Analysis of Phase Retrieval from Multiple Images

Graham N Craik

Submitted for the degree of Doctor of Philosophy

Heriot Watt University
School of Engineering and Physical Science
December 2010

The copyright in this thesis is owned by the author. Any quotation from this thesis or use of any of the information contained in it must acknowledge this thesis as the source of the quotation or information.

ACADEMIC REGISTRY
Research Thesis Submission

Name*:*	Graham Craik		
School/PGI:	EPS		
Version: *(i.e. First, Resubmission, Final)*	Final	Degree Sought (Award **and** Subject area)	PhD Physics

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 and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, 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 *(name in capitals):*	Graham Craik
Signature of Individual Submitting:	
Date Submitted:	

For Completion in Academic Registry

Received in the Academic Registry by *(name in capitals):*			
Method of Submission *(Handed in to Academic Registry; posted through internal/external mail):*			
*E-thesis Submitted (**mandatory for final theses from January 2009)***			
Signature:		Date:	

Abstract

This thesis considers the calculation of phase from sets of phase contrast and defocused images.

An improvement to phase contrast imaging is developed that combines three phase contrast images. This method results in a reduction in the phase error by a factor of up to 20 in comparison to using a single image. Additionally the method offers the potential for optimisation and the extension to an arbitrary number of images.

Phase diversity using defocused images is considered in more depth where the intensity transport equation is used to calculate the phase. First a Green's function approach to solving this equation was considered. One of the Green's functions stated in the literature is shown to be incorrect, the other two are shown to be correct both giving equivalent phase estimates. A further improvement is made to this method by removing the singularities in the phase calculation process.

As an alternative to the Green's function solution a Fourier transform approach is also considered. A complete solution to the intensity transport equation is derived with inclusion of the boundary conditions. This completes the method incompletely described in the literature.

Through simulation, generic key factors are identified for the potential optimisation of experimental and numerical process to improve the estimated phase.

Determining 3D structural information of an object from the phase calculated in a single plane is considered using an iterative process. It is shown that this process is limited but can be used, in some cases, to generate an approximate representation of the object.

