \quad (5.5)$$

The illumination function was created by using a simple piston term (top hat) to define the d.c offset (with a height of 1) in addition to a Sine wave ac ripple. The overall PV height of the Sine wave was varied to alter the ‘strength’ of the non-uniformity of the input illumination.

Figure 5.17 shows the SAE retrieved phases calculated for the phase profile in 5.17(a) as the PV height of the Sine wave was increased. All scales shown are measured in waves of error. The number of cycles across the pupil was held constant at 2 for all of these simulations, except in 5.17(b) where the input illumination function contained only the d.c component. Figures 5.17(c) to (f) clearly show the deterioration of the retrieved phase solution as the size of the ac ripple is increased.

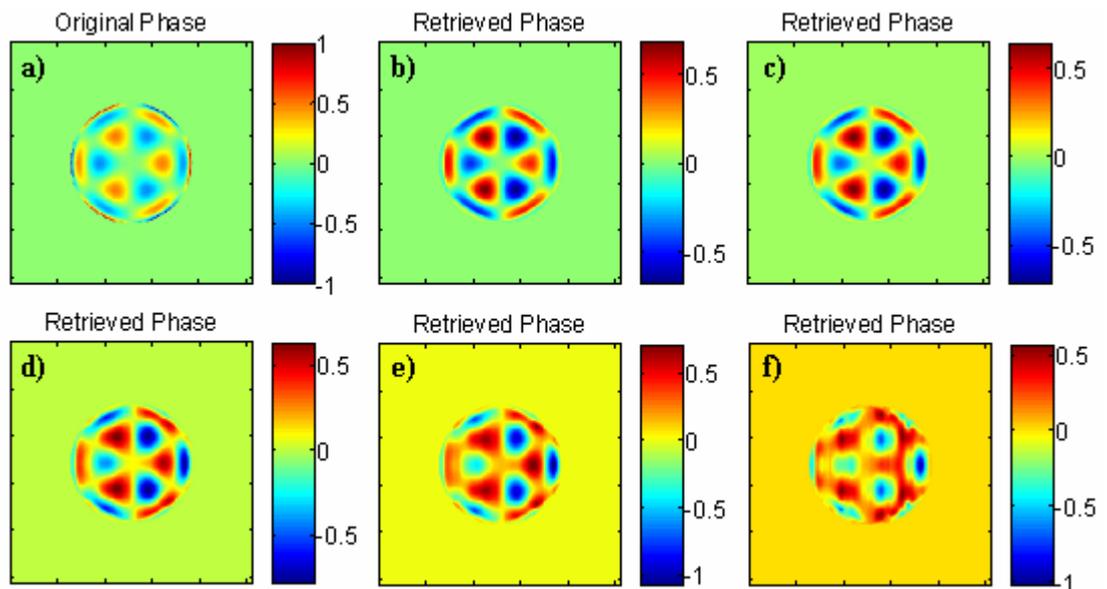


Figure 5.17 The SAE retrieved phase profiles for the original phase profile (a) as the PV height of the ac ripple on the illumination is varied. The results shown are for PV values of (b) zero, (c) 0.25, (d) 0.5, (e) 0.75, (f) 2.

The deterioration in the quality of the phase solution can be quantified in a number of ways. In this case the EM for each solution was calculated and is recorded in Table 5.2 in addition to the PV height of the Sine wave in each case. The PV heights of the Sine wave are a measure of amplitude and therefore strictly should correspond to a number of photons. Since this is a simulation assigning units to these PV values would involve an arbitrary choice of how many photons generate a given level of signal. Therefore in these examples the units are not expressed but should be taken to be $\sqrt{\text{photons}}$. In Chapter 4 an iterative refinement of the SAE solution was described and implemented. This iterative solution used the measured data as the ‘ground truth’ with which to refine the accuracy of the final solution. In a practical situation the user has only this data and the computed solution, therefore these are the tools which must be used to decide whether or not to trust the solution. One method is to take the phase solution and calculate from this what the data should look like (following the procedure described in Chapter 3 Figure 3.5). The Root Mean Square (RMS) error between the measured and calculated versions of the data can then be calculated and used to assess the quality of the solution. This method was used in these simulations and the RMS values for the solutions shown in Figure 5.17(b)-(f) are also given in Table 5.2.

Figure	PV height of the ac ripple (Sine wave) on a d.c term of height 1	EM	RMS
5.17(b)	0	0.886	0.519
5.17(c)	0.25	0.876	0.526
5.17(d)	0.5	0.841	0.630
5.17(e)	0.75	0.737	0.632
5.17(f)	2	0.691	0.696

Table 5.2 A table showing the RMS fits of the forward simulated data (calculated from the SAE retrieved phase) to the ‘measured’ data and the calculated EM’s of the results shown in Figure 5.17.

The results given in Table 5.2 clearly show that the quality of the SAE solution for the input phase decreases as the size of the ripple on the illumination increases. This rate of deterioration in the quality of the retrieved phase does not seem an unreasonable, however the plots shown in Figure 5.17 show that visually the solution bears little resemblance to the original phase when the PV of the ripple exceeds 0.5.

The addition of the a.c ripple has had an adverse effect on the quality of the solution but that is to be expected when dealing with scintillation. The results presented here show that the SAE algorithm has not immediately collapsed with the introduction of a ripple in the input illumination but has undergone a gradual decline in accuracy. This is encouraging and warrants further investigation.

5.3.1.2 Scintillation by Obscuration

A final simulation was conducted to study the effects of an obscuration in the input illumination, which is essentially a form of scintillation. In this example a small, circular, region of the input intensity is set to zero and the phase solution for this wavefront is computed using the SAE algorithm. This situation is analogous to measuring the surface profile of a test object with a partial obscuration caused for example by damage or marks on the surface (deliberate or otherwise). This obscuration can be seen in Figure 5.18(c) where the intensity image from the +1 diffraction order is shown. The position of this obscuration corresponds to a peak in the input phase and was chosen to see how the algorithm interprets the phase solution across an area where the error signal has been artificially set to zero.

The original and retrieved phase profiles are given in Figure 5.18(a) and (b) respectively. In 5.18(d) cross sections of the original and retrieved phase profiles, taken through the position of the obscuration (and shown by the black dotted lines in plots (a) and (b)), are compared. The retrieved phase was calculated both with and without the obscuration and both these results are included in 5.18(d).

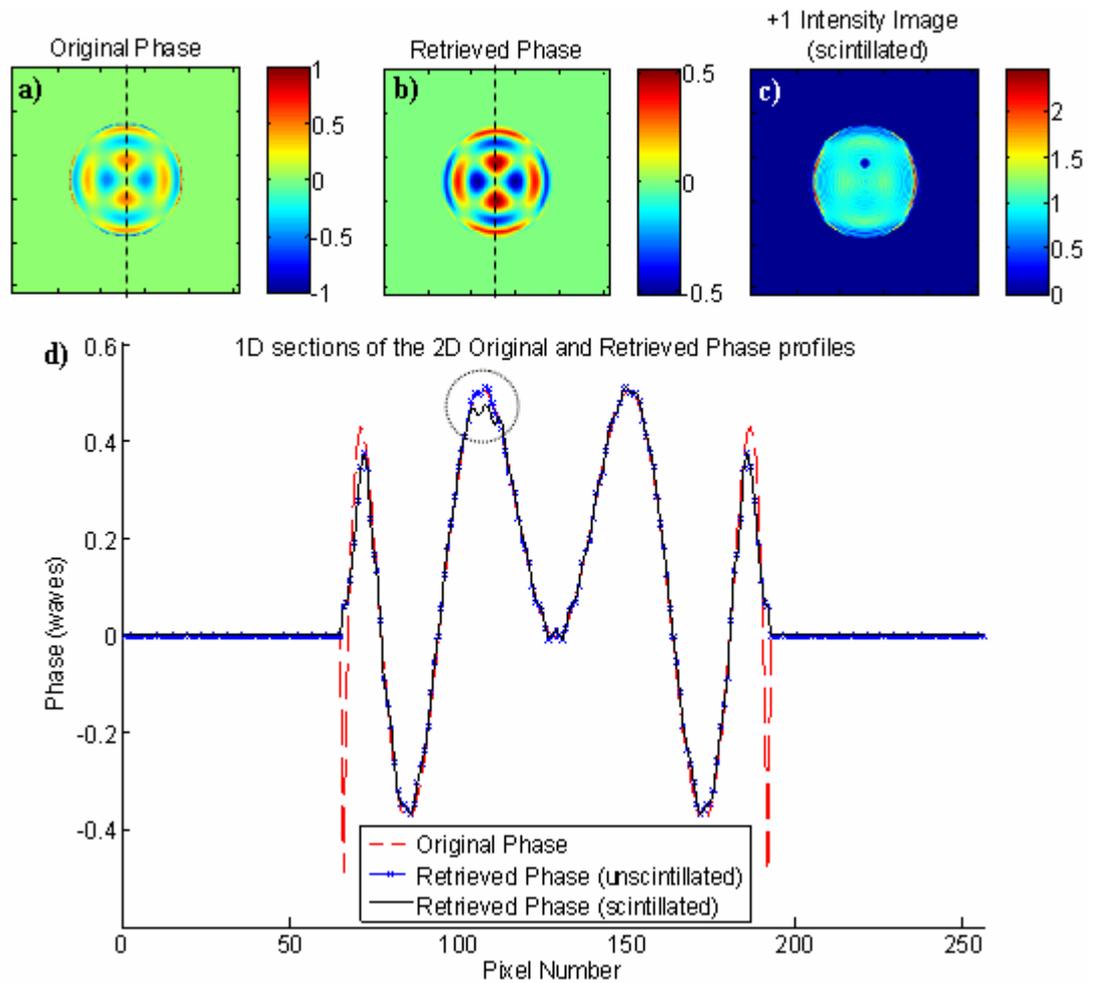


Figure 5.18 Original (a) and retrieved (b) phase profiles for a wavefront whose input illumination is partially obscured [scales are in waves]. (c) The +1 intensity image showing the position of the obscuration [scale is a measure of contrast]. Cross sections of the original phase profile and the retrieved phase calculated both with and without the obscuration are given in (d). The position of the obscuration is shown by the dotted circle. The dotted lines in (a) and (b) show the cross section position.

The results in Figure 5.18(d) show that where the obscuration has set the input intensity (and thus the error signal) to zero the SAE has retrieved a sensible solution in this region. The result of the obscuration (as seen in the dotted circle in 5.18(d)) is a slight error in the amplitude of the peak. Since there is no information in this region to change the phase solution ideally when the error signal drops to zero the SAE should fit a plane wave (or a slope) connecting points in the retrieved phase either side of the obscured region. In practice diffraction rings within the obscured region will lead to a ripple in

the reconstruction (as seen in 5.18(d)) instead of a plane wave. If however the size of the blur function is sufficient to bridge the gap in the signal (which is not the case here) we would expect a phase to be retrieved across that region. Though there will inevitably be errors in the amplitude of the computed solution it provides a good starting point for an iterative refinement of the solution (such as the one presented in Chapter 4).

The results shown in Figures 5.17 and 5.18 suggest that the SAE is capable of handling some level of scintillation. The next step is to test the performance of the GPD sensor and SAE algorithm with properly modelled scintillation. There was insufficient time during this project to do this, therefore this is held over for future work.

5.3.2 Use of the SAE with Discontinuous Wavefronts

In Chapter 4 the SAE algorithm was shown to work well with a variety of continuous wavefronts. In Chapter 2 several examples of applications which involve analysis of discontinuous wavefronts were discussed, these included metrology of integrated circuitry and phasing of segmented telescopes. In this section the accuracy of the SAE algorithm will be tested with a simulated segmented telescope pupil. The telescope pupil comprises tessellated hexagonal elements (the number of which can be varied) which each have an individual (randomly chosen) tilt and piston error. Accurate measurement of the tilt and piston in a real telescope is critical for co-phasing. Edge sensors are commonly used in real telescopes to judge when adjacent segments are accurately aligned. These have been designed to work in space-based telescopes, under extreme environmental conditions, as well as in ground based telescopes [7-11]. In a telescope the size of the Euro50 (50m primary diameter) it is anticipated that 3504 edge sensors would be required [12]. Therefore a single device like the GPD wavefront sensor, if capable of sensing phasing errors, could offer a cheaper alternative or allow the number of edge sensors to be reduced.

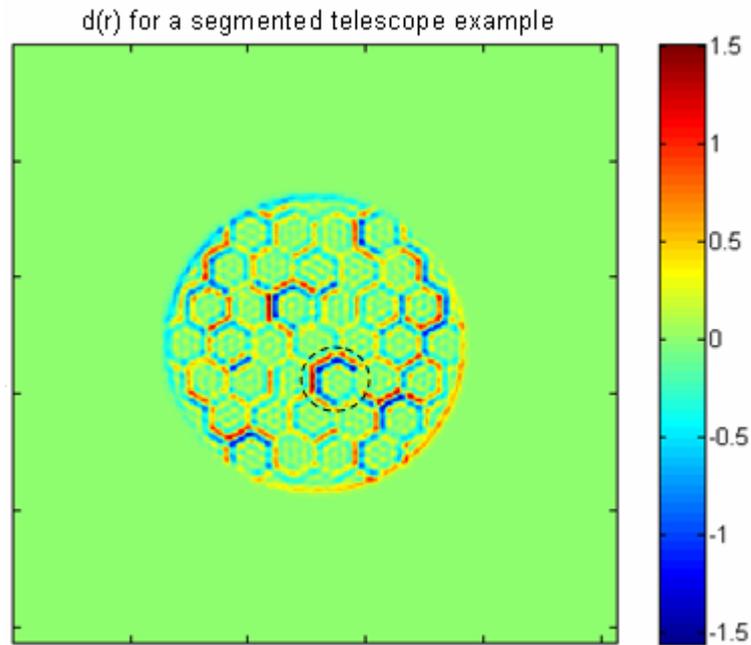


Figure 5.19 An example of the difference image $d(r)$ generated by a simulated segmented telescope with defocus used as the phase diversity. The dotted circle highlights interference at the segment edges [scale is a measure of contrast].

In previous chapters the GPD technique has been likened to Wavefront Shearing Interferometry (WSI). In WSI points on a wavefront are interfered by duplicating the wavefront and shifting the duplicate with respect to the original. Interference occurs only at points where the original and sheared duplicate overlap. When performing GPD wavefront sensing, points on the wavefront which overlap with the blur function will interfere with each other across this region. Applied to the segmented telescope this technique allows interference between adjacent mirror segments, if the width of the blur function is sufficient to bridge the gap between them. This can be seen visually in Figure 5.19, an example of the difference image $d(r)$ generated by a segmented input wavefront with defocus phase diversity. In this plot adjacent segments with relatively large misalignment have generated significant positive (shown in red) and negative (blue) signal. Interference between the edges of these segments has created the periodic signal which may be seen along the edge. An area where this can be seen is highlighted by the black dotted circle in Figure 5.19. These nodes are the same phenomena as the more easily recognised interference fringes, the greater they are in number the larger the phase error. This difference image could be used to directly correct the phasing errors

between segments. The user would first seek to minimise the number of nodes present at the edges by adjusting the relative tilt of the segments until there is only one fringe across each edge (i.e. the error signal is a constant on each edge). At this point only piston errors remain and can be corrected by moving the segments until the constant signal values across the edges are reduced to zero.

A series of Matlab simulations were conducted to assess the quality of the SAE solution when given a segmented input wavefront. Two examples of these are given in Figures 5.20 and 5.21.

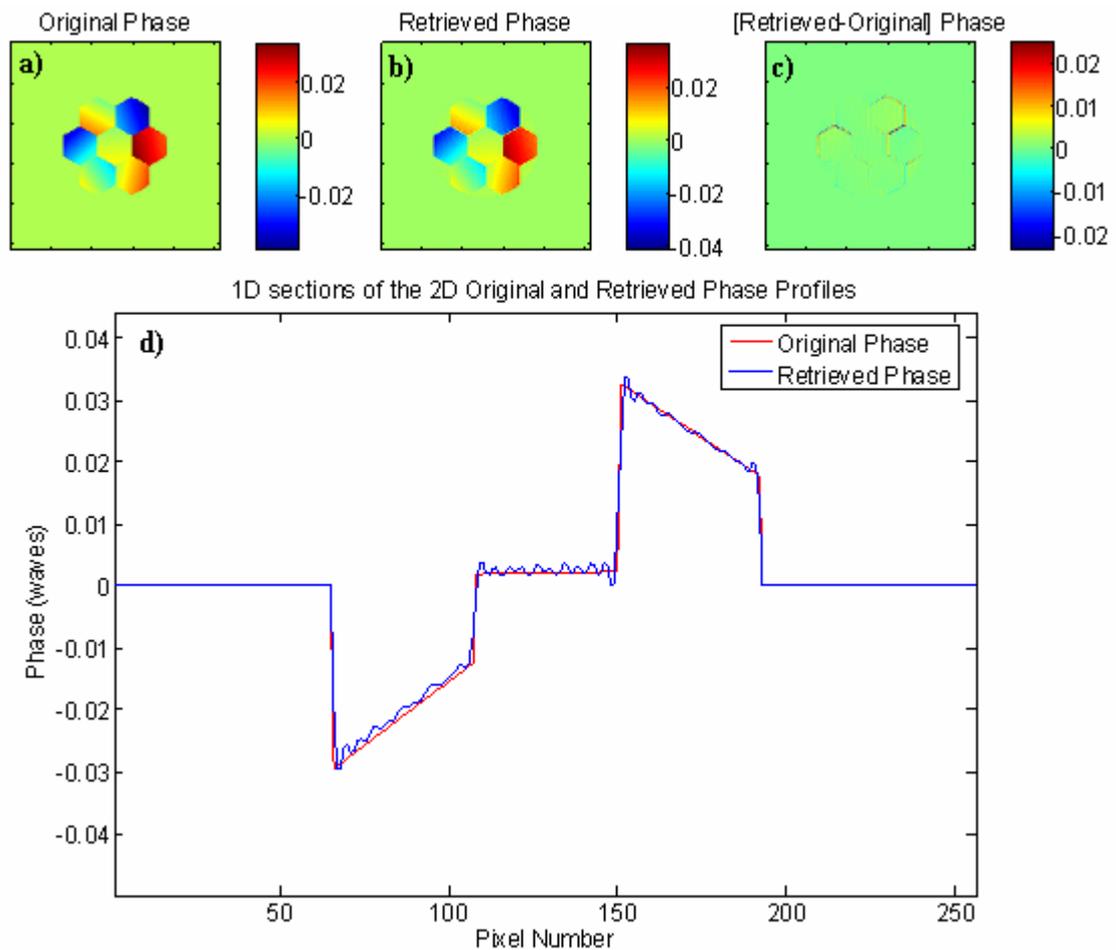


Figure 5.20 The SAE reconstruction (b) of segmented input wavefront (a). The difference between these is given in (c). A cross section of the original and retrieved phase profiles is plotted in (d) [all scales are in waves].