Contents Page

Abstract ..... **i**

Contents Page ..... **ii**

Figure List ..... **v**

List of Symbols ..... **xi**

List of Phase Functions ..... **xiii**

Chapter 1 – Introduction **1**

1.1. Introduction 1

1.2. General Wavefront Sensing Methods 2

1.2.1. Interferometry 2

1.2.2. Common Path Interferometry 3

1.2.3. Gradient Measurements 5

1.2.4. Phase Diversity 7

1.3. Aims and Objective of this Thesis 9

1.3.1. Comparison of the General Wavefront Sensing Methods 10

1.4. Outline of the Thesis 11

Chapter 2 – Phase Contrast Imaging **14**

2.1. Introduction 14

2.2. Background – Phase Contrast Imaging 16

2.2.1. Physical Description of General Common Path Interferometer 16

2.2.2. Mathematical Representation of an Optical System 19

2.3. Theory 20

2.3.1. Mathematical Description of a Common Path Interferometer 20

2.3.2. Calculating the Phase from a Single Image 22

2.3.3. Calculating the Phase from Multiple Images 25

2.4. Implementation and Evaluation of the Numerical Simulations 27

2.4.1. Numerical Implementation of the Theory 27

2.4.2. Sampling Continuous Functions 28

2.4.3. Maximum Scattering Angle 30

2.4.4. Evaluation of the Calculated Phase 31

2.5. Benefits and Limitations 34

2.5.1. Benefits of the 3 Image Method 34

2.6. Iterative Refinement of the Phase Solution 38

2.7. Further Development of Phase Contrast Imaging 43

2.8. Conclusions 46
Chapter 3 – Theory and Implementation of Defocus Phase Diversity 48
3.1. Introduction 48
3.1.1. Aims and Objectives 48
3.2. Theory 51
3.2.1. Derivation of the Intensity Transport Equation 51
3.2.2. Green's Function Solution 58
3.2.3. Derivation of the First Green's Function 60
3.2.4. Derivation of the Second Green's Function 62
3.2.5. Validity of the Green's Functions 68
3.2.6. Fourier Transform Solution to the Intensity Transport Equation 70
3.3. Numerical Implementation of the Theory 73
3.3.1. Generation of the Simulated Data 74
3.3.2. Description of the Inversion Process using Fourier Analysis 74
3.3.3. Definitions of the Boundary Data 78
3.3.4. Implementing the Green's Function and Removing the Singularities 83
3.4. Conclusions 88
Chapter 4 – Experimental Testing of the Theory 90
4.1. Introduction 90
4.2. Background – Simultaneous Multi-Plane Imaging 90
4.3. Experiment Conducted in the Laboratory 92
4.4. AMO / Wavefront Sciences Inc Experiment 95
4.5. Conclusions 98
Chapter 5 – Improvements and Limitation 100
5.1. Introduction 100
5.2. Experimental Optimisation 100
5.2.1. Resolution 100
5.2.2. Approximating the Intensity, $I(\mathbf{x})$ 103
5.2.3. Improving the Approximation of the Axial Derivative 105
5.2.4. Asymmetric Measurement Planes 109
5.2.5. Summary of Physical Options 111
5.3. Numerical Processing 111
5.3.1. Incorrectly Choosing the Disk Radius 112
5.3.2. Misaligned Data 114
5.3.3. Sampling in Fourier Space 116

5.3.4. Computation Time 117
5.4. Conclusions 119
Chapter 6 – Determining a 3D Object 121
6.1. Introduction 121
6.1.1. Literature Review – 3D Imaging 121
6.1.2. Aims and Objectives 124
6.2. Model of an Object Viewed with a Microscope 125
6.2.1. Simulating the Imaging Process 126
6.3. Iterative Calculation of the Object from the Data 134
6.3.1. Simulations 136
6.4. Calculating a 3D Object from Phase Contrast and Defocused Images 145
6.5. Conclusions 156
Chapter 7 – Conclusions and Future Work 158
7.1. Future Work 161
Appendix A – Second Order Approximation of the ITE 163
Appendix B – Comparison of Terms in the Green's function and Fourier Transform Solution 165
References 166

Figure List

Figure 1.1 Schematic of a Mach-Zender interferometer 3

Figure 1.2 Schematic of a general common path interferometer. The Black lines represent the path of the light when there is no object. 4

Figure 1.3 Example of phase contrast images. a) show the set phase. b) an image generated with no phase shift and c) an image generated with a phase shift of $\pi/2$ radians 5

Figure 1.4 Schematic of a Shack Hartmann wavefront sensor. a) shows the instrument illuminated by a plane wave for calibration. b) shows an illuminated object which modifies the phase and the position of focal points. 6

Figure 1.5 Schematic of the principle of phase diversity using defocused images 7

Figure 1.6 An example of defocused measurements. a) shows the set phase. b) shows the measurement from plane B 8

Figure 2.1 a) A schematic of a common path interferometer. b) shows the physical shape of pieces of glass that can be used to apply a phase shifts to the reference beam, in plane C. The phase shifts are in the same direction. c) shown the shape of two pieces of glass that can be used to give equal and opposite phase shifts. d) shows the division of the area of the reference wave and the scattered wave 17

Figure 2.2 Schematic of phase contrast imaging using conical illumination. The black lines show the path of the light 18

Figure 2.3 Projection of Z_2^0 on to a subset of the Zernike polynomials specified along the *x* axis 32

Figure 2.4 Fit of a set of test functions defined along the x axis to Z_2^0 33

Figure 2.5 RMS error plotted as a function of the input phase coefficient for difference phase shifts and combinations of images indicated in the legend. 35

Figure 2.6 RMS error between the set and calculated phase as a function of the phase shift, γ. The phase is calculated from one, two and three images. The phase coefficient is set at 0.1 radians. 36

Figure 2.7 Contrast of the difference data plotted as a function of phase shift, γ 38

Figure 2.8 RMS Error plotted as a function of the number of iterations. The error is between the measured and the recalculated positive phase shifted