Figure 5.20 shows an input wavefront consisting of 7 hexagonal segments, each with tip/tilt and piston errors. As these plots show the accuracy of the solution is excellent, the EM calculated for this reconstruction is 0.99996. In 5.20(d) the 1D cross section of the 2D original and retrieved phase profiles is shown. From this it is clear to see that the largest error on the wavefront is caused by Gibb's effect. This is a well known phenomenon and occurs when the Fourier transform of a discontinuous function is calculated. Oscillations will occur in the region of the discontinuity [13-15]. This in itself, and the removal of these oscillations, is a topic of great research interest in a wide variety of applications such a Magnetic Resonance Imaging (MRI) and Computer Aero Acoustics (CAA) [16-18]. At this point the Gibb's oscillations are not causing a serious problem in the SAE solution, so the added effort of implementing a removal technique would not noticeably improve the accuracy. It is something that might be of interest in the future, in applications which demand much higher resolution, in this case there are a number of techniques which may be used to remove these oscillations [13-15, 17].

The next generation of ground based telescopes, with primary mirror diameters from 30-100m will contain a great many mirror segments. Therefore, several sets of simulations were conducted to test the SAE accuracy with larger numbers of segments. One example, with 37 segments, is shown in Figure 5.21. The Keck telescopes, which form the world's current largest telescope, contain 36 segments each [8, 9], therefore this number was chosen to simulate the algorithm's performance for a practical situation. As these plots show the SAE has performed well with the larger number of segments, the EM for this reconstruction is 0.998.

The Gibbs phenomena appears worse in this reconstruction (see Figure 5.21(d)) due to the smaller size of the segments. In the previous example (Figure 5.20) each segment was 43 pixels wide whereas in this example each segment is 18 pixels across. Therefore the oscillations do not have sufficient space to die down across the segments. However, the reconstruction is still very good even with this apparent increase in the strength of the Gibbs oscillations.

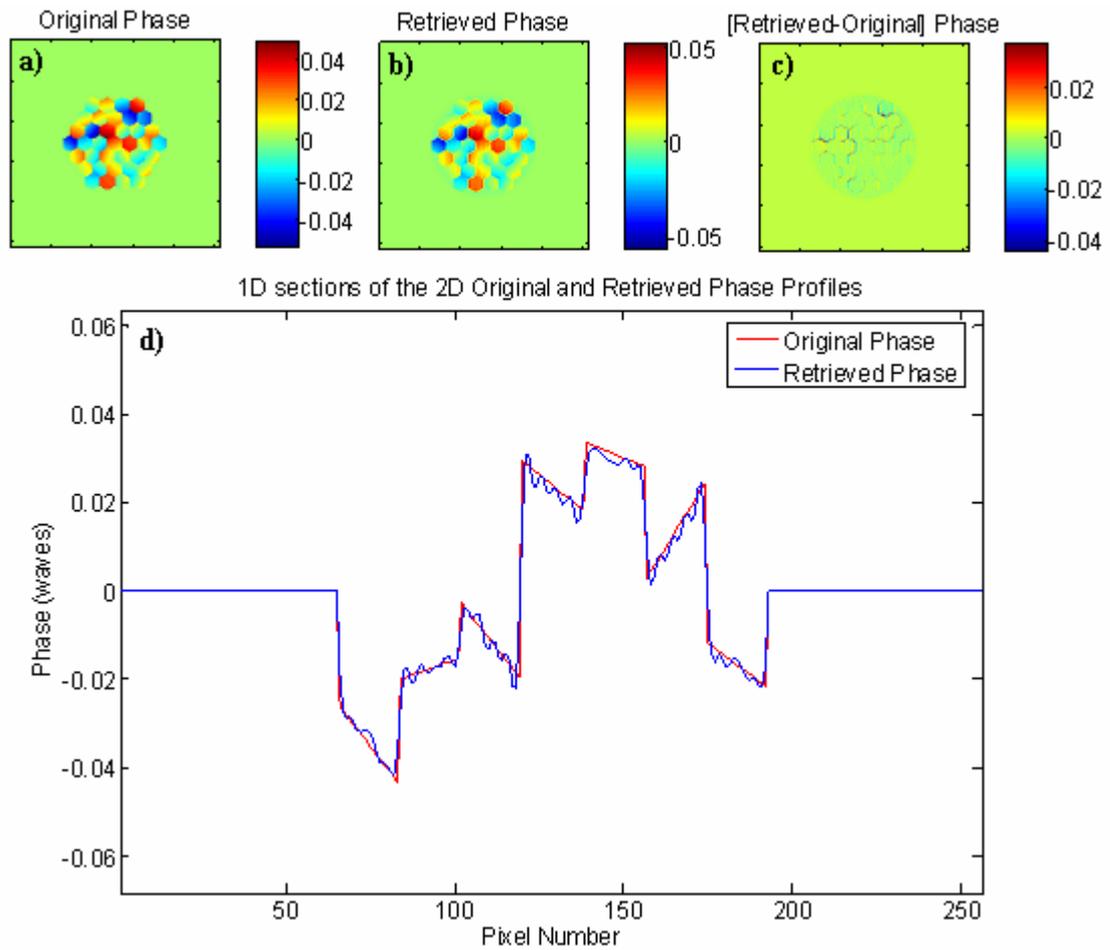


Figure 5.21 The SAE reconstruction (b) of segmented input wavefront (a). The difference between these is given in (c). A cross section of the original and retrieved phase profiles is plotted in (d) [all scales are in waves].

The examples shown in Figures 5.20 and 5.21 are both for small phase errors between the segments. Figure 5.22 is a larger angle example, with the maximum input misalignment set at 0.46λ PV, which is still within the monotonic region for the error signal (see Chapter 4 for details). As this plot shows, there are much larger errors between the original phase profile and the retrieved solution. Figure 5.22(c), the difference between the original and retrieved phase profile, shows that the largest errors in the retrieved phase are from the segments which had the most misalignment to begin with. The accurate co-phasing of the mirror segments is likely to be an iterative process, in which case as the misalignment is reduced so the accuracy of the SAE solution will increase. Therefore it is hoped that this process will converge upon the correct solution within just a few iterations.

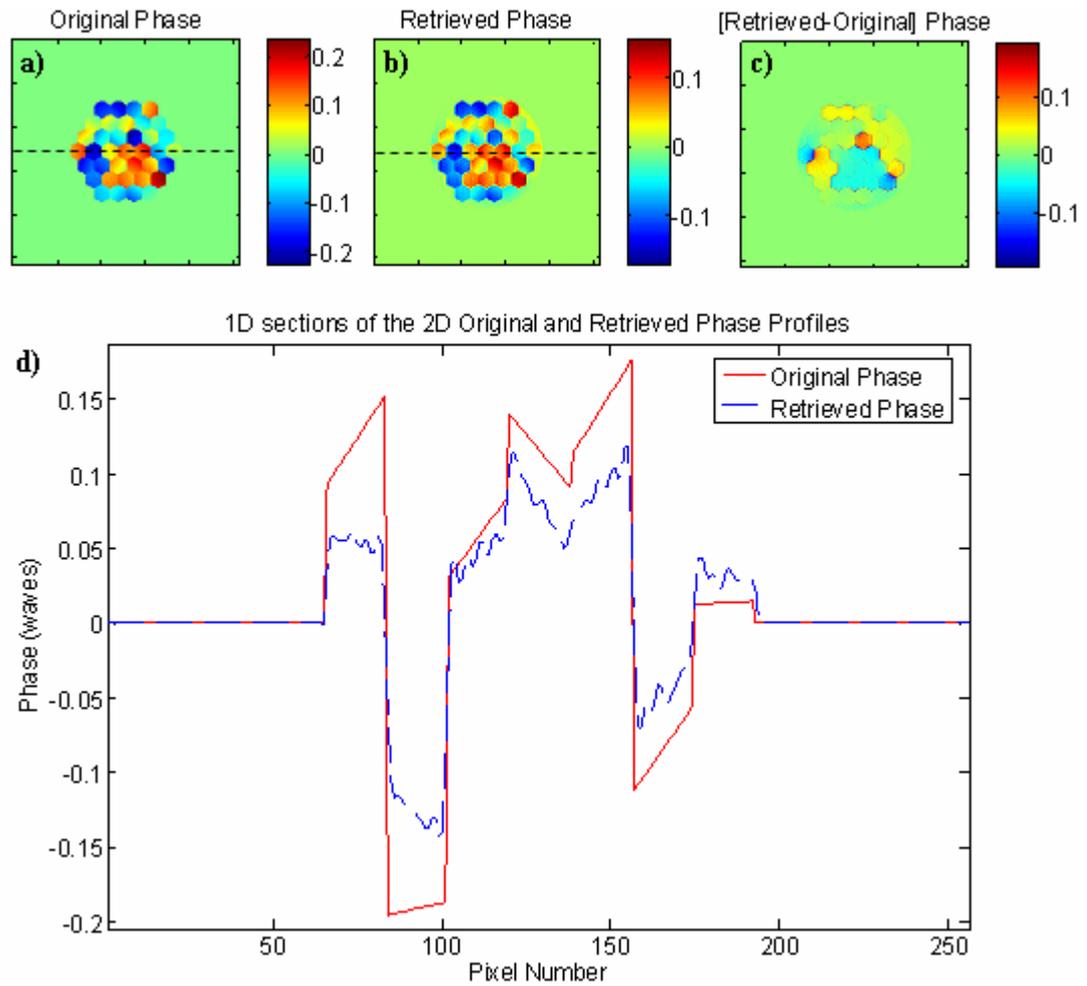


Figure 5.22 (a) Original and (b) retrieved phase profiles for a segmented telescope with larger phase errors, (c) shows the difference between these profiles. Cross sections of these profiles (taken along the dashed lines in plots (a) and (b)) are given in (d) [All scales are in waves].

It was also found in simulation that globally adding multiples of 0.25λ piston error to the same input phase profile will yield the same retrieved phase in each case (which will be the phase solution from the monotonic region e.g. shown in 5.22(d)). This is the same effect as described in Chapter 4 (Section 4.4) and therefore some iteration will be required to find the correct solution if the phase errors in the telescope alignment exceed 0.5λ PV. Perhaps another solution would be to reduce the co-phasing errors using current techniques to reach a rough alignment and to refine the solution using the SAE. As part of the future work for this project an experiment will be conducted using a small

segmented mirror to test the practicality of using this method to align a real telescope mirror.

In conclusion, in all the simulations conducted with discontinuous segmented mirrors the SAE was shown to perform reasonably well. With increased computing power these simulations could be extended to create input wavefronts with similar numbers of segments to the next generation telescopes. This was not possible during this project, but from these simulations it is concluded that GPD wavefront sensing of large primary segmented telescopes could provide useful information on co-phasing errors and should be investigated further.

5.4 Conclusions

In this chapter the SAE algorithm, and factors affecting its accuracy, have been studied in more detail. It has been shown that the properties of the PTF are vital in determining the performance of the GPD sensor and the regularisation issues that will have to be addressed. The user has complete control over the choice of diversity function (and therefore the PTF) so in this chapter the question of how to make an informed choice has been addressed.

It was found that ideally the PTF should contain as few zero's as possible to minimise the loss of information, since spatial frequencies of the input wavefront won't be sampled at these points. Also, fewer zeros mean less regularisation problems. 'PTF Engineering' whereby the user seeks to combine the favourable PTF properties of more than one diversity function in linear combination was discussed. It was found that, in all cases, at least one combination contained the desirable qualities from each of its constituent phases individual PTF's. In these examples only linear combinations of two Zernike diversity function were considered, these alone provide a vast number of possible diversity functions. Perhaps other, non-Zernike, allowable GPD functions or other combination methods would yield better results. It is for the user to decide whether the improvements gained in sensitivity and performance would warrant the added complexity in design and implementation of a combination diversity function.

These results have shown that there is a great deal of choice of combination functions, and the likelihood of finding one which will optimise the sensitivity of the GPD sensor to the user's particular application is very good. Analysis of the PTF is just one tool which the user can employ to make the best choice for their needs.

PCM has also been demonstrated as a suitable diversity function for a GPD null sensor. It is encouraging to find that the analysis in this project can be applied to such a well known example. The PTF generated by this diversity function matches the requirements for a 'good' function as outlined in this chapter and PCM was shown to work particularly well with wavefronts containing high spatial frequency information.

Finally the performance of the SAE reconstruction when dealing with discontinuous wavefronts and wavefronts with non-uniform illumination was studied. It was found that, as expected, introduction of non-uniform illumination does lead to a reduced accuracy in the solution for the wavefront phase. The gradual rate of decline in accuracy as the strength of the non-uniformity was increased suggests that the SAE should be capable of providing a reasonable solution for some level of scintillation. The algorithm was also shown to work well with a small obscuration of the input intensity. Further study of this, with properly modelled scintillation, will be part of the future work for this project. It is impossible to retrieve a perfect solution for the phase since, by definition, information is lost in the null regions when scintillation is present. What constitutes a 'reasonable' solution will depend on the accuracy required for each particular application. In this chapter the GPD SAE system was shown to perform well in sensing the tip/tilt and piston errors of an example of a discontinuous wavefront from a segmented telescope. The main errors in the solution arose from Gibb's oscillations caused by the discontinuities but these are not considered to be a large enough problem at this point to warrant extra effort to remove them. Despite this effect it is believed that with sufficient computer power it would be possible to prove the effectiveness of this technique on the large number of segments required by the next generation telescopes. The GPD sensor would provide a compact versatile solution to the daunting task of co-phasing such huge numbers of mirror segments and could perhaps replace or reduce the need for edge sensors on each segment.

Thus far the GPD SAE sensor system has been developed and tested extensively in simulation. The next important stage is to see how this system performs experimentally; this will be the topic of Chapter 6.

5.5 References

1. Dammann, H., *Colour Separation Gratings*. Applied Optics, 1978. **17**(15): p. 2273-2279.
2. Layet, B., I.G. Cormack, and M. Taghizadeh, *Stripe Colour Separation with Diffractive Optics*. Applied Optics, 1999. **38**(35): p. 7193-7201.
3. Murphy, D.B., *Fundamentals of Light Microscopy and Electronic Imaging*. 1st ed. 2001: John Wiley and Sons.
4. Nikon, *Living Cells in Brightfield and Phase Contrast*. 2006.
5. Huang, T., *Discovered That the Phase Contrast Microscope Satisfies the Gpd Null Sensor Conditions*. Personal Communication 2005
6. Woods, S.C. and A.H. Greenaway, *Wave-Front Sensing by Use of a Green's Function Solution to the Intensity Transport Equation*. J.Opt.Soc.Am.A, 2003. **20**(3): p. 508-512.
7. Gilmozzi, R., et al. *The Future of Filled Aperture Telescopes: Is a 100m Feasible?* in *Advanced Technology Optical/IR Telescopes VI*. 1998. Kona, Hawaii: Proc. SPIE. Vol. 3352. p. 778-791
8. Waldrop, M.M., *Keck Telescope Ushers in a New Era*. Science, 1990. **249**(4974): p. 1244-1246.
9. Medwadowski, S.J., *Structure of the Keck Telescope - an Overview*. Astrophysics and Space Science, 1989. **160**(1-2): p. 33-43.
10. Johns, M., R.P. Angel, S. Sheckman, R. Bernstein, D.G. Fabricant, P. McCarthy, and M. Phillips. *Status of the Giant Magellan Telescope (Gmt) Project*. in *Ground Based Telescopes*. 2004: Proc.SPIE. Vol. 5489. p. 441-453
11. Nelson, J.E. *Design Concepts for the California Extremely Large Telescope*. in *Telescope Structures, Enclosures, Controls, Assembly/Integration, Validation, and Commissioning*. 2000: Proc.SPIE. Vol. 4004. p. 282-289
12. Andersen, T., et al. *The Euro50 Extremely Large Telescope*. in *SPIE*. 2003: Proc. SPIE. Vol. 4840. p. 214-225
13. Jung, J.-H. and B.D. Shizgal, *Generalization of the Inverse Polynomial Reconstruction Method in the Resolution of the Gibbs Phenomenon*. J. of Comp. and Appl. Math., 2004. **172**: p. 131-151.
14. Shizgal, B.D. and J.-H. Jung, *Towards the Resolution of Gibbs Phenomena*. J. of Comp. and Appl. Math., 2003. **161**: p. 41-65.
15. Pasquetti, R., *On Inverse Methods for the Resolution of the Gibbs Phenomenon*. J. of Comp. and Appl. Math., 2004. **170**: p. 303-315.
16. Tam, C.K.W., *Computational Aeroacoustics - Issues and Methods*. Am.Inst. of Aeronautics and Astronautics, 1995. **33**(10): p. 1788-1796.
17. Archibald, R. and A. Gelb, *A Method to Reduce the Gibbs Ringing Artifact in Mri Scans While Keeping Tissue Boundary Integrity*. IEEE Transactions on Medical Imaging, 2002. **21**(4): p. 305-319.
18. Meadows, K.R., D.A. Caughey, and J. Casper, *Computing Unsteady Shock Waves for Aeroacoustic Applications*. Am.Inst. of Aeronautics and Astronautics, 1994. **32**(7): p. 1360-1366.

Chapter 6

Experimental Testing, Practical Considerations and Applications of the Small Angle Expansion Algorithm

Summary of Chapter 6

In Chapters 4 and 5 the Small Angle Expansion (SAE) method of phase retrieval was developed and tested. It was shown to work well in simulation with both continuous and discontinuous wavefronts as well as in the case of non-uniform illumination. In this chapter we will consider the modifications required and practical issues involved in using this algorithm on real experimental data. This involves a more rigorous approach to the regularisation problem identified in Chapters 4 and 5, and further study of the data conditioning. This chapter shows how these problems were approached, and demonstrates the use of the algorithm on real experimental data. Ongoing issues are identified and will be discussed here and in the final chapter as part of the future work for this project. Finally, some particular applications of this new sensor system which are of interest to our industrial collaborators will be presented.

6.1 Introduction

In Chapter 5 the Small Angle Expansion (SAE) algorithm was shown to work with non-uniform illumination and discontinuous wavefronts. This addressed one of the main science drivers for the development of a new wavefront sensor system. However, what is required is a system that works experimentally and not just in simulation. The next step, and the subject of the present chapter, is how this algorithm is modified to be suitable for use with real experimental data. It is often not a trivial process to turn a simulation problem into a practical solution, and the discussions presented in this chapter will show that there are many issues to be considered and resolved. Some of these will be addressed here and others are held over for future work which will be discussed in the final chapter.