intensity for various input phase coefficients (radians) defined in the legend. 40
Figure 2.9 Results of the iterative scheme applied to a system where there is a discontinuity in the phase and the amplitude. a) shows the RMS error plotted as a function of the number of iterations for different input coefficients. b) shows cross sections of the phase during the iterative process. Purple represents the set phase, green represents the phase after 1 iteration, blue represents the phase after 3 iterations and red represents the phase after 5 iterations. 41
Figure 2.10 Results of the iterative scheme applied to a system where there is a discontinuity in the phase and the amplitude. The third image is replaced with the average intensity calculated from this image. a) shows the RMS error plotted as a function of the number of iterations for different input coefficients. b) shows cross sections of the phase during the iterative process. Purple represents the set phase, green represents the phase after 1 iteration, blue represents the phase after 3 iterations and red represents the phase after 5 iterations. 42
Figure 3.1 Schematic of the optical system for recording defocused images 52
Figure 3.2 Schematic of the division of the data over the aperture according to the ITE. 56
Figure 3.3 Examples of the way the phase is encoded in to the intensity derivative. The top line shows the phase a) shows defocus Z_2^0 b) shows Z_5^3 and c) shows Z_3^3. The bottom line shows the corresponding derivative calculated using the ITE. 56
Figure 3.4 Flow diagram of the procedure for inverting the data using the Fourier method. 76
Figure 3.5 Example of the way the coefficients of the harmonic modes are encoded into the boundary data. The left figures represent the data along the circular boundary so the data is by definition 2π periodic. 77
Figure 3.6 Schematic of the data points used in the bicubic interpolation 79
Figure 3.7 Boundary data used to calculate the coefficients of the harmonic functions. Red represents the data calculated using the nearest neighbour method, blue represents the bicubic interpolated data, green represents the sinc interpolated data and purple represents the boundary data calculated using the ITE.. 81

Figure 3.8 Calculated spectrum of coefficients. a) shows the real part of the spectrum. b) shows the imaginary part of the spectrum. In both graphs, the coefficients calculated from the nearest neighbour data are shown in red, blue represents the coefficients calculated from the bicubic interpolated data and green represents the coefficients from the sinc interpolated data 81
Figure 4.1 Schematic of an imaging system using a) a standard straight line grating and b) a distorted diffraction grating. The thicker black line represents a stop between which the phase is to be determined. 91
Figure 4.2 Schematic of the experimental setup. The image of plane B is inverted. 93
Figure 4.3 Schematic of the model to derive the defocus coefficients 94
Figure 4.4 Defocus coefficients plotted as a function of the position of the light source. The coefficients calculated using the ITE are marked in blue and the coefficients calculated using the dimensions of the system are marked in red 95
Figure 4.5 Experimental setup (Figure supplied by AMO / Wavefront Sciences Inc) 97
Figure 4.6 An example of the experimental data provided by AMO / Wavefront Sciences Inc 97
Figure 4.7 Experimental results from a set of data provided by AMO / Wavefront Sciences Inc. The calculated coefficients are plotted as a function of the light source position. 98
Figure 5.1 Output coefficient plotted as a function of the number of sample in the radius of the circle. a) shows the coefficients calculated using the Green's function and b) shows the coefficients calculated using the Fourier transform method. 101
Figure 5.2 Output coefficient plotted as a function of the number of sample in the radius of the circle. The coefficients are calculated using the Green's function and the set phase is defined by Zernike polynomials with increasing order. 103
Figure 5.3 RMS error between the set and calculated phase plotted as a function of the input phase coefficients. Three different normalisation methods are used in the process of calculating the phase; the average intensity, I_A, the point by point average, $I_P(\mathbf{x})$ and the third image, $I(\mathbf{x})$. 105
Figure 5.4 Schematic of the simulation setup showing the relative position of the measurement planes to the plane where the phase is estimated 106

Figure 5.5 Output phase coefficient as a function of the plane spacing Δz. a) shows the phase calculated using the Green's function and b) shows the phase calculated using the Fourier transform method. The exact axial derivative and the derivative approximated using different numbers of measurements are used as input data and the set phases is Z_5^3 108
Figure 5.6 Schematic showing the relative position of the reconstruction plane to the assumed measurement plane marked by the blue dashed line and actual measurement planes in green. 110
Figure 5.7 Input and output phase coefficients plotted as a function of the measurement plane translation. 110
Figure 5.8 An example of intensity data, with an estimation of the boundary marked by the purple circle 113
Figure 5.9 a) shows the input and output phase coefficient plotted as a function of the evaluation radius when the set phase is Z_5^3. b) shows the input and output phase coefficient plotted as a function of the evaluation radius when the set phase is Z_2^{-2}. The correct radius is 85 pixels. 113
Figure 5.10 Illustration of misalignment of the intensity data 115
Figure 5.11 Output phase as a function of the displacement of one of the images. For different input phase distributions, Z_2^0, Z_5^3 and Z_2^{-2}. 116
Figure 5.12 Example of the phase calculated from misaligned data. The example corresponds to the defocus case in Figure 5.11. Purple represent the phase calculated when the misalignment is zero pixels, green represent a misalignment of 2 pixels, blue represent a misalignment of 4 pixels and red represent a misalignment of 6 pixels 116
Figure 5.13 Output coefficient plotted as a function of the grid size used when calculating the phase using the Fourier transform method 117
Figure 5.14 Time to complete the calculation of the phase from data generated at different resolutions. The time from the calculation (seconds) is plotted as a function of the grid size. The data is defined over a circle a quarter of the grid size. 118
Figure 6.1 Schematic of a confocal microscope 122
Figure 6.2 Schematic of a wide field epi-fluorescence microscope 122
Figure 6.3 Schematic of a microscope, which is defocused to adjust the imaged plane 124