The principle difference between real and simulated data is that real data will contain noise created by variations in environmental conditions (such as background light levels) at the time the data is recorded, as well as instrumental noise and experimental errors. The inclusion of this noise in the recorded data results in a classic regularisation problem, which was first discussed in Chapter 4 (section 4.3.4). This discussion is continued in this chapter and a more rigorous approach to regularisation, using a Wiener filter, will be described and implemented in the SAE algorithm.

In addition to the regularisation required to analyse experimental data the issues of data manipulation must be considered, namely normalisation and conditioning. Normalisation of the data is required to ensure that a universal change in the intensity (for example caused by increasing the exposure time) does not alter the retrieved phase solution. Modification of the experimental data by $I_0(r)$, as specified in Chapter 4 (see Equation 4.13) will be referred to here as data conditioning, and will affect the stability of the solution. It is not immediately clear which value should be used for I_0 , which is the intensity in the pupil plane. For example, should the intensity in the 0th order be used? Or should it be formed using a combination of the intensity values in the ± 1 orders $((I_1 + I_2)/2)$? These are just two options for the treatment of the data conditioning, the merits of each and the possibility of other solutions will form the basis of Section 6.4.

Finally, the first experimental results using the SAE algorithm will be given and the practical issues for successful data collection and reduction will be discussed. Issues such as proper sampling and correct centring of the intensity images will be examined. As is often the case problems are encountered when trying to implement a real experiment for the first time that are not anticipated from experience with simulated data. For comparison the preliminary experimental results will also be analysed using the Gureyev-Nugent (GN) algorithm (which was first introduced in Chapter 2).

By the end of this chapter it will have been shown that a Generalised Phase Diversity (GPD) sensor, coupled with the SAE algorithm, has the potential to be a useful and versatile wavefront sensing system. Possible applications of this device will be

considered, focussing on those which are of particular interest to our industrial collaborators. These include thin-film metrology, ‘designer beam shape’ lasers, military imaging, astronomy and biomedical engineering applications.

6.2 Experimental Testing of the GPD Sensor and SAE Algorithm

Following the success of the GPD method and SAE algorithm on simulated data the next step was to extend this to a real experiment. After discussions with our collaborators at Zeeko and University College London (UCL) it was decided that the wavefront sensor should be configured to image a pupil plane at a distance of 1m or less from the wavefront sensor. The reasoning for this is that Zeeko, who build polishing machines, would eventually like to mount the GPD system on the end of a swing arm profilometer. In this position the sensor would be arranged to study the surface of the test piece as it is profiled, at a height of about 1m above the sample. The experimental set-up, which will be detailed in the next section, was built with this in mind.

Adapting the GPD+SAE system for experimental testing raised several interesting problems and these will be discussed throughout this chapter.

6.2.1 Experimental Set-Up

Figure 6.1 shows the experimental set-up for the initial tests of the GPD sensor and SAE algorithm. In this experiment a single mode fiber, coupled to a 2mW 633nm Helium Neon (HeNe) laser, was used to create a rapidly diverging point source. Figure 6.1(a) demonstrates how the fiber was mounted on a translation stage to allow it to be moved with respect to lens L1 ($f=250\text{mm}$) which, depending on the source-L1 separation, would either collimate the beam or allow it to be slightly converging/diverging. The reason for this will be discussed later in section 6.2.2 where practical issues arising from this arrangement are discussed. The camera lens, L2 ($f=35\text{mm}$), was positioned so the pupil plane (see Figure 6.1) was in focus. In the 0th order an image of this plane was obtained, and in the ± 1 orders the intensity images created were images of the pupil plane convolved with the blur function of the defocus grating.

The ± 1 order images, for this experimental setup, were approximately 90 pixels in diameter. In Figure 6.1(b) the physical distances separating the various components are detailed. In this diagram $\pm \Delta$ mm refers to the translation of the source about the position shown. The source was translated by a total amount of ± 10 mm (i.e. $\Delta=5$ mm) and the series of data images was captured at 1mm intervals along this translation. The distances shown here were chosen to maximise the amount of optical bench space available for this experiment. The CCD camera used to capture the data was an electronically cooled Retiga EX Qimaging camera which has 12 bit digital output, a 1360(h) x 1036(v) array of $6.45\mu\text{m}$ square pixels, and a 60dB Signal-to-Noise ratio. The quoted readout noise and dark current for this camera are $8 e^-$ and $0.15 e^-$ /pixel/second cooled. These figures have been experimentally measured for this camera to be 3.8 counts/pixel, and the dark count is 0.05 counts/pixel/second respectively [1].

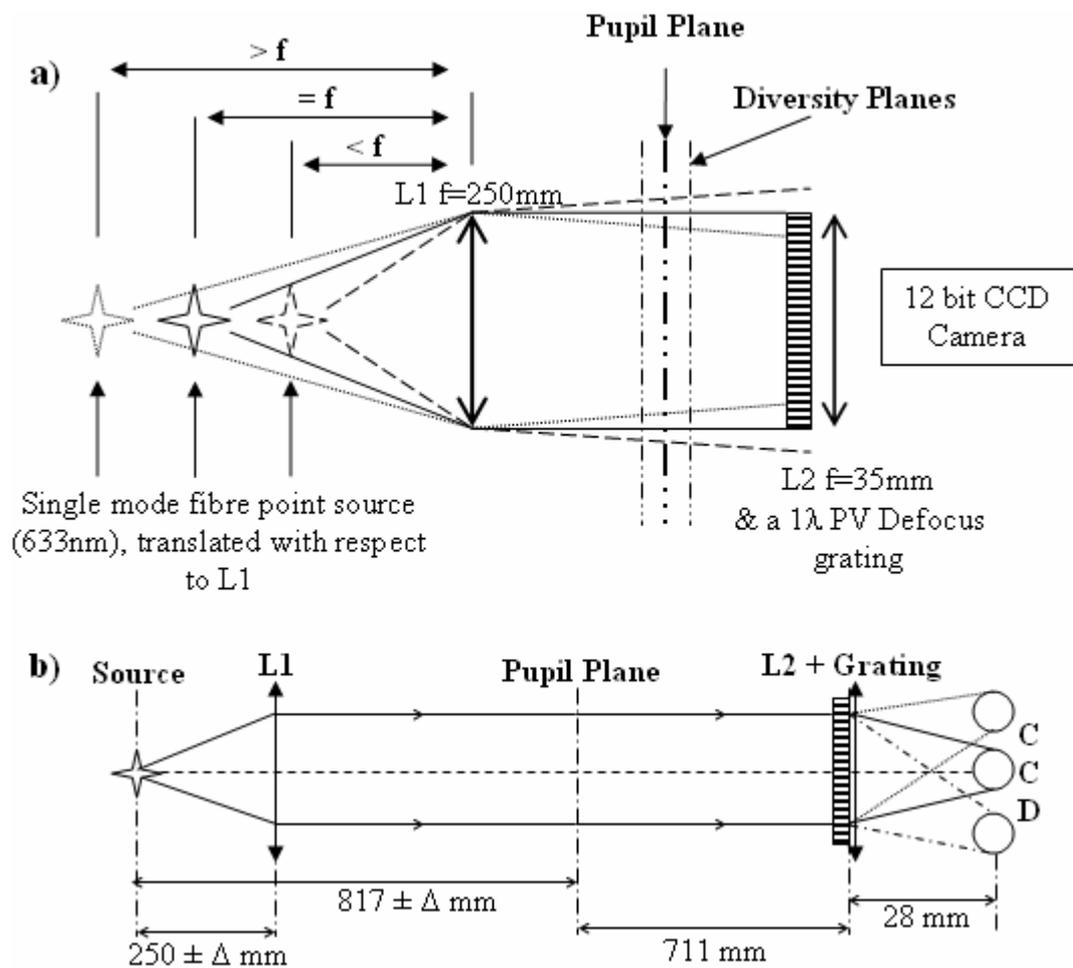


Figure 6.1 (a) The experimental set-up for the first practical test of the GPD sensor and SAE algorithm (b) A schematic showing the distances between the components in (a).

Although one of the main reasons for developing GPD was to use aberrations other than defocus this was chosen to be the first diversity function tested. Using this function would allow the data to also be analysed by the GN algorithm (see Chapter 2 for details). It was hoped that direct comparison of the SAE with an Intensity Transport Equation (ITE) based method would give a good indication of the accuracy of the new algorithm.

The ‘Pupil Plane’ as shown in Figure 6.1 was an arbitrary plane in the optical system that, due to our collaborator’s application, was required to be less than 1m from the grating and lens combination. Altering the position of this chosen plane or the focal length of L2 could be used to change the magnification of the data on the camera. For Zeeko and UCL’s application the ‘pupil plane’ would be the plane of their sample. The light from the point source could be directed onto the sample and reflected off at an angle to be collected by the GPD sensor (the L2-grating-CCD combination) which would be focussed on the sample plane. In future experiments it is planned to assess the surface flatness of a small plane mirror, using a collimated probe beam, and imaging the plane of the mirror as the pupil.

6.2.2 Practical Issues

In Chapter 4 the relationship between the width of the blur function and the linearity of the error signal produced was discussed in detail. It was found that the error signal has a sinusoidal response to the size of the phase step across the width of the blur function. The error signal will only be monotonic in the region $\pm\pi/2$ ($\lambda/4$). It was concluded that a sufficiently narrow blur function could allow phase retrieval of input wavefronts with large Peak-to-Valley (PV) amplitude if the local slope across the blur width did not exceed the monotonic limit. In a practical experiment this must also be linked to the sampling of the intensity images.

The size of the images in the ± 1 orders is related to the strength of the diversity and input wavefront phases as well as the magnification introduced by any other lenses in the optical set-up. Care must be taken when designing and setting up the GPD sensor so

that the wavefront slope from pixel to pixel does not exceed the $\pm\pi/2$ limit. In general, it is best to make the images as large as is practically possible. This consideration, plus the level of noise in the data, will limit the dynamic range of this method. It should be noted that this is a practical limit that applies to all Phase Diversity (PD) phase reconstruction methods, since it is the data that becomes ambiguous rather than a feature of the algorithm.

In this experiment, to keep the PV amplitude of the wavefront well within the monotonic region L1 was used to create a beam which was close to collimated. In an ideal experiment it would be preferable to remove this lens and thus reduce the overall aberrations of the optical system itself. However, in practice it was found that doing so would produce a beam in the pupil plane whose curvature would exceed the monotonic limit for the SAE. The curvature of this beam could not be reduced by simply placing the source at a greater distance from the pupil plane due to space restraints in the lab. By introducing the collimating lens, using Gaussian ray transfer matrix methods [2, 3], it was calculated that the change in curvature over the range of translation used in the experiment (shown in Figure 6.1) was 7λ in total. The blur function was only 2 camera pixels in width, therefore over the 90 pixel sized spots the wavefront could be well sampled and the error signal monotonic.

Another practical consideration is to avoid saturation of the camera whenever possible. The SAE solution, as with many other algorithms, is dependent on the accurate calculation of the difference between two intensity images. For example, if the camera is saturated then there is no way of knowing if the real intensity value is just outwith the camera's sensitive range or 10 times the maximum value. It is impossible to get a meaningful difference from these saturated areas in the image. It may be necessary to readjust the camera settings in real-time to prevent saturation if the intensity values are expected to vary by large amounts. For any given camera providing a given bit precision in output this also provides a limit to the dynamic range.

6.2.2.1 Centring Accuracy of the Data Images

The error signal which is analysed by the SAE algorithm to retrieve the phase is calculated by taking the difference of the intensity images in the ± 1 diffraction orders. In a practical experiment, unlike the simulated case, a method for finding the centres of these images must be chosen so that they may be cropped and subtracted as accurately as possible.

A common centroiding method is to find the centre of mass of the given image. However, if the aberrations present in the beam cause an uneven distribution of intensity in the data images (for example coma produces a bright crescent on one edge) then the centroid calculated will not be the geometric centre of the spot. To calculate the geometric centre the boundaries of the intensity image must be calculated to high precision. When noise is included in the data this may not be a trivial task. With any practical method it is quite possible that the centres of the data images will only be accurate to within a few pixels. The question is, how will this affect the retrieved solution? In this section the affects of misalignment are explored through simulation. These simulations uncovered an unexpected ambiguity in the SAE solution and this will also be discussed here.

A simulation was constructed to test the effect of misalignment between the centres of the two data images. The simulation is designed to allow one of the data images to be shifted by 1 pixel in the x-direction and the error signal generated for a plane wave is calculated with and without this shift. For diversity phases with suitable symmetry properties for a GPD null sensor the error signal for a plane wave should be zero to within computer rounding error, so any structure present after the shift is caused by the misalignment. Figure 6.2 shows the results of this simulation, which uses defocus as the diversity phase. In 6.2(a) the error signal created by a plane wave is shown and in 6.2(b) the error signal created by the same data but with a 1 pixel shift is shown.

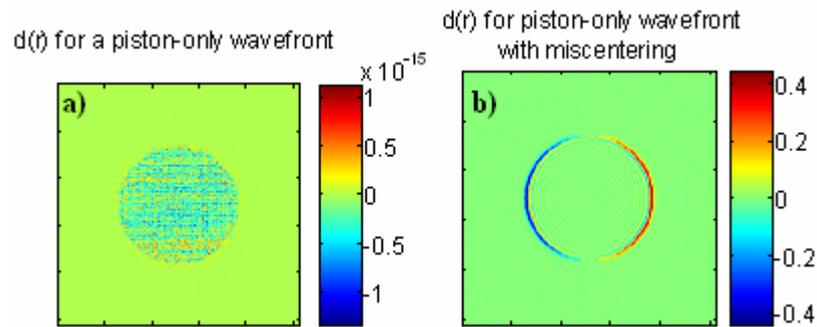


Figure 6.2 The error signal generated for a piston-only (plane) wavefront with (a) no misalignment and (b) 1 pixel of misalignment (in the x-direction), between the data images in the \pm diffraction orders.

As expected, the perfectly aligned simulation with plane wave input has resulted in a null output to within computer rounding error (see Figure 6.2 (a)). A non-null output for the same plane wave input has been calculated (and is shown in 6.2 (b)) when a shift of just 1 pixel is applied to one of the data images. The $d(r)$ generated looks like the error signal from a wavefront with tilt aberration. To test this further the error signal in 6.2 (b) was input into the SAE algorithm and the result decomposed into a set of Zernike coefficients. Interestingly it was found that the two largest aberrations calculated from this error signal input were coma and tilt, with the largest contribution from coma. The reason for this can be understood by considering Figure 6.3. In this figure the error signal generated by a tilted and a comatic wavefront are plotted in 6.3 (a) and (b) respectively. The similarity between these two is immediately apparent. For the tilted wavefront the error signal comprises a bright negative crescent on one edge, and an equally bright positive crescent on the opposite edge. The internal structure in this error signal is due purely to diffraction effects from the hard edged pupil used. The comatic wavefront shares the same crescent-like features at the edges with a small amount of signal in the centre from a pair of positive and negative peaks. In Figure 6.3 (c) an example of coma phase is shown in profile and it can be seen from this that this aberration contains a large tilt-like element. This is related to the fact that coma can be caused by misalignment of lenses in an optical system such that incident light does not travel parallel to the optic axis. Rays from an off-axis object passing through the outer margins of the lenses will not be focussed to a single tight spot but a series of comatic circles, the sum of which produces the characteristic V-shaped flare in the focal plane.

Tilt is also characteristic of off-axis objects, a wavefront containing a pure tilt term will appear as a plane wave which has originated from an off-axis source and is travelling through the optical system at an angle to the optic axis [2, 4]. With this in mind it is not surprising that the coma profile shown in Figure 6.3(c) contains a large tilt-like component. The confusion of the SAE algorithm between tilt and coma is likely to be caused by the similarity between the error signals generated by each, as demonstrated in Figure 6.3(a) and (b). Note that the structure of $d(r)$ is dependent on the diversity phase used, as well as the wavefront aberration. Further tests below will show that this problem is not suffered by all GPD diversity functions.

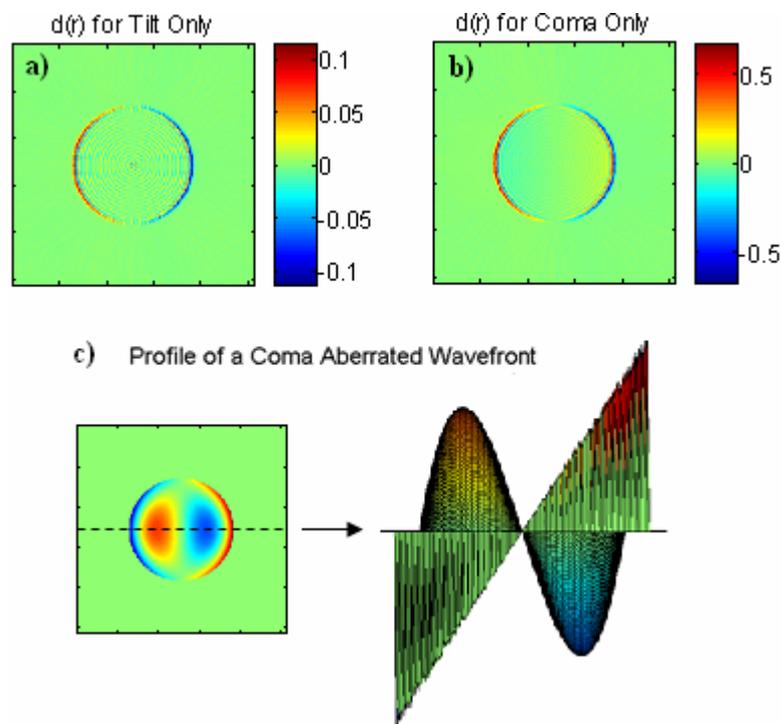


Figure 6.3 The error signal $d(r)$ generated for a wavefront with (a) only tilt phase errors (b) only coma phase errors [scales are a measure of contrast]. A plot of a coma phase error is shown in profile in (c), it demonstrates a large tilt-like component.

To confirm that the SAE has difficulty distinguishing tilt errors from coma errors a further simulation was carried out. In each case the diversity phase used was defocus and the retrieved phase for a wavefront with either tilt or coma only phase errors was

calculated. These results are plotted in Figure 6.4 and confirm that the SAE will retrieve almost the same phase profile for the two different input aberrations.

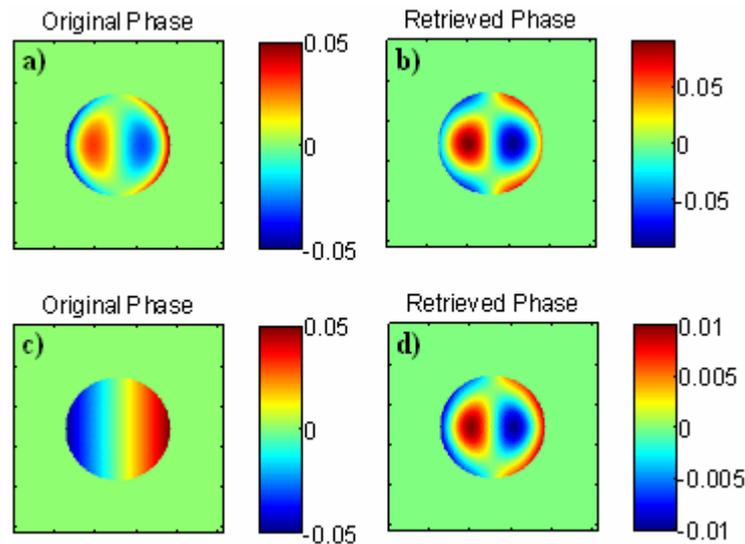


Figure 6.4 The original (a) and retrieved (b) phase profiles for a wavefront with Coma. The original (c) and retrieved (d) phase profiles for a wavefront with only tilt aberration. All scales are in waves.

The results presented here have shown that in a practical situation, and with defocus diversity, misalignment of the intensity images prior to their subtraction will result in an extra tilt and coma term being introduced to the solution. This was found to be the case with aberrated wavefronts as well as in the plane wave case presented here. This ambiguity in the SAE solution for tilt and coma was an unexpected outcome of these simulations and prompted further study.

Further simulations were conducted to determine whether this confusion between tilt and coma is a universal problem for the SAE algorithm, or whether it is confined to certain diversity functions. In a number of allowable diversity functions tested (including spherical aberration) the quality of the reconstruction was extremely poor. This, as discussed in Chapter 5, is a result of this technique's occasional difficulty in reconstructing low order aberration modes. Aberration modes (such as tilt) whose FT is concentrated at low spatial frequencies (close to the origin) will not be properly sensed

when using diversity functions whose Phase Transfer Function (PTF) is zero at the origin. As shown in Chapter 5, all Zernike functions with suitable symmetries for use in a GPD null sensor will have a zero in the PTF at the origin. The phase diverse data (and therefore $d(r)$) is created by convolution of the FT of the input wavefront with this function, therefore wherever the PTF goes to zero those spatial frequencies in the wavefront will be lost. For some diversity functions the PTF has a steep slope about the origin and recovers quickly to some significant value. In these cases the SAE will be more sensitive to low order aberration modes in the input wavefront. As mentioned above, a number of the diversity functions tested did not have sufficient sensitivity to the low order mode to retrieve an identifiable wavefront tilt. Of the diversity functions which did return a solution for the wavefront phase all, with the exception of defocus, were able to tell the difference between tilt and coma. Two examples of this are shown here. The diversity phases used were Z_{10}^0 (Figure 6.5) and the Phase Contrast Microscope (PCM) diversity function from Chapter 5 (Figure 6.6).

Figure 6.5(c) shows that, although the form of the retrieved phase term is correct, there are large amplitude errors in the solution. This again is linked to the low-order mode sensitivity problem discussed earlier. Z_{10}^0 , like defocus, is one of the radially symmetric Zernike diversity functions. Figure 6.5 shows that although the error signals this function generates for tilt and coma are similar (6.5 (a) and (d)) the SAE is still able to distinguish them. Comparing these error signals to those shown for defocus in Figure 6.3 (a) and (b) it may be seen that, although similar, the error signals generated by Z_{10}^0 have much higher contrast. This is evident in the strong signal in 6.5(d), forming a pair of reverse lobes. The increased signal level has resulted in extra sensitivity when using this diversity function when compared to the defocus diversity used earlier.

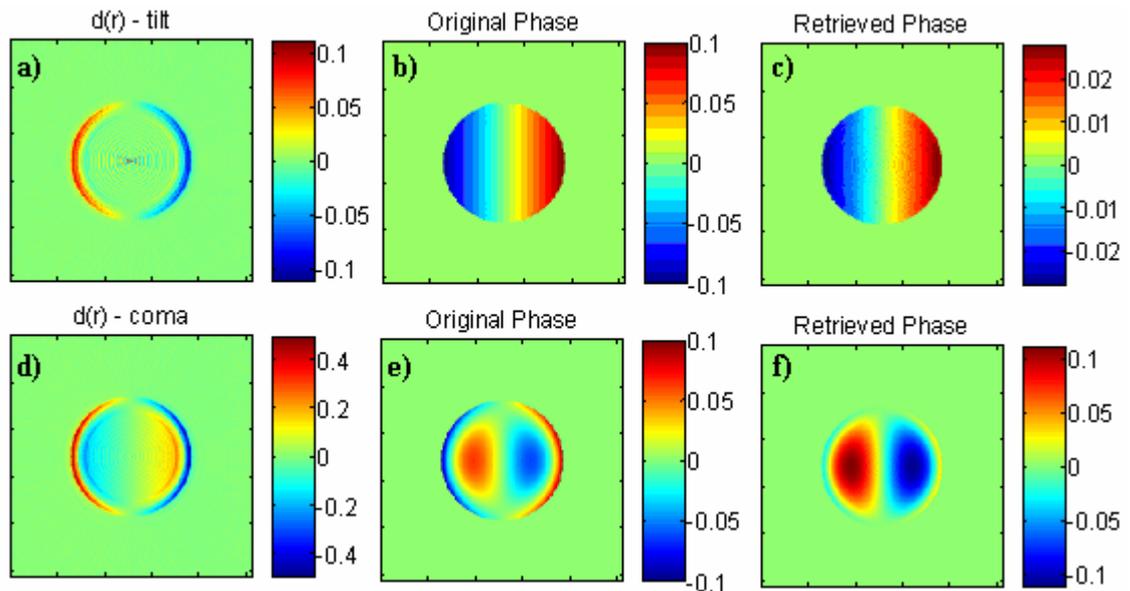


Figure 6.5 The error signals (a,d) generated using Z_{10}^0 diversity for two wavefronts, one with tilt (b) aberration and another with coma (e). The SAE retrieved phase profiles for the tilt and coma aberrated wavefronts are given in (c) and (f) respectively [scales in plots b,c,e,f are in waves. Scales of plots a and d are a measure of contrast].

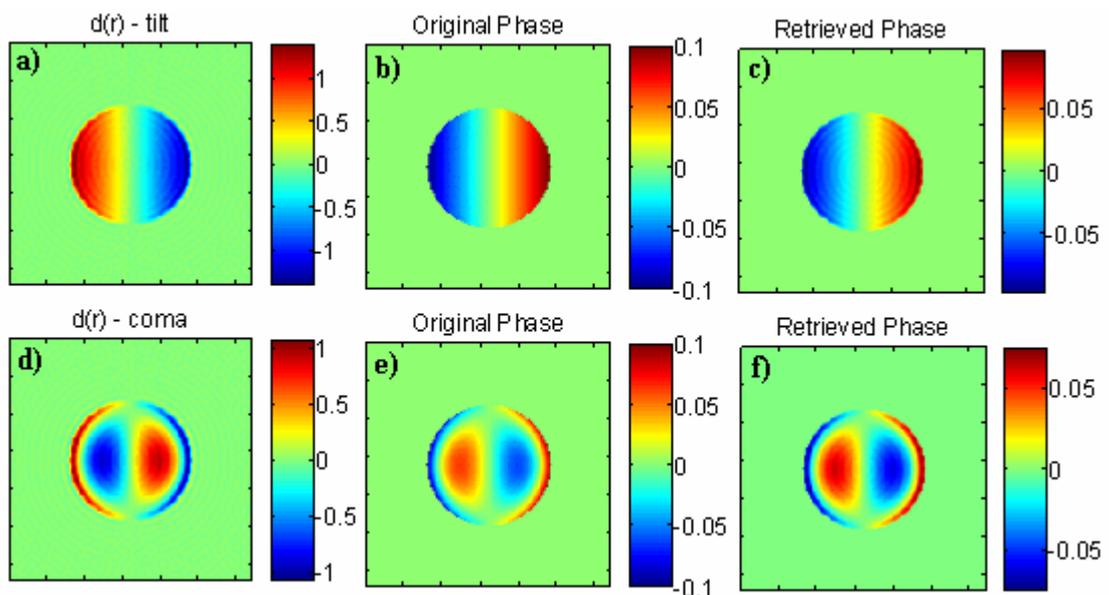


Figure 6.6 The error signals (a,d) generated using PCM diversity for two wavefronts, one with tilt (b) aberration and another with coma (e). The SAE retrieved phase profiles for the tilt and coma aberrated wavefronts are given in (c) and (f) respectively [scales in plots b,c,e,f are in waves. Scales of plots a and d are a measure of contrast].

Figure 6.6 shows the results obtained using the same PCM diversity function which was presented in Chapter 5. Here, as in that chapter, it has been shown to perform very well. Due to the form of its PTF the PCM diversity is more suited to dealing with wavefronts with low order modes than some of the Zernike based diversity functions tested (see Chapter 5 for details).

In conclusion, unless dealing with defocus as the diversity function, a misalignment of the data images prior to their subtraction will create an artificial tilt term in the retrieved solution. When using defocus, in addition to the tilt term, coma will also be present in the solution. It is believed that this ambiguity is due to the very similar form of the error signal generated by both of these wavefront errors. In both cases the error signal is dominated by the edges, and the difference between them is a very low contrast signal in the centre. Since the Green's function and Gureyev-Nugent (GN) algorithms do not suffer similar problems and since the data itself is unique (albeit low contrast) the SAE should be able to tell the difference between coma and tilt. Further investigation of this phenomenon should be conducted as part of the future work of this project.

6.3 Initial Results

Figure 6.1 shows the experimental set-up which was used to obtain the initial results for use with the SAE algorithm. As 6.1(a) shows, the point source was translated with respect to L1 such that the beam produced should go from being converging, to collimated, to diverging as the fibre is moved towards the lens. If the initial position is taken to be when the fibre is f_{L1} from the lens then data was obtained $\pm 5\text{mm}$ from this point. From this arrangement it is expected that the largest aberration on the input wavefront will be curvature. The data images were background subtracted and then processed using the SAE algorithm, and the GN algorithm.

In Figure 6.7 (d) and (e) the retrieved phase profiles from one of these datasets is shown for both the SAE and GN algorithm respectively. In 6.7 (a)-(c) the actual camera data which was used to generate these retrieved phase profiles is shown. In the following section these results will be compared in detail.

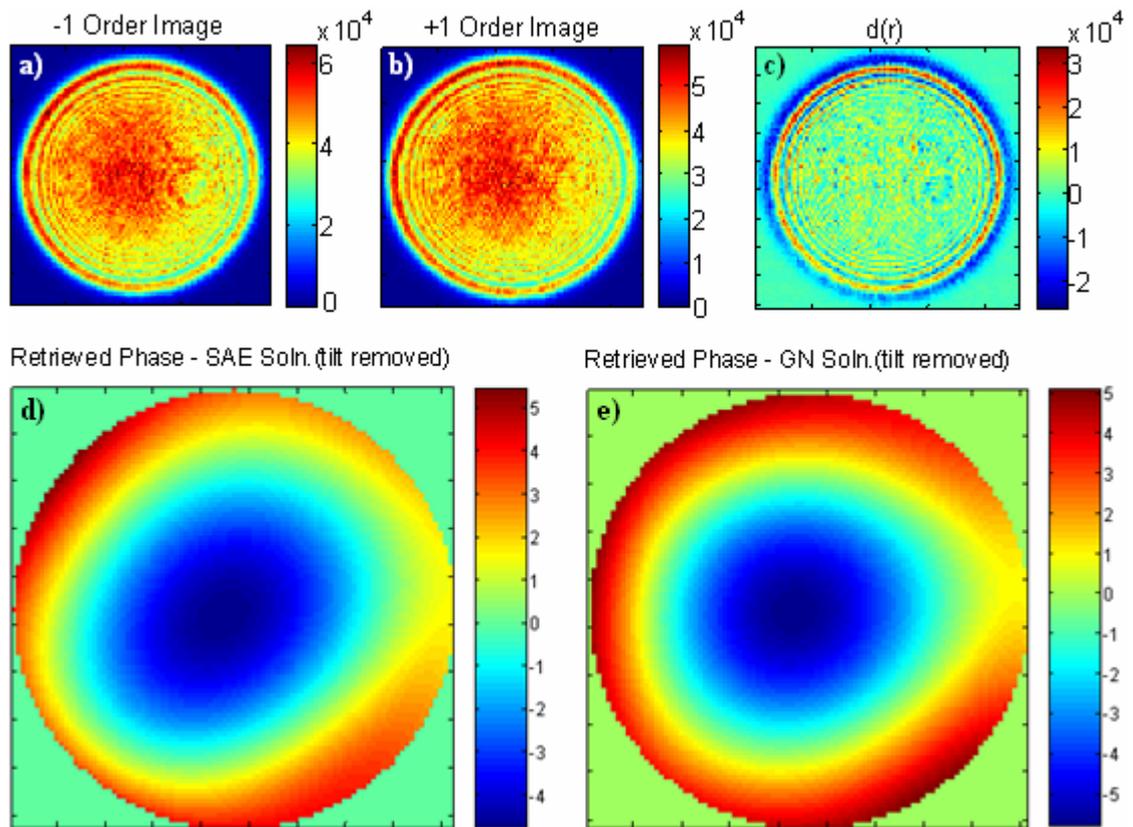
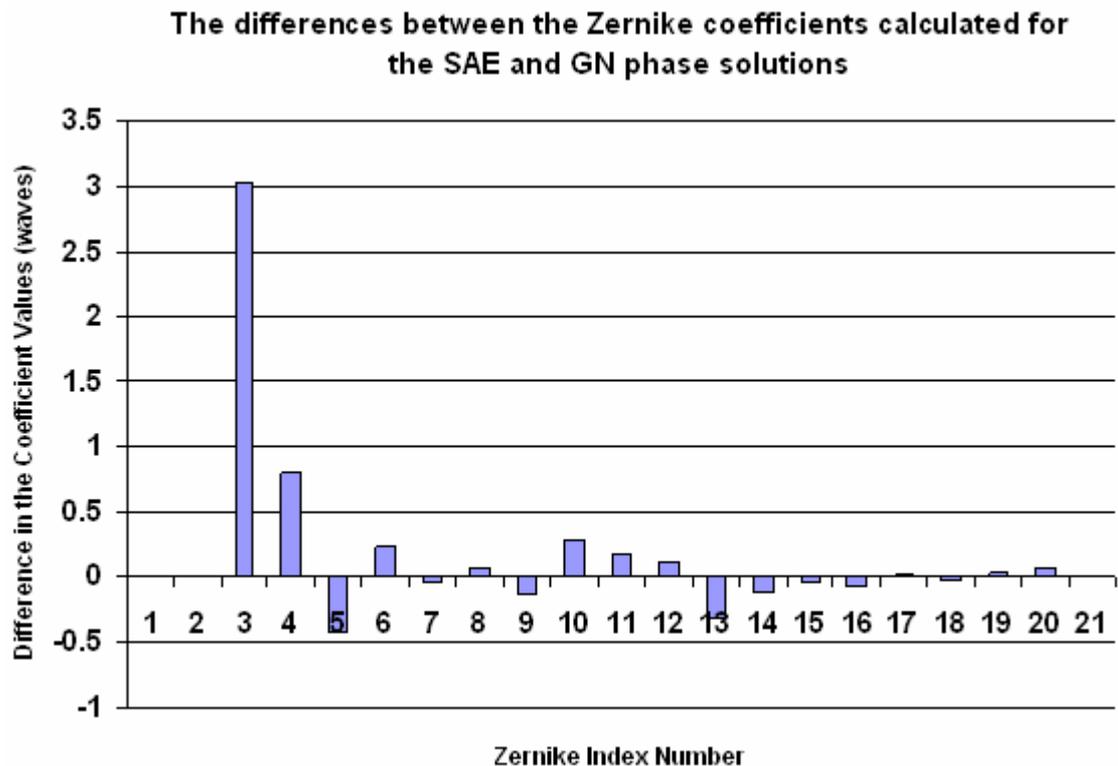


Figure 6.7 Initial experimental results from the set-up detailed in Figure 6.1. The SAE (d) and GN (e) solutions for the retrieved phase both have tilt removed [scales are in waves]. The phase diverse images used to calculate these retrieved phase profiles are shown in (a) and (b), and their difference, $d(r)$ is shown in (c).

6.3.1 Comparison with the GN Algorithm Solution

Upon first inspection of the results (in Figure 6.7(d) and (e)) it may be seen that both algorithms have returned a defocus dominated solution for the wavefront phase. This agrees with expectations given the way in which the experiment was set-up (see Figure 6.1). The PV aberration of the retrieved phase profiles is also similar for both algorithms: 10.2λ for the SAE solution and 10.9λ for the GN result. Although the results displayed in Figure 6.7 are tilt removed a Zernike decomposition of both sets of results, without tilt removal, was conducted. In both solutions the largest contribution

came from tilt, followed by defocus. To compare the SAE and GN phase solutions the difference between each retrieved Zernike coefficient was calculated.



Graph 6.1 A histogram plotting the differences between the SAE and GN solutions for each of the Zernike coefficients. Since the GN algorithm does not calculate piston or tip (index numbers 1 and 2) no values for the difference are plotted for these coefficients.

Graph 6.1 is a histogram of these differences for each Zernike coefficients. The Zernike coefficients are labelled by ‘index number’ from 1 to 21, and a table which relates these index numbers to their Zernike Polynomials is given in the Appendix (see Section 6.9). Note that there is no difference plotted for Zernike indices 1 and 2. These correspond to piston (1) and tip (2) which are not calculated by the GN algorithm. Therefore no meaningful difference could be plotted for these first two coefficients and they are left blank in Graph 6.1 for this reason. Another point to note here is that the coefficients retrieved, for any Zernike decomposition, depend highly on the accurate definition of the position of the unit disc (the area over which the polynomials should be fitted). A difference of only a single pixel in the declared position of the data boundary will result in a significant change in the values of the calculated coefficients. Therefore, to minimise the problems this could create, the GN algorithm was given the same cropped

data images as the SAE. Any error in the position of the boundary would therefore be the same for both algorithms so, while the absolute values of the coefficients may be incorrect due to this error, the relative values of the coefficients from the SAE and GN algorithm solutions should be the same.