Figure 6.4 Schematic of the imaging system 126
Figure 6.5 a) 3D view of the phase object b) Images showing the phase of the back propagated complex amplitude in different planes. c) Corresponding slices of the phase object. The depth of the volume is 30 μm and the slices in b) and c) are, from left to right, 9.6 μm, 11.25 μm, 17.34 μm and 19.68 μm from the entrance plane. 128
Figure 6.6 a) Schematic of the imaging system showing a phase object one volume element in size is placed 2.5 microns into the volume from the entrance plane. b) Across section of the back propagated phase (radians). c) a cross section of the back propagated intensity. The focal points are marked in yellow, 2.5 microns from the entrance plane. The entrance plane is at the left and the exit plane at the right. 129
Figure 6.7. a) shows a schematic of the imaging system where two objects are placed in the volume directly behind each other. Each phase object is one volume element and both apply the same phase change of 0.1 radians. b) shows a plot of the phase through the volume along the dashed line shown in a)..... 131
Figure 6.8 a) Schematic of the imaging system where two objects placed in the volume and separated laterally. Each object is one volume element in one size and causes a phase change of 0.1 radians the illuminating wave is propagating out of the page. b) shows a plot of the phase through the volume along the dashed line in a)..... 132
Figure 6.9 a) Schematic of a lens placed in a volume at 2.5 μm from the entrance plane and illuminated by a plane wave. b) shows a cross sections of the intensity, on a log scale, of the plane marked in red in a) 133
Figure 6.10 Parameter used in the simulations 137
Figure 6.11 Additional quantities describing the simulated systems 137
Figure 6.12 Phase objects: a) a helix and b) a hollow sphere. The colour scale indicates the position of the object in depth, along the z axis. 138
Figure 6.13 Results of the iterative process to estimate the object shown in Figure 6.12 a). a) shows the estimated object. b) shows the phase in the exit plane after illuminating the estimated object. c) shows the residual error between the phases in the exit plane corresponding to the original and estimated objects. NA = 1..... 139
Figure 6.14 Convergence of the iterative scheme for example in Figure 6.13 140

Figure 6.15 Results of the iterative process to estimate the object shown in Figure 6.12 a). a) shows the estimated object. b) shows the phase in the exit plane after illuminating the estimated object. c) shows the residual error between the phases in the exit plane corresponding to the original and estimated objects. NA = $1/8$ 141

Figure 6.16 a) shows a cross section of the set phase object for the example in Figure 6.15. b) shows the same cross section through the estimated phase object for the example in Figure 6.15. The cross sections are through the plane $x = -32.9\mu m$ 142

Figure 6.17 Results of the iterative process to estimate the object shown in Figure 6.12 b). a) shows the estimated object. b) shows the phase in the exit plane after illuminating the estimated object. c) shows the residual error between the phases in the exit plane corresponding to the original and estimated objects. NA = 1 143

Figure 6.18 Results of the iterative process to estimate the object shown in Figure 6.12b). a) shows the estimated object. b) shows the phase in the exit plane after illuminating the estimated object. c) shows the residual error between the phases in the exit plane corresponding to the original and estimated objects. NA = $1/8$. 144