As Graph 6.1 demonstrates the largest disparities between the two solutions are for tilt (index number 3), defocus (4), astigmatism modes (5, 6, 9 and 10) and for the modes surrounding spherical aberration (Z_4^0 , index number 11); Z_4^2 (12), Z_4^{-2} (13) and Z_4^4 (14). Apart from tilt and defocus the differences between the solutions for all other modes is less than 0.5λ , and in most cases much less than that. The largest difference, in the tilt value, could be attributed to the way in which each algorithm deals with misalignment of the data image. This misalignment will be discussed further later in this section. Since tilt is routinely corrected for in a large number of AO systems, using a tilt correction stage usually comprising an adjustable mirror to redirect the beam, the absolute value of this term is of less importance than other modes in a practical situation. The difference of the GN and SAE solutions on the value of the defocus coefficient is of more immediate concern. Since the GN algorithm has been shown to perform extremely accurately in the retrieval of defocus [5] (see Chapter 2 for details) the error in this term could indicate a problem with the accuracy of the SAE solution. This will also be discussed further later in this chapter.

In most respects the SAE and GN solutions compare favourably well and the results are sensible given our knowledge of the system. Aside from the differences in the computed Zernike coefficients there is another very noticeable difference between these solutions; their orientation with respect to each other. Both algorithms were carefully studied to identify any steps within the code where a rotation was employed. In neither algorithm was the data or solution deliberately rotated. Study of the difference image used in both cases (see Figure 6.8) reveals a misalignment of the data images consistent with a $\frac{1}{2}$ pixel shift in both the x and y directions. The dotted line in Figure 6.8 shows the direction of this misalignment. Comparing this with the retrieved phase profiles in Figure 6.7 it can be seen that the SAE solution matches the axis of symmetry seen in the difference image.

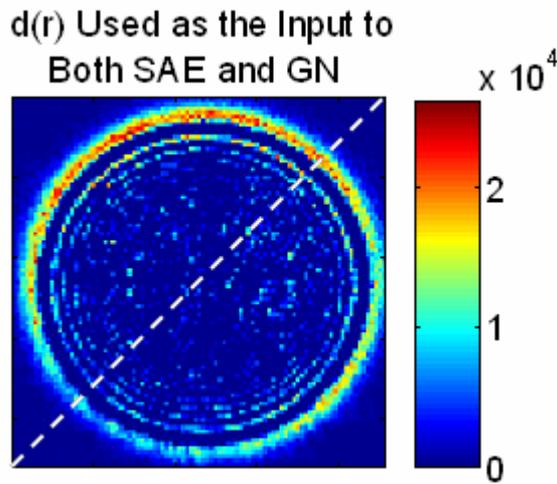


Figure 6.8 The difference image $d(r)$ from the set of real data used to calculate the retrieved phase profiles in Figure 6.7. The dotted line shows the axis of symmetry created by a $\frac{1}{2}$ pixel misalignment in x and y of the data images subtracted to create this image.

The misalignment of the data shown in Figure 6.8 would support the finding of a large tilt term in both the SAE and GN solutions. It was thought that perhaps the apparent rotation between the SAE and GN solutions is a result of the difference in astigmatism modes between the two solutions. As a simple check a simulation was conducted in which a phase profile similar to the form of the SAE solution in Figure 6.7(a) was input into both algorithms. This simulated phase profile was created using a mixture of defocus and astigmatism and is shown in Figure 6.9.

The astigmatism component caused the simulated phase profile to have an axis of symmetry identical to the one shown in Figure 6.8. Therefore it is reasonable to suppose that the difference in the astigmatism modes of the SAE and GN solutions (see Graph 6.1) will affect the direction of the axis of symmetry in each solution (see Figure 6.7). Given the simulated input phase, both the SAE and GN algorithms were able to retrieve the solution with the correct orientation. This shows that the apparent rotation between the SAE and GN solutions is not caused by some simple fault in the code of either and would suggest that the difference in the astigmatism modes is the likely source of the problem.

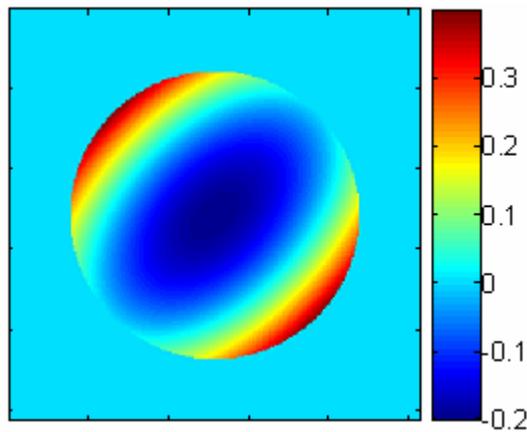


Figure 6.9 A simulated phase profile created using a mixture of defocus and astigmatism. This was created to resemble the experimentally measured SAE phase solution (Figure 6.7(a)) [scale is in waves].

These initial results are encouraging, as they show that the SAE algorithm has retrieved reasonable results (as compared to the trusted GN solution) from real experimental data. However, there are several issues which must be addressed before this sensor system can be used to retrieve the unknown wavefront phase with repeatable accuracies that are comparable (or better than) the GN and Green's function solutions. Normalisation, data conditioning and regularisation will be considered in this chapter and other issues will be identified and held over for future work.

6.4 Normalisation and Data Conditioning

In Chapter 4, when the equations for the SAE algorithm were initially derived, it was shown that the error signal $d(r)$ must be modified (by the factor $4I_0(r)$) to retrieve a solution for the wavefront phase. This is demonstrated by Equation 4.13 which is reproduced here for convenience, the part of (4.13) which describes the PTF has been replaced by the abbreviation for this term introduced in Equation (4.14):

$$\frac{d(r)}{4I_0(r)} = \int d\xi \Phi(\xi) T(\xi) e^{-ir\xi} , \quad (4.13)$$

where $d(r)$, the error signal, is formed by taking the difference of the phase diverse images in the ± 1 diffraction orders and $I_0(r)$ is the intensity of the input wavefront in the pupil plane. The FT of the wavefront phase is $\Phi(\xi)$, and the PTF (which is created by the properties of the diversity function programmed into the grating) is represented by $T(\xi)$ (see Chapter 4 for details). The modification of the error signal by $4I_0(r)$ will henceforth be referred to as data conditioning.

If $I_0(r)$ should tend to zero at any point r then the value of the phase solution at that point does not matter but the effect on neighbouring regions is a cause for concern. The simulations presented in Section 5.3.1.2, where an obscuration was used to artificially set the error signal to zero in a small region, suggest that this will not be a serious problem. When noise is present in the data, which will certainly be the case in any real situation, as $I_0(r) \rightarrow 0$ the noise term will be greatly amplified ($n(r)/I_0(r) \rightarrow \infty$). Regularisation should be used to increase the stability of the data conditioning and this forms the subject of the following section. To minimise the effects of this division by zero requires careful choice and manipulation of the function $I_0(r)$.

When choosing the function to represent the intensity in the pupil plane, $I_0(r)$, care must be taken to preserve the structure of $d(r)$ so that only its overall amplitude is affected by the division ($d(r)/4I_0$). From a data conditioning point of view this division should be limited to the area covered by the signal, to extend beyond this region would be to simply amplify the noise and reduce the stability of the solution. Given the information available to the user, how should I_0 be chosen?

Without additional equipment, which may not even be practically reasonable to include in the optical setup, the user will not be able to measure the intensity in the pupil plane

directly. However, in the detector plane the 0th order image, in the pupil plane configuration, will be an image of the pupil blurred by the optical system. In the ± 1 orders the user has access to images of the blurred pupil (as seen in the 0th order image) which has been further blurred by the addition of the diversity phase.

In Chapter 4 it was shown that the error signal, due to the blurring effect of adding the diversity phase, will always extend further than the support of the wavefront (i.e. it is always larger than the geometric pupil). This is shown schematically in the example shown in Figure 6.10. In this case (based on geometric optics) the input wavefront contains only a defocus term so that the intensity images in the ± 1 orders contain no structure and one is smaller and brighter than the other. This leads to a difference image $d(r)$ as shown, and the cross section of this error signal is also given. The overall width of this signal is set by the size of the largest diffraction order image ('X' in 6.10), which is greater than the width of the pupil (\approx 'Y').

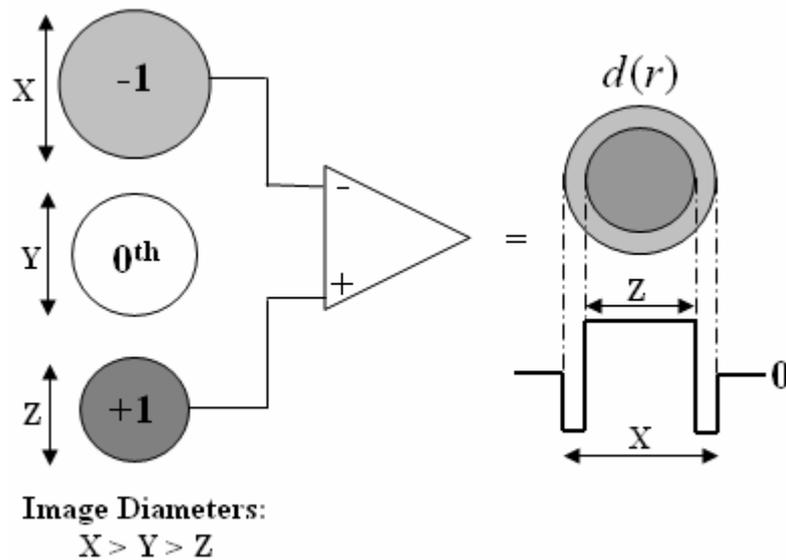


Figure 6.10 A schematic showing the diffraction orders produced in the image plane for a pupil plane configured GPD sensor with defocus phase diversity and a curved input wavefront. The difference of the images in the ± 1 orders is shown, as is its cross section.

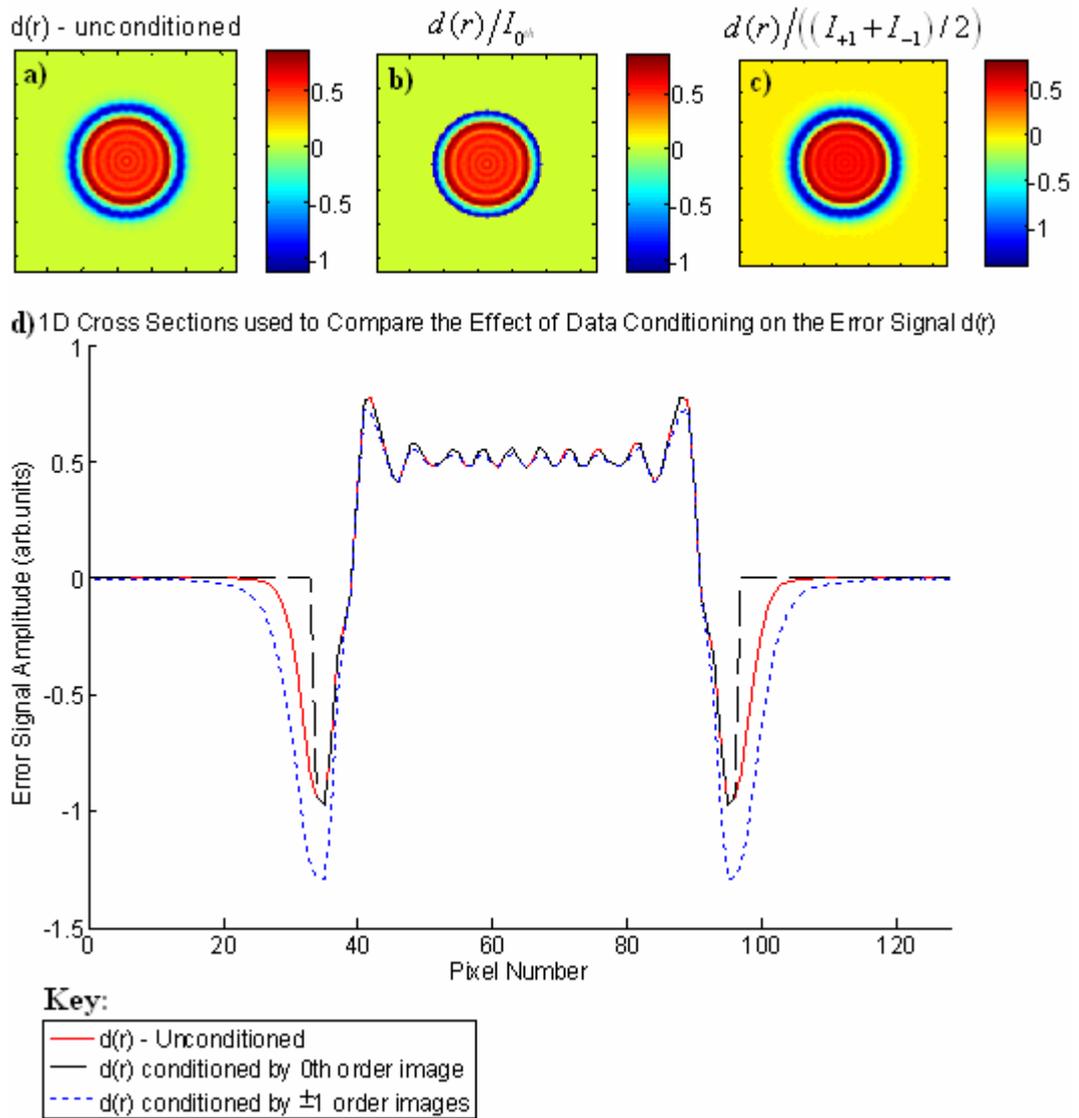


Figure 6.11 (a) the unconditioned error signal $d(r)$. This error signal conditioned by (b) the 0th order image I_0^{th} and (c) the average of the ± 1 diffraction orders $((I_+ + I_-)/2)$ (d) cross sections of the original error signal and the data conditioned versions. [all scales are a measure of contrast].

In Figure 6.11 the effect of the choice of $I_0(r)$ is considered using a simple simulation to illustrate the problem. This figure demonstrates the effect on the error signal $d(r)$ when conditioned by the 2 options for $I_0(r)$; The 0th order image (I_0^{th}) or the average intensity of the two diffraction orders $((I_{+1} + I_{-1})/2)$. In 6.11(a) the unconditioned error signal is shown. The error signals produced by dividing the

unconditioned signal by either $I_{0^{th}}$ or $(I_{+1} + I_{-1})/2$ are plotted in 6.11(b) and (c) respectively. Cross sections of the unconditioned data and both sets of conditioned data are given in 6.11(d). These plots (b-d) show the main difference between these versions of conditioned data is at the boundary of the error signal.

As 6.11(b) shows the effect of conditioning by the 0th order image is truncation of the difference image. This is an undesirable effect which involves loss of information from the boundary of the error signal. In 6.11(c) the error signal conditioned by $(I_{+} + I_{-})/2$ shows that, although this function has the correct width, its structure has altered the error signal at the boundary. In both cases (6.11 (b) and (c)) the change seen in the error signal will alter the phase profile that will be retrieved using the SAE. This figure demonstrates that neither option for $I_0(r)$ is the ideal solution. Instead, it would be better to create a hybrid of the two functions using the smooth top-hat intensity structure of the 0th order image but artificially extended to match the width of the $(I_{+} + I_{-})/2$ function. By doing so the structure of $d(r)$ will be unchanged by its division by $I_0(r)$. Furthermore, the stability of the solution will be increased since this new $I_0(r)$ will limit the manipulation of the data to the region containing only the data, and not the noisy background.

In previous chapters it has been mentioned that, instead of the phase, it may be desirable to reconstruct the imaginary part of the wavefront. The reason for this may be seen here. As Equation (4.14) showed retrieval of the wavefront phase requires that the data be conditioned by $4I_0(r)$ but retrieval of the imaginary part does not. By removing the need to perform data conditioning it should be more straightforward to reconstruct $a(r)$, the difficulty will then lie in converting this to the phase. The stability of the solution will be increased since wherever the intensity tends to zero so too will the imaginary part $a(r)$. Part of the future work for this project will be to investigate this route further.

Finally, in addition to these considerations the data must also be correctly normalised. Normalisation is necessary for several reasons. The first is to mitigate the effects of

etch depth errors in the grating resulting in uneven division of incident intensity into the relevant diffraction orders. With careful design and manufacture such errors should be very small. Secondly, and more importantly, it must be ensured that simply changing the exposure time to increase/decrease the intensity on the detector (without making any other changes to the input or system) does not result in a different phase being retrieved. This second function is performed, in part, by using the pupil plane intensity $I_0(r)$.

6.5 Regularisation

In Chapter 4 the requirement for proper regularisation was identified in Equation 4.16, reproduced here.

$$\frac{D_i(\xi) + N_i(\xi)}{T(\xi)} = \Phi(\xi) \quad (4.16)$$

Where $N_i(\xi) = \mathfrak{F} \left[\frac{n(r)}{4I_0(r)} \right]$,

where $D_i(\xi)$ is the FT of the modified error signal ($D_i(\xi) = \mathfrak{F}\{d(r)/4I_0(r)\}$), $N_i(\xi)$ is the FT of the additive noise in the error signal, $T(\xi)$ is the PTF and $\Phi(\xi)$ is the unknown wavefront phase. As discussed in Chapter 4 this equation demonstrates the existence of a classic regularisation problem within the SAE. While $D_i(\xi) \rightarrow 0$ wherever the PTF $T(\xi) \rightarrow 0$, the noise term will not. Thus $N_i(\xi)/T(\xi) \rightarrow \infty$ as $T(\xi) \rightarrow 0$.

In all previous chapters the data presented has been simulated, and contained no noise. As such sufficient accuracy in the phase solution was obtained by simply discarding the solution at points where the PTF is known to be zero. A ‘threshold’ constant, close to zero, was defined and at points where the modulus of the PTF fell below this value the phase solution was not calculated. This method however will not be suitable in real situations containing noise.

In this experiment all noise is assumed to be additive and, as such, independent of the data [6]. Modelling of multiplicative noise sources is a complicated process, and beyond the scope of this project. By assuming all noise to be simply added to $d(r)$ the system will be easier to analyse. In the experimental data presented in this chapter the only attempt thus far to reduce the noise was to perform background subtraction of the data images prior to their inversion. The ‘threshold constant’ method of regularisation was used to obtain the SAE results presented in Figure 6.7(a). Without this the SAE was unable to retrieve a solution at all since the signal was completely lost beneath huge noise spikes at high spatial frequencies. A more strict regularisation method is required, and for this the Wiener filter was chosen as the most suitable option.

The Wiener filter is a well known method for retrieving an estimate of some original signal based on a measured signal that has been corrupted by noise [7]. The presence of the noise in the measured signal means it is impossible to ever retrieve the original signal perfectly, but with a optimum filter it should be possible to obtain a good approximation. There are many methods for deriving such filters, and in the case of the Wiener filter it is considered ‘optimum’ when the mean squared error between the measured signal and the input signal is minimised [8].

For regularisation of the experimental data a Wiener filter was constructed based on the equation put forward by Champeney [9], which is given in Equation (6.1):

$$H(\omega) = \frac{P_S(\omega)}{P_S(\omega) + P_N(\omega)} , \quad (6.1)$$

where $H(\omega)$ is the transfer function of the Wiener filter, $P_S(\omega)$ is the power spectrum of the signal (in this case the data image) and $P_N(\omega)$ is the power spectrum of the noise (in this case the background image). The filter passes frequencies where power in the signal dominates the power in the noise and attenuates frequencies where the power in the noise exceeds that in the signal. The power spectrums $P_S(\omega)$ and $P_N(\omega)$ were calculated by taking the modulus squared of the FT of the data image and background image respectively. The filter is applied in Fourier space by first taking the FT of the

original image and then multiplying it by the filter $H(\omega)$. The inverse FT of this product produces the Wiener filtered image. Figure 6.12 shows the result of Wiener filtering the data image used to compute the phase solutions (using the SAE and GN algorithms) that are presented in Figure 6.7. In both the images shown in Figure 6.12 the Matlab colour scheme was chosen to clearly show the noise effects in the original image which is why these figures have a different appearance to all other data and results in this thesis. In Matlab the colour scheme is linked to the scaling of the plot, and in this case the scale has no relevant meaning and has been omitted. This data also shows that many diffraction orders are visible at the camera plane, and the ± 1 order images are cropped from this larger image as part of the data processing. In 6.12(a) the original data image is shown and the striped effect of the noise is readily observed. It was found, much later, that this striped effect was caused by defects in the protective glass plate covering the CCD sensitive chip. Figure 6.12(b) shows the same data image after the application of the Wiener filter. This image shows that the noise has been very effectively removed producing a much cleaner image. This newly filtered image was input into the SAE algorithm to study the effect of removing the noise on the phase solution.

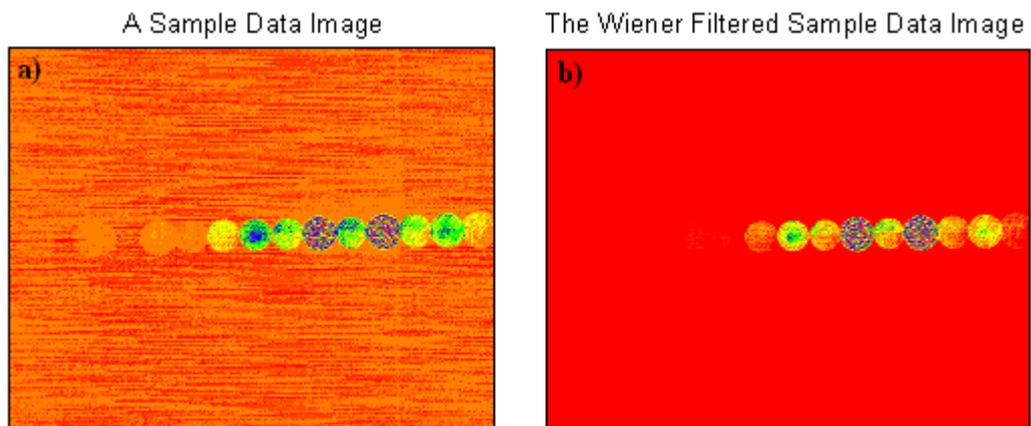


Figure 6.12 A sample experimental data image before (a) and after (b) Wiener filtering to remove the effects of additive noise.

SAE analysis of the Wiener filtered data showed some small differences in the profile of the retrieved phase. In Figure 6.13 the difference between the SAE phase solutions for the filtered and unfiltered data images is shown (the phase solution for the unfiltered data is shown in Figure 6.7(a)). In 6.13(a) this ‘difference phase’ is shown to have a large tilt component. In 6.13(b) the same difference phase is shown, but with the tilt removed so that the changes to other aberration modes may be seen more easily.

To quantify the changes in the retrieved phase caused by the Wiener filtering of the data, the ‘difference phase’ profile (6.13(a)) was decomposed into a series of Zernike polynomials. The coefficients of these are plotted in the histogram in Graph 6.2. Note that these values represent the change in each of the coefficients when comparing the SAE phase solution computed for unfiltered and Wiener filtered versions of the same dataset.

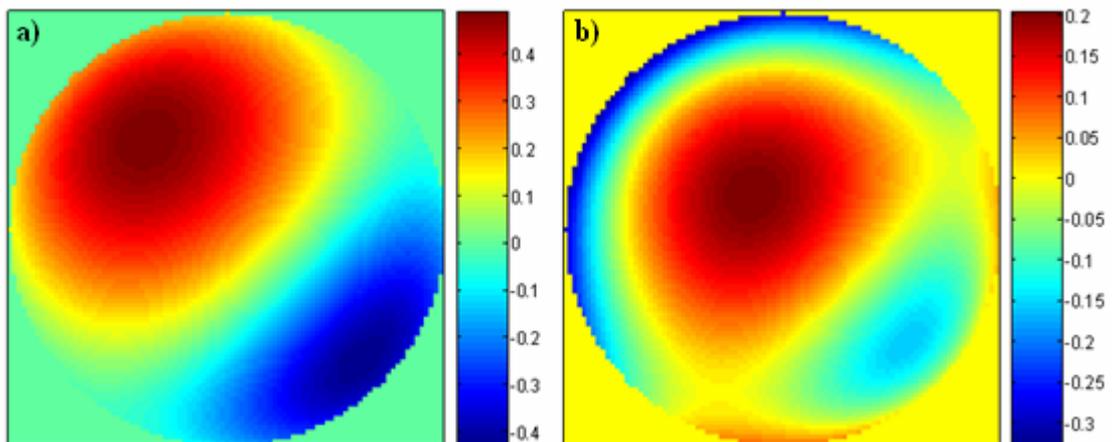
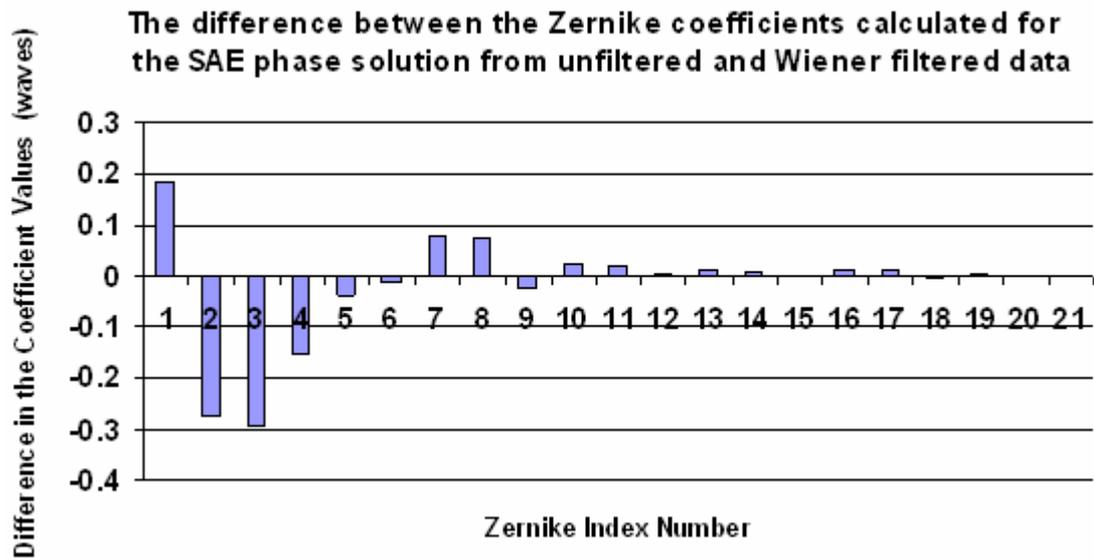
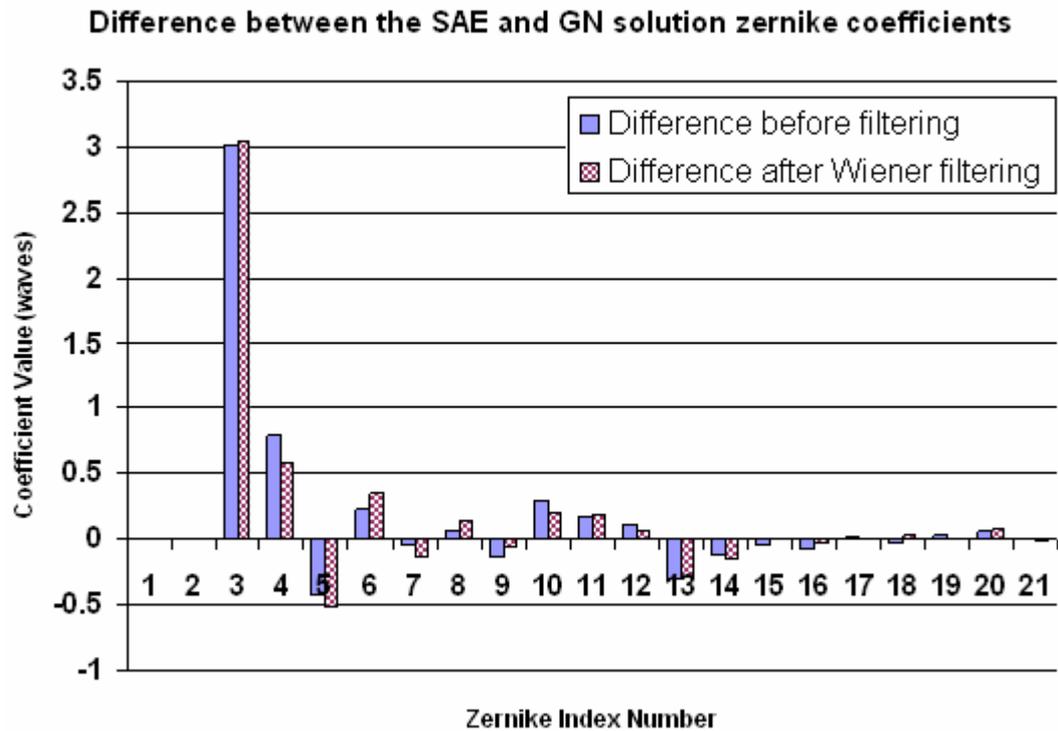


Figure 6.13 The difference between the SAE retrieved phase profiles for unfiltered data and the Wiener filtered data with (a) tilt included and (b) tilt removed. [scale is in waves].



Graph 6.2 Histogram showing the difference in Zernike coefficients for the SAE solutions calculated from both the original data and the Wiener filtered data (unfiltered – filtered).

The histogram in Graph 6.2 shows that the main Zernike coefficients affected by the filtering process are (in order largest-smallest change): tip/tilt (labelled ‘index numbers’ 2 and 3), piston (1), defocus (4) and coma (7 and 8). Refer to the Appendix in Section 6.9 for a full list of the Zernike equations used in this chapter, and the ‘index numbers’ these relate to. The newly computed Zernike decomposition of the SAE phase solution from Wiener filtered data was then compared to the GN algorithm equivalents. As in Graph 6.1, the differences between the SAE and GN solution coefficients were calculated. This was done for the solutions from both the original (unfiltered) and Wiener filtered datasets. This comparison will give an indication of the effect of filtering the data on improving the fit of the solutions from the two algorithms. This information is plotted in Graph 6.3.



Graph 6.3 Histogram depicting the differences between the SAE and GN solution Zernike coefficients, both before and after Wiener filtering. Since the GN algorithm does not calculate piston or tip (index numbers 1 and 2) no values for the difference are plotted for these coefficients.

Graph 6.3 shows there is an even divide between the number of modes whose fit was improved (i.e. the difference was reduced) and those modes whose fit worsened after filtering. In most cases the change in the fit was very slight. One point to note is that one of the largest improvements was seen in the defocus mode, which in this experiment is the mode of greatest interest. These results appear to suggest that most improvement gained is through the mitigation of weak noise components that affect low-frequency reconstruction only (where $T(\xi)$ is small).

While it is interesting to find that the process of filtering the data has some effect on final phase solution itself it should be noted that this is not the primary function of the regularisation. The most important feature should be to remove noise from the data and in so doing increase the stability of the algorithm. Figure 6.12 shows that the Wiener filter employed here has been most effective in the removal of unwanted noise, and this

will serve to help stabilise the SAE solution. With improved modelling of the noise processes involved in the experimental system it may be that the Wiener filter will prove not to be the optimum filter to use for regularisation but it does appear to have made a significant improvement in this case. Further investigation of the regularisation issues will be part of the future work for this project.

6.6 Treatment of the Boundary – An Ongoing Issue

It has been shown that the SAE, depending on the diversity function used, can have difficulty reconstructing low frequency aberration modes (see Chapter 5 for details). It was believed that this could be part of the cause of the mismatch between the defocus coefficients calculated using the SAE and GN algorithms (as discussed in section 6.3.1). In the example of a defocused wavefront and a defocus diversity function (as we have in this experiment), experience suggests that the Green's function and GN solutions do not suffer from the same accuracy problems as the SAE when given the same data input. There must be something missing within the SAE algorithm which is causing this inaccuracy.

To investigate this further a simple simulation was created to look at the error signal created for a defocused wavefront with defocus as the diversity function. The results of this simulation are plotted in Figure 6.14. In Figure 6.14(a) the FT of the error signal $d(r)$ is shown. The error signal is generated, as discussed in Chapter 4, by the convolution of the input wavefront with the diversity filter function programmed into the grating. The rings of zero's evident in 6.14(a) are therefore caused where the PTF of the diversity function goes to zero. This is demonstrated in Figure 6.14(b), where 1D cross sections of the PTF and the FT of $d(r)$ are shown; the zero points in the PTF match the rings of zeros in the FT of $d(r)$. The interesting feature of these plots is the high peak in the FT of $d(r)$ near the origin. The error signal has recovered quickly from the zero in the PTF at the origin to produce this peak. This would suggest that noise present at low spatial frequencies will also be pushed to high values by this PTF and that the solution will therefore be very sensitive to noise. This would offer one explanation for the SAE's inaccuracies when attempting to reconstruct the wavefront phase from this data. However, the Green's function solution when given the same type

of input data was capable of extremely high accuracies [10]. This raises the question, what does the Green's function solution do, that the SAE doesn't?

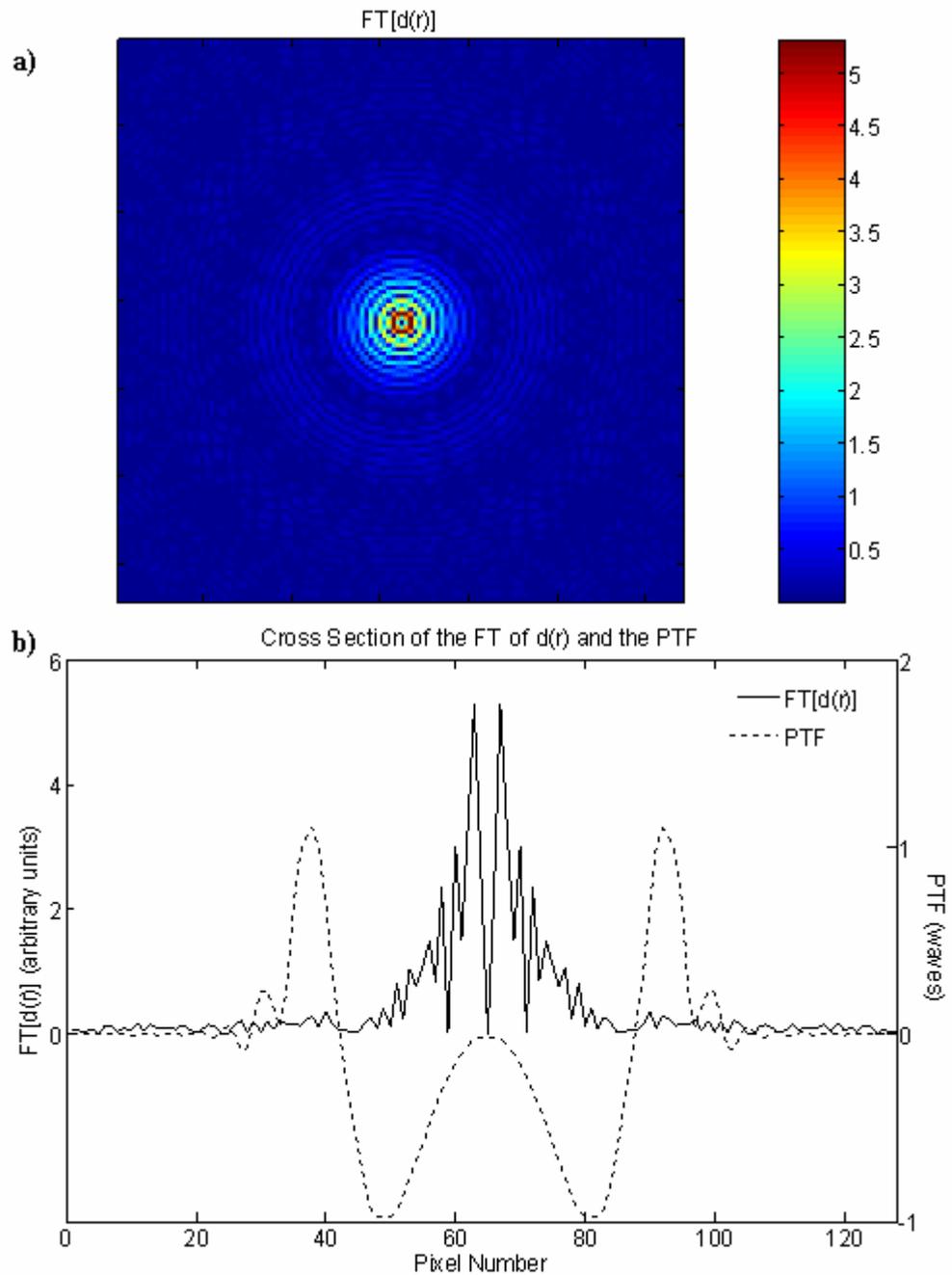


Figure 6.14 a) The FT of the error signal $d(r)$ generated for a defocused wavefront with defocus diversity phase b) a 1D cross section of $FT[d(r)]$ and the PTF for this simulated system.