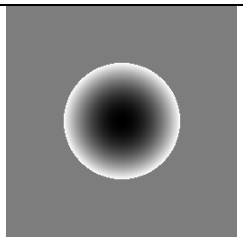
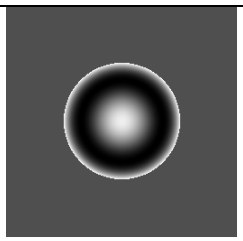
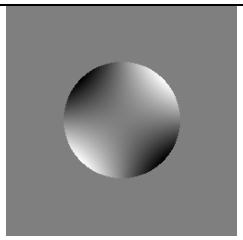
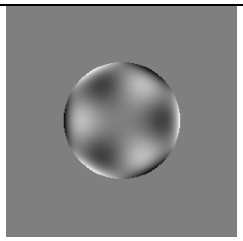
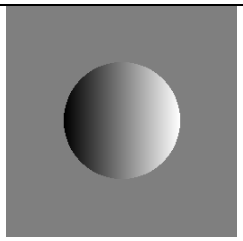
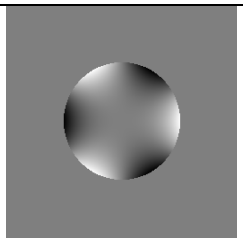
Figure 6.22 a) shows the projection of the set object along the x axis. b) shows the projection along the x axis of the object calculated from the known phase in the exit plane. c) shows the projection along the x axis of the object calculated from phase contrast images. d) shows the projection along the x axis of the object calculated from defocused images. The colour scale is arbitrary to allow the fainter parts of the object to be visualised. 153

List of Symbols

$g(\xi,\tau)$	Function to apply an arbitrary phase shift
k	Wavenumber
$\hat{\mathbf{n}}$	Outward pointing unit vector normal to edge of the aperture
$\mathbf{x}$	Position vector in real space
$\mathbf{x}'$	Position vector in real space
A	Transparency of the area of the scattered wave
B	Transparency of the area of the reference wave
$Data(\mathbf{x})$	Input data from the defocused images
$G(\mathbf{x},\mathbf{x}')$	Green's function
I_0	Constant intensity
$I(\mathbf{x},\tau)$	Intensity
$P(\mathbf{x})$	Aperture
$R(\mathbf{x})$	Reference Wave
Z_m^n	Zernike polynomial of order m and angular frequency n
$\alpha(\mathbf{x})$	Known phase
δ	Dirac delta
γ	Constant phase shift
$\gamma(\xi)$	Arbitrary phase shift
$\phi(\mathbf{x})$	Unknown phase
τ	Coefficient of the arbitrary phase shift
λ	Wavelength
ξ	Position vector in Fourier space
ξ'	Position vector in Fourier space
$\psi(\mathbf{x},\tau)$	Complex amplitude of a wave in real space
∇^2	The Laplacian which represents $\partial^2/\partial x^2+\partial^2/\partial y^2$ in Cartesian coordinates
Δz	Plane spacing
$\Delta I(\mathbf{x})$	Difference between two phase contrast images
$\Sigma I(\mathbf{x})$	Sum of two phase contrast images

$\Psi(\xi\tau)$	Fourier transform of $\psi(\xi)$
$\mathfrak{F}$	Fourier transform
$\mathfrak{F}^{-1}$	Inverse Fourier transform
$\Re\{\ \}$	Real part
$\Im\{\ \}$	Imaginary part
$\otimes$	Convolution
$\forall$	For all
$\in$	Is an element of
$[\ ]$	Closed interval including the end points

List of Phase Functions

Name	Symbol	Equation, $\phi(\mathbf{x})$	Picture	Harmonic[1]/Non Harmonic
Defocus	Z_2^0	$2r^2 - 1$		Non harmonic
Spherical Aberration	Z_4^0	$6r^4 - 6r^2 + 1$		Non harmonic
Astigmatism	Z_2^{-2}	$r^2 \sin(2\theta)$		Harmonic
	Z_5^3	$(5r^5 - 4r^3)\cos(3\theta)$		Non harmonic
Tilt	Z_1^1	$r\cos\theta$		Harmonic
	Z_3^3	$r^3\cos(3\theta)$		Harmonic

1. A function is harmonic if $\nabla^2\phi(\mathbf{x}) = 0$ inside the area of the disk, section 3.2.1.

Chapter 1 – Introduction

1.1. Introduction

Light can be described as a wave, which can be characterised by amplitude and phase. When the light undergoes transmission, reflection or diffraction there is a change in both the amplitude and phase of the wave.

In general, when light is detected by, for example, the eye or a camera, the intensity is measured. For most practical optical measurements the intensity is proportional to energy and can be calculated from the square of the amplitude but does not contain information about the phase.

There are situations where light interacts with an object and there is little or no change to the amplitude of the wave. In these cases, the