One major difference between the SAE and Green's function algorithms is that the Green's function includes a treatment of the boundary conditions. It incorporates a slope measurement at the boundaries and can extrapolate these slopes to find the shape of the wavefront. This is demonstrated in Figure 6.15.

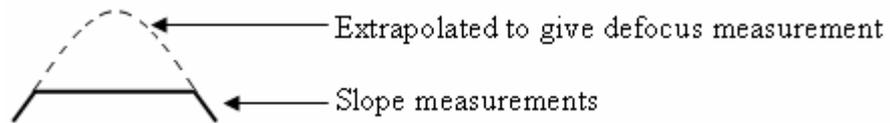


Figure 6.15 Schematic showing how defocus can be extrapolated from measurements of the slope at the boundary of the error signal. This process is included in the Green's function phase retrieval algorithm.

The SAE algorithm does not include any special treatment of the boundaries, perhaps further investigation of the boundary conditions would allow the accuracy of the SAE to be increased. The treatment of the boundary is at least partly tied in with the choice of $I_0(r)$ for the data conditioning. This is an issue which will require a lot of work to address and is an ongoing issue that will be part of the future research for this project.

6.7 Applications of the New Sensor System

Throughout this thesis the concept of GPD, its operation, and the inversion of the resulting data using the SAE algorithm has been presented. The GPD wavefront sensor, in combination with the SAE algorithm, forms a new wavefront sensing system that could potentially be applied to a great number of situations. This system was developed as part of a larger collaborative project (OMAM – Optical Manipulation And Metrology) involving several industrial partners, Heriot-Watt University and UCL. Each partner joined OMAM in the hope of applying the GPD sensor system to their own particular applications. To conclude this chapter and put the work of this thesis into context these applications will be outlined briefly in this section.

6.7.1 Military Imaging and Free-Space Communications

Industrial collaborators Selex (formerly BAE Systems) and my colleague Clare Dillon are interested in using the GPD sensor system for applications involving horizontal propagation paths. In astronomy applications the object of interest is at a great distance from the observer and therefore the turbulence of the Earth's atmosphere may be considered to be relatively close to the receiver. In military imaging and free-space communications, where horizontal propagation paths are often used, turbulence will affect the entire path of the beam and both the object and receiver are embedded in turbulence. Anisoplanatism and scintillation effects have a much more detrimental effect on horizontal path systems than vertical path systems (like astronomy) [11]. There have been many studies of the effects of anisoplanatism on laser beam propagation through turbulent media [11-16], mostly involving measurements of atmospheric properties such as the C_n^2 parameter. Where it is hoped the GPD sensor system can make a novel contribution is in the simultaneous measurement of scintillation and phase information.

Operating the GPD sensor in the pupil plane configuration an image of the pupil lens can be obtained in the 0th order image. Where scintillation is present at the pupil this will also be recorded in the 0th order and this image can be used to retrieve information about the scintillation statistics. Figure 6.16 shows how the GPD sensor could in principle be modified to retrieve phase information from several sources simultaneously. This figure shows how light from two separate sources propagates through the GPD pupil plane sensor. A Fresnel biprism is placed at the focal plane of the pupil lens L_1 such that the images of each source are incident on different facets of the prism. The result of this is that after propagating through the camera lens/grating combination (L_2) the phase diverse data from each source is spatially separated on the CCD camera. This allows the simultaneous measurement of the turbulence induced phase fluctuations of the two propagation paths.

With this system it should be possible to measure both the scintillation statistics and phase fluctuations of multiple propagation paths. With this information this system

could be incorporated into a multi-conjugate adaptive optics system for military imaging and free-space communications.

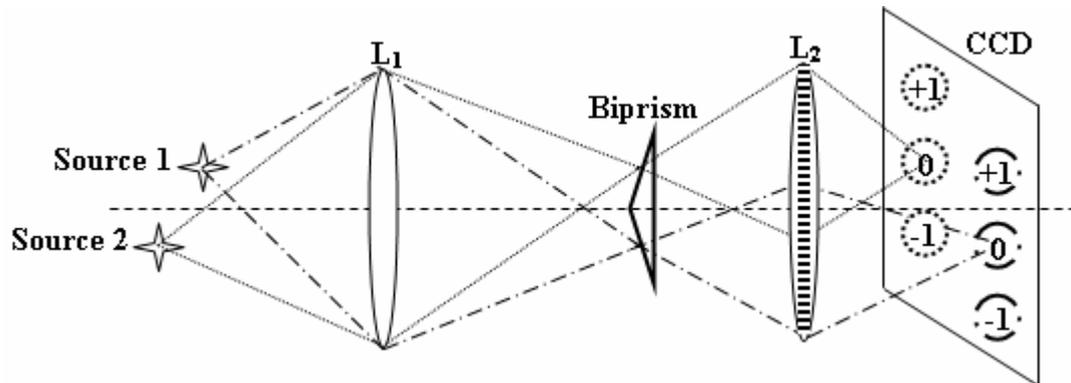


Figure 6.16 A modification of the pupil plane GPD sensor to measure the turbulence induced phase fluctuations along two different propagation paths [11].

6.7.2 Thin Film Metrology

Industrial partners Scalar Technologies and my colleague David Faichnie are interested in incorporating the GPD sensor into an inline thin film metrology system. This system is intended to measure the thickness of thin films, and multilayered laminate structures, while simultaneously monitoring the shape of the film surface using the GPD wavefront sensor.

The ability to measure optical surface properties is of critical importance in a wide range of industries; from microelectronics and semiconductor manufacture to optical coatings and biomedicine. With increasing process complexity comes the demand for more accurate, flexible, compact and reliable metrology solutions for incorporation into inline manufacturing processes [17]. These are often required to operate in real-time and be compatible with the production environment. This is in stark contrast to many previous technologies where samples could be taken to sophisticated test labs where conditions could be controlled and factors such as the size of the instrument and the time taken to make the measurement were of much less importance. Optical metrology,

as an intrinsically non-contact method, is an attractive option for inline surface measurements. There are a number of technologies currently available that offer real time analysis of thin films in the production environment. These include scatterometers, profilometers, ellipsometers and reflectometers [17, 18]. The metrology system being developed by Scalar and David, which will include the GPD wavefront sensor, will offer the added benefit of being capable of taking thickness and surface profile measurements simultaneously. As a non-contact method it would also be possible to use this to measure wet surfaces (the fluid/wetting film would appear as a separate layer) [19, 20].

The GPD sensor will be used as an add-on to the already working thickness monitor. The principle of the thickness measurement is a simple one, using the separation on the CCD of Fresnel reflections from several layers of laminated structure to determine the thickness of each layer. A point source is focussed by a lens to a spot approx $2\mu\text{m}$ in width, onto the surface of the laminate structure. At the air-structure interface some light is reflected while the remainder is transmitted into the sample. This process continues through the sample, with a portion of the light being reflected by each interface. These reflections are focussed onto the detector using a second lens. It has been shown that there is a linear relationship between the thickness of the layers and the separation of their reflections. Scanning the sample allows the thickness to be profiled across the width of the sample [19]. To incorporate the GPD sensor each of the Fresnel reflections is considered to be a separate source and, similar to the multi-source concept shown in Figure 6.16, multiple sets of diffraction orders will be detected on the CCD. These sets of data can then be analysed to retrieve surface profile information of each particular laminate layer at the point the reflection was generated.

Studies are continuing to look at the effects of material refractive index on the thickness measurement and on the aberrations induced by the thin film sample itself. Initial modelling of the system suggests that the most critical aberrations induced by the film are tilt, astigmatism, coma and defocus. With further work on the GPD sensor it may be possible to design a grating which is optimised for accuracy in sensing these aberrations in particular. Initial results are promising and suggest that this system is an attractive option for a real time in-line industrial thickness and surface shape sensor [19, 20].

6.7.3 Designer Beam Shaping of Femtosecond Lasers

In collaboration with the Ultrafast Optics group at Heriot-Watt University, the GPD sensor is being considered for inclusion in a system to alter, and monitor, the spatial properties of femtosecond laser pulses.

The ability to adaptively alter the spatial and temporal properties of laser beams is a highly desirable technology with significance in a diverse range of research and industrial applications. Control of the spatial characteristics of laser beam and pulses could be used, for example, to machine a specific shape of hole while manipulation of the temporal properties could allow the depth of the machined hole and the extent of the heating of the surrounding material to be carefully controlled.

Industrial and Market drive to produce faster computers, using ever smaller circuit components, has pushed imaging techniques to their limits and has led to a need for accurate laser beam shaping for efficient high quality results. This is just one, albeit very important, industrial application of laser beam shaping. Applications in laser printing, materials processing, optical data storage, optical metrology and fibre injection are just a few examples where laser beam shaping can be used to improve results [21].

To alter the spatial properties of a laser beam, shaping techniques generally fall into 3 broad categories; aperturing, field mapping and beam integration. In aperturing the beam is expanded and then apertured to select only a smooth portion of the beam. This method however does mean that light is lost from the system, and it can be difficult to find an appropriate area of the expanded beam to aperture off. Beam integration is best used with relatively incoherent lasers, like Excimers, and involves splitting the beam into beamlets by use of a lenslet array that applies phase errors to each beamlet. These beamlets are recombined in an image plane by a focussing lens, and by controlling the phase errors that are applied the shape in the output plane can be manipulated. This method is sensitive to diffraction effects and in some cases, alignment errors. Finally, field mapping involves insertion of a phase element to modify the phase of the laser beam and a focussing lens to scale the output beam. Like beam integrators this is also

prone to diffraction and alignment errors but both these methods are more efficient than aperturing in terms of light loss [21].

Temporal control of Ultrafast laser pulses is an important technique for the optimisation of materials processing [22]. Adaptive control of the temporal profile would be ideal when machining a specimen containing several different materials, such as brittle dielectrics, which respond differently to high energy input [23]. Temporally altered pulses can be generated for use in laser fusion experiments [24], and also for femtochemistry [25]. These are just a few examples of what is a very fertile area of cross-discipline research.

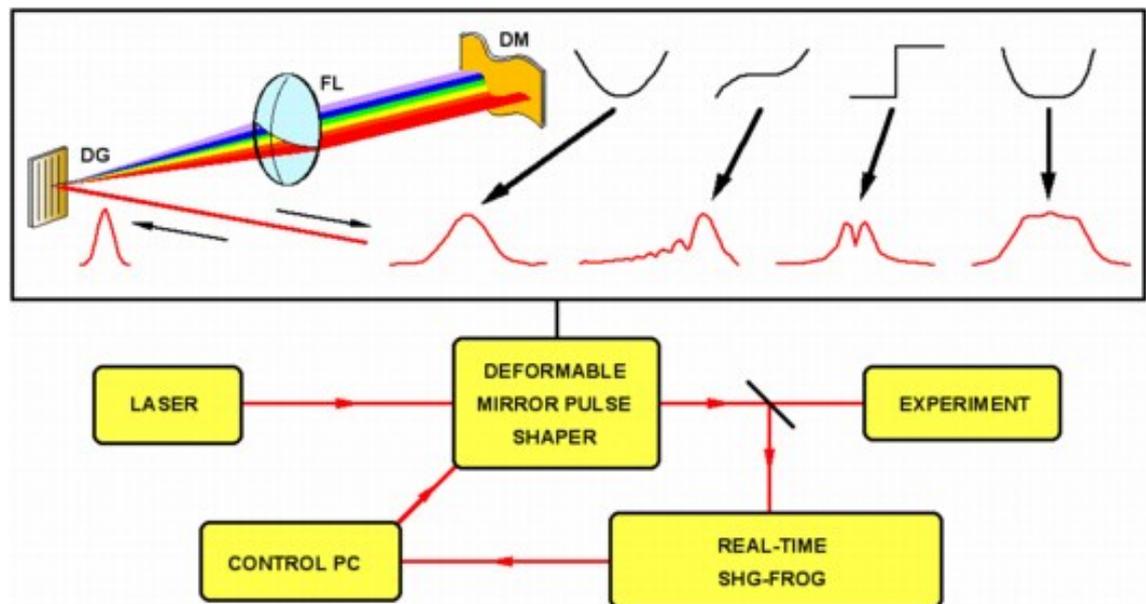


Figure 6.17 A possible field mapping system to alter the spectral phase, and therefore the temporal profile, of a femtosecond laser pulse. Examples of pulse shapes are shown in red, and the corresponding DM surface shape is sketched in black [26].

Ultrafast laser pulses can be shaped temporally if the phase profile of their constituent wavelength components can be modified [27]. Figure 6.17 shows how this might work in principle. A femtosecond pulse from a Ti:Sapphire laser is reflected off a diffraction grating which acts as a dispersive element, separating out the component wavelengths in the pulse. A deformable mirror (DM) is placed at the Fourier plane of the lens FL. By

altering the surface shape of the DM the user can selectively retard/advance separate wavelengths by altering the path length of that wavelength. In the past DM's have been demonstrated to be useful for femtosecond laser pulse compression using similar methods [28]. The altered wavefront is reflected off the mirror, and focussed by FL to overlap all the wavelengths on the grating which recombines them. What exits the system is a pulse whose temporal profile has been altered by manipulation of the spatial phase of the dispersed pulse [27, 29]. The Real Time SHG-FROG system (Second Harmonic Generation – Frequency Resolved Optical Gating) is very sensitive and provides full characterisation of the generated ultrashort pulse, by measuring the intensity and the phase of the spectral and temporal components [30]. Examples of the temporal pulse shapes which could in principle be obtained are illustrated by the red curves in Figure 6.17. The black curves correspond to an estimate of the mirror shape which would be required to generate these pulse shapes. The question is, how could a GPD wavefront sensor and the SAE algorithm be incorporated to be of use in this system?

In Chapter 4 (Section 4.6.4) an iterative extension to the SAE algorithm was used to refine the retrieved phase profile solution to match the experimental data. A natural progression for this method would be to try matching the retrieved phase PSF data to a set of 'target' data (instead of measured data). If these target data (spatial intensities) were chosen to match the data which would be produced by a particular beam shape then this algorithm could provide the information needed by the DM to alter the spectral phase of the laser beam. If the phase profile required to impose the correct spatial phase changes onto the dispersed beam could be calculated by the SAE, a 'recipe book' of phase profiles and beam shapes could be constructed. This recipe book would include details of the phase shape required to alter the beam sufficiently, and the electrode potentials which should be applied to the mirror to create this shape (which would have to be calculated). Since the SHG-FROG system is capable of monitoring the temporal and spatial phase profile of the beam there is little need for a GPD wavefront sensor to be included in this system, although one could be added using another beam splitter to take some of the light returning from the DM (before it reaches FL). Carefully selecting a diversity filter function which can operate with relatively poor signal to noise would mean that the beamsplitter reflectivities could be chosen to split off only a small portion of the light. The GPD wavefront sensor would be able to monitor the spatial phase

properties of the reflected beam from the DM, leaving measurement of the temporal properties to the SHG-FROG system.

6.7.4 Polishing Applications: Manufacturing Telescope Mirror Segments and Prosthetic Joints

Industrial collaborators Zeeko, who specialise in precision optical polishing, and National Physical Laboratories (NPL) working in conjunction with UCL plan to mount the GPD sensor on a swing-arm profilometer for in-situ real-time metrology of samples during the polishing process. Their interest lies particularly in the accurate polishing of ‘free-form’ objects which they define to be ones that deviate significantly from being simple rotationally-symmetric forms [31]. Under this definition free-form surfaces include aspherics, off-axis paraboloids and hyperboloids which, for example, could be polished mirror segments for next-generation Extremely Large Telescopes (ELT’s). Also of particular interest to Zeeko is the polishing of prosthetic knee and hip joints which are highly complex free-form surfaces which require very high levels of precision in surface shape.

The next generation of telescopes, with primary diameters of 30-100m, will rely on segmentation to achieve such large mirror diameters. This has been discussed in previous chapters from the point of view of segment co-phasing, which presents a significant challenge. The manufacture of the huge number of segments required to build these telescopes is an essential factor in the success of building such large telescopes. The European Southern Observatory (ESO) is currently working on the Overwhelmingly Large (OWL) telescope which will have a 100m segmented mirror containing 3048 segments, and also a segmented secondary mirror requiring a further 216 segments. In comparison the twin Keck telescopes, currently the worlds largest, comprise just 36 segments in total [32, 33]. To build the OWL to schedule will require the production of 1.5 segments on average per day [34]. This is not outwith the scope of current industrial capabilities but the ability to monitor the surface shape in real-time as the segments are polished could give significant savings in both time and money.

Zeeko and NPL are also interested in the precision polishing of prosthetic hip and knee joints, which is a challenging application requiring very high accuracies in surface shape. Hip replacement procedures are becoming increasingly commonplace as medical technology improves and life expectancies increase. Over 30,000 operations are performed per year in England alone usually on patients over the age of 50 [35]. For older patients the replacement joint will usually last for the rest of their life, but in younger patients revision surgery can be necessary when the first joint fails. This surgery is more complex, more expensive, and failure rates are much higher [35]. The non-surgical alternative to hip replacement (medication) is extremely expensive over the lifetime of the patient and may leave the patient debilitated or disabled and in considerable pain. It is therefore of vital importance that prosthetic joints are manufactured to last as long as possible, to reduce costs and the likelihood of a second replacement operation. It is believed that the typical lifetime of a knee or hip prosthesis, which is currently around 10-15 years, is limited by the precision or the surface shape and quality [31]. Due to their function hip and knee joints are in constant use and under considerable pressure, this causes wear of the articulating joint surfaces over time. This wear is also dependent on the level of activity of the patient, which is another reason why younger patient's replacement joints do not last as long. If during manufacture even micron-level high spots are left on the surface pressure will be concentrated in these regions resulting in a higher wear rate and in some cases fracture of the joint [31, 36]. Debris particles from the joint as it is worn down are not tolerated by the human body and cause normal cell growth around the joint to cease. It can also create inflammation, bone reabsorption, loosening of the joint and severe pain. In some cases this debris has been linked to cancer at the site of the joint replacement [37]. Increasing the precision with which these joints are polished is of vital importance for both economic and medical reasons. Extending the lifetime of a prosthetic joint by just a few years could mean the difference between a young patient in their 30's needing just one revision replacement instead of two [36].

By incorporating the GPD sensor into their polishing system Zeeko plan to improve the precision and efficiency of the polishing process by performing real-time metrology of the form under polish. This will help to save time and remove the added errors caused by demounting the form for testing and then remounting it for further polishing. The

simple and lightweight design of the GPD sensor will allow it to be easily mounted on the end of the swing arm profilometer.

6.8 Conclusions

In this chapter the first experimental results from the GPD SAE sensor system have been presented. It is often not a trivial process to turn a system which has worked well in simulation into one which works just as well in practice. This has been demonstrated here, and some of the practical considerations and modifications necessary to get a working experimental version of this system have been given. There is however much more work required before this is a simple ‘plug and play’ wavefront sensing system and this will be discussed briefly in the concluding chapter.

The initial results from the GPD sensor are encouraging. While it is not at this stage achieving accuracies comparable to the GN or Green’s function solutions it is capable of retrieving the correct wavefront shape, with errors in the scaling of the solution. These scaling errors may in part be linked to the regularisation and stability of the algorithm as well as to treatment of the boundary conditions. These are ongoing issues to be tackled.

It has been shown in this chapter that use of a Wiener filter greatly reduced the amount of noise in the experimental data, which will help to stabilise the solution. In time it may prove that the Wiener filter is not the optimum filter to use for this system, and this will require more sophisticated modelling of the noise processes involved. However, as an initial approach the Wiener filter appears to work well.

To conclude this chapter, and the main body of this thesis, the GPD sensor system was considered within the context of the OMAM collaboration and the industrial partners who plan to use it. From free-space communications to the manufacture of prosthetic joints there is a wide range of applications which will benefit from this new wavefront sensor within this collaboration alone. It is hoped that, with further development, this wavefront sensing system will be implemented in a range of practical situations and will

prove to be an attractive and competitive technology for currently available wavefront sensors technologies.

6.9 Appendix: List of Zernike Polynomials Used

In this appendix details of the Zernike polynomials used in Chapter 6, and their corresponding index numbers which are used as simple identifiers for each of them, are tabulated. These polynomials, as with all the Zernikes used in this thesis have been taken from the definitions given by Wavefront Sciences [38].

Index No.	Notation	Equation (Polar Form)	Description (where applicable)
1	Z_0^0	$1/\sqrt{\pi}$	Piston
2	Z_1^1	$\rho \cos(\theta)$	Tilt (about x-axis), a.k.a. Tip
3	Z_1^{-1}	$\rho \sin(\theta)$	Tilt (about y-axis)
4	Z_2^0	$2\rho^2 - 1$	Defocus
5	Z_2^{-2}	$\rho^2 \sin(2\theta)$	Astigmatism (axis at $\pm \pi/4$)
6	Z_2^2	$\rho^2 \cos(2\theta)$	Astigmatism (axis at 0 or $\pi/2$)
7	Z_3^{-1}	$(3\rho^3 - 2\rho) \sin(\theta)$	3 rd order Coma (along x-axis)
8	Z_3^1	$(3\rho^3 - 2\rho) \cos(\theta)$	3 rd order Coma (along y-axis)
9	Z_3^{-3}	$\rho^3 \sin(3\theta)$	Triangular Astigmatism (base on x-axis)
10	Z_3^3	$\rho^3 \cos(3\theta)$	Triangular Astigmatism (base on y-axis)

Index No.	Notation	Equation (Polar Form)	Description (where applicable)
11	Z_4^0	$6\rho^4 - 6\rho^2 + 1$	3 rd order Spherical Aberration
12	Z_4^2	$(4\rho^4 - 3\rho^2) \text{Cos}(2\theta)$	
13	Z_4^{-2}	$(4\rho^4 - 3\rho^2) \text{Sin}(2\theta)$	
14	Z_4^4	$\rho^4 \text{Cos}(4\theta)$	
15	Z_4^{-4}	$\rho^4 \text{Sin}(4\theta)$	
16	Z_5^1	$(10\rho^5 - 12\rho^3 + 3\rho) \text{Cos}(\theta)$	
17	Z_5^{-1}	$(10\rho^5 - 12\rho^3 + 3\rho) \text{Sin}(\theta)$	
18	Z_5^3	$(5\rho^5 - 4\rho^3) \text{Cos}(3\theta)$	
19	Z_5^{-3}	$(5\rho^5 - 4\rho^3) \text{Sin}(3\theta)$	
20	Z_5^5	$\rho^5 \text{Cos}(5\theta)$	
21	Z_5^{-5}	$\rho^5 \text{Sin}(5\theta)$	

6.10 References

1. Dalgarno, P.A., *Performed Measurement of the Readout Noise and Dark Count of the Retiga Ex Qimaging Camera*. Personal Communication 2006
2. Hecht, E., *Optics*. 3rd ed. 1997: Addison Wesley Publishing Co.
3. Yariv, A., *Optical Electronics in Modern Communications*. 5th ed. 1997: Oxford University Press.
4. Born, M. and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. 7th (expanded) ed. 2001: Cambridge University Press.
5. Campbell, H.I., S. Zhang, A. Brun, and A.H. Greenaway. *Generalised Phase Diversity Wavefront Sensing*. in *AMOS Technical Conference*. 2004. Wailea,

Maui: Published on CD, contact amostech@maui.com or visit www.maui.afmc.af.mil for a copy

6. Storey, N., *Electronics - a Systems Approach*. 2nd ed. 1998: Prentice Hall.
7. Jan, J., *Digital Signal Filtering, Analysis and Restoration*. 1st English translation ed. Iee Telecommunications Series, ed. G. White. Vol. 44. 2000, London: The Institution of Electrical Engineers.
8. Cooper, G.R. and C.D. McGillem, *Probabilistic Methods of Signal and System Analysis*. 3rd ed. 1999, New York: Oxford University Press.
9. Champeney, D.C., *Fourier Transforms and Their Physical Applications*. 1st ed. Techniques in Physics. 1973: Academic Press.
10. Djidel, S. and A.H. Greenaway. *Nanometric Wavefront Sensing*. in *3rd International Workshop on Adaptive Optics in Industry and Medicine*. 2002: Starline Printing Inc. p. 213-219
11. Dillon, C.E., H.I. Campbell, S. Zhang, and A.H. Greenaway. *Anisoplanatic Atmospheric Measurements for Terrestrial Imaging*. in *The 5th International Workshop on Adaptive Optics in Industry and Medicine*. 2005. Beijing, China: Proc. SPIE. Vol. 6018. p. 60180W-1 60180W-8
12. Fried, D.L., *Anisoplanatism in Adaptive Optics*. J.Opt.Soc.Am.A, 1982. **72**(1): p. 52-61.
13. Johnston, R.A., N.J. Wooder, F.C. Reavell, M. Bernhardt, and J.C. Dainty, *Horizontal Scintillation Detection and Ranging $C_n^2(Z)$ Estimation*. Applied Optics, 2003. **42**(18): p. 3451-3459.
14. Travis, M.T. and H.T. Yura, *Scintillation Effects on Centroid Anisoplanatism*. J.Opt.Soc.Am.A, 1987. **4**(1): p. 57-59.
15. Vernin, J. and F. Roddier, *Experimental Determination of Two-Dimensional Spatiotemporal Power Spectra of Stellar Light Scintillation. Evidence for a Multilayer Structure of the Air Turbulence in the Upper Troposphere*. J.Opt.Soc.Am.A, 1973. **63**(3): p. 270-273.
16. Wilson, R.W., N.J. Wooder, F. Rigal, and J.C. Dainty, *Estimation of Anisoplanatism in Adaptive Optics by Generalized Scidar Profiling*. Monthly Notices of the Royal Astronomical Society, 2003. **339**: p. 491-494.
17. Kincade, K., *Optical Surface Analyzers Become Precision Manufacturing Tools*. Laser Focus World, 2005. **41**(2): p. 89-93.
18. Braun, A.E., *Metrology Develops to Measure Thinner Films Better*. Semiconductor International, 2004(January Issue).
19. Faichnie, D., I. Bain, and A.H. Greenaway. *Thickness Measurement and Surface Profiling Using Principles of Wavefront Sensing*. in *The 5th International Workshop on Adaptive Optics in Industry and Medicine*. 2005. Beijing, China: Proc. SPIE. Vol. 6018. p. 60180T-1 60180T-11
20. Faichnie, D., A.H. Greenaway, K. Karstad, and I. Bain, *Thin Film Metrology Using Modal Wavefront Sensing*. J.Opt.Soc.Am.A, 2005. **7**(6): p. S290-297.
21. Dickey, F.M., *Laser Beam Shaping*. Optics and Photonics News, 2003(April 2003 issue): p. 30-35.
22. Stoian, R., M. Boyle, A. Thoss, A. Rosenfeld, G. Korn, and I.V. Hertel, *Dynamic Temporal Pulse Shaping in Advanced Ultrafast Laser Material Processing*. Applied Physics A: Materials Science and Processing, 2003. **77**(2): p. 265-269.
23. Stoian, R., et al. *Temporal Pulse Shaping and Optimization in Ultrafast Laser Ablation of Materials*. in *MATERIALS RESEARCH SOCIETY SYMPOSIUM - PROCEEDINGS: Advanced Optical Processing of Materials, Apr 22-23 2003*. 2003. San Francisco, CA, United States: Materials Research Society. Vol. 780. p. 183-194

24. Jackel, S., et al., *Temporal Pulse-Shaping for Laser-Fusion Experiments Using a Cavity-Dumped Q-Switched Oscillator*. 1982. **15**(2): p. 255-259.
25. Mendham, K.J., J.W.G. Tisch, M.B. Mason, N. Hay, and J.P. Marangos, *Control of Laser Heating in Clusters through Variation in Temporal Pulse Shape*. Optics Express, 2003. **11**(12): p. 1357-1364.
26. Reid, D., *Dynamic Pulse Shaping - an Introduction*. 2005.
27. El-Agmy, R., H. Bulte, A.H. Greenaway, and D.T. Reid, *Adaptive Beam Profile Control Using a Simulated Annealing Algorithm*. Optics Express, 2005. **13**(16): p. 6085-6091.
28. Zeek, E., K. Maginnis, S. Backus, and e. al, *Pulse Compression by Use of Deformable Mirrors*. Applied Optics, 1999. **24**(7): p. 493-495.
29. Ramsay, E., *Discussion of the Frog Technique and Other Ultrafast Optics Issues*. Personal Communication 2005
30. Cormack, I.G., *Rapid Techniques for Ultrashort Optical Pulse Characterisation*. Doctoral Thesis 2001. University of St Andrews
31. Walker, D., et al. *First Results on Free-Form Polishing Using the Precessions Process*. in *Free-Form Optics: Design, Fabrication, Metrology*. 2004. North Carolina, USA: ASPE. p. 29-34
32. Medwadowski, S.J., *Structure of the Keck Telescope - an Overview*. Astrophysics and Space Science, 1989. **160**(1-2): p. 33-43.
33. Chanan, G., M. Troy, and C. Ohara. *Phasing the Primary Mirror Segments of the Keck Telescopes: A Comparison of Different Techniques*. 2000: Proc. SPIE. Vol. 4003. p. 188-201
34. Gilmozzi, R., et al. *The Future of Filled Aperture Telescopes: Is a 100m Feasible?* in *Advanced Technology Optical/IR Telescopes VI*. 1998. Kona, Hawaii: Proc. SPIE. Vol. 3352. p. 778-791
35. NHS, *Hip Replacement - Getting It Right First Time*. 2000 Report Number: HC 417 Session 1999-00
36. Davis, P.J., *Artificial Hip Joints: Applying Weapons Expertise to Medical Technology*. Energy and Technology Review, 1994(October Edition): p. 16-21.
37. Brown, B.H., R.H. Smallwood, D.C. Barber, P.V. Lawford, and D.R. Hose, *Medical Physics and Biomedical Engineering*. Series in Medical Physics and Biomedical Engineering. 1998, London: IOP Publishing.
38. WavefrontSciences, *Periodic Table of the Zernike Polynomials*. 2001.

Chapter 7

Conclusions and Future Work

It has been shown that Phase Diversity (PD) wavefront sensing does not have to be limited to the use of defocus as the diversity phase. Any function, or combination of functions, which is complex and whose real and imaginary parts are both non-zero and share the same symmetry may be used to build a null sensor and perform Generalised Phase Diversity (GPD) wavefront sensing. One example is the Phase Contrast Microscope (PCM), a familiar application whose phase shift technique was found to satisfy the conditions for a GPD null sensor [1]. The GPD sensor's sensitivity will be determined largely by the diversity function chosen, and with careful selection the user may find a function which will give optimum sensitivity to the aberrations of most interest in their given application.

The motivation for this project was to design a compact, versatile wavefront sensor capable of providing highly accurate results with continuous, discontinuous and scintillated wavefronts. There is no fundamental problem with using defocus as the diversity phase in any of these cases; the limitations of previous sensing methods have been caused by their data inversion techniques. Defocus PD (DPD) has traditionally been used with the Intensity Transport Equation (ITE) to retrieve the curvature of the input wavefront. It is the assumptions made in solving the ITE which limit the use of the wavefront sensor to continuous, unscintillated, wavefronts [2]. Having established that it is possible to use diversity functions other than defocus to perform PD a new phase retrieval algorithm, not based on the ITE, was required. The Small Angle Expansion (SAE), essentially a deconvolution algorithm, was developed for use with the GPD sensor and could also be used as an alternative to the currently used ITE based applications. The SAE computes an analytic solution to the wavefront phase which has also been shown to work well with an additional iterative refinement stage using the measured data as the 'ground truth' against which the accuracy of the solution is measured.

Chapter 7: Conclusions and Future Work

To perform GPD wavefront sensing all that is required is an error signal created by taking the difference of two images containing equal and opposite amounts of the diversity phase, and knowledge of the diversity function used. The method of applying the PD may be chosen to suit the user's application. Throughout this thesis the GPD sensor presented has been a variation of the diffraction grating based technique proposed by Blanchard *et al* [3-5]. This method was selected for its simple, compact and robust common path design.

The process by which the phase diverse data is formed has been studied, and described as essentially a convolution between the input wavefront phase and a Phase Transfer Function (PTF) which is created using the known properties of the PD function. By studying the form of the PTF the user is able to determine which spatial frequencies that particular diversity function will be most sensitive to, and which it will be blind to. Since the data inversion process is a deconvolution between the measured data and this PTF its structure is also directly linked to the stability of the inverse problem.

It has been shown that the error signal, for all GPD diversity functions, has a sinusoidal response to the size of the phase step in the input wavefront. Therefore the wavefront sensor is fundamentally blind to piston errors of exact multiples of π , and outwith the linear region $\pm\pi/2$ (about the origin) the error signal will become multi-valued and therefore ambiguous. This is a universal problem, not unique to GPD. With ambiguous data no phase retrieval algorithm will be able to compute the correct input phase. Instead it is the corresponding phase from the monotonic $\pm\pi/2$ region which will be reconstructed. The nature of the convolution process which generates the data means the SAE requires the local wavefront slope across the width of the blur function be restricted to small angles, within the monotonic region. This is an important step forward from previous 'small angle' phase retrieval algorithms which have required that the Peak to Valley (PV) of the wavefront itself be limited to small angles for accurate reconstruction [2, 6]. This advantage can be maximised by increasing the sampling (increase the number of pixels/samples across the input wavefront) or reducing the width of the blur function (through careful choice of the diversity function). In this way it would be possible to push this algorithm to deal with extremely aberrated input wavefronts.

Chapter 7: Conclusions and Future Work

The SAE algorithm has been shown to work well with discontinuous wavefronts generated by a simulated segmented telescope pupil. It is believed that this technique would provide a useful tool for the accurate co-phasing of the next generation of Extremely Large Telescopes (ELT). The next stage of this research will be to test this experimentally, using a 7 segment piezo driven mirror and the GPD wavefront sensor. The performance of the SAE with non-uniform illumination was also tested, using a sine wave to modulate the wavefront amplitude and also an artificial obscuration to study how the algorithm copes in the absence of an error signal. In both cases the results were encouraging and part of the future work for this project will be to introduce properly modelled scintillation into the simulations to test the accuracy of the wavefront sensor in these conditions.

Finally the GPD wavefront sensor and SAE algorithm were tested experimentally and the results compared to the Gureyev-Nugent (GN) algorithm, an ITE based solution [7, 8]. The SAE algorithm performed well, retrieving the correct wavefront shape but with scaling errors which it is believed are linked to problems with the treatment of the boundary conditions. Wiener filtering was used to help increase the stability of the solution and reduce the noise in the data. These results, while encouraging, show that there is more work required before this system can be used as a simple ‘plug and play’ sensor capable of measurement accuracies to rival currently favoured techniques. This future work will include more rigorous study of the treatment of the boundary conditions, investigation of Signal to Noise (SNR) issues and work to increase the stability of the solution to noise.

The GPD wavefront sensor, coupled with the SAE algorithm, is a lightweight, versatile and robust sensor system for wavefront metrology. With further development it is planned that this sensor will be incorporated into a number of our collaborators industrial and research applications from military imaging to bioengineering. It is hoped that, in the future, this new wavefront sensor will go on to become a competitive technology compared to currently favoured wavefront sensing techniques.

7.1 References

1. Huang, T., *Discovered That the Phase Contrast Microscope Satisfies the Gpd Null Sensor Conditions*. Personal Communication 2005
2. Woods, S.C. and A.H. Greenaway, *Wave-Front Sensing by Use of a Green's Function Solution to the Intensity Transport Equation*. J.Opt.Soc.Am.A, 2003. **20**(3): p. 508-512.
3. Blanchard, P.M. and A.H. Greenaway, *Broadband Simultaneous Multiplane Imaging*. Optics Communications, 2000. **183**(1-4): p. 29-36.
4. Blanchard, P.M., et al., *Phase-Diversity Wave-Front Sensing with a Distorted Diffraction Grating*. Applied Optics, 2000. **39**(35): p. 6649-6655.
5. Blanchard, P.M. and A.H. Greenaway, *Simultaneous Multiplane Imaging with a Distorted Diffraction Grating*. Applied Optics, 1999. **38**(32): p. 6692-6699.
6. Gonsalves, R.A., *Small-Phase Solution to the Phase-Retrieval Problem*. Optics Letters, 2001. **26**(10): p. 684-685.
7. Gureyev, T.E. and K.A. Nugent, *Phase Retrieval with the Transport-of-Intensity Equation. Ii. Orthogonal Series Solution for Nonuniform Illumination*. J.Opt.Soc.Am.A, 1996. **13**(8): p. 1670-1682.
8. Gureyev, T.E., A. Roberts, and K.A. Nugent, *Phase Retrieval with the Transport-of-Intensity Equation: Matrix Solution with Use of Zernike Polynomials*. J.Opt.Soc.Am.A, 1995. **12**(9): p. 1932-1941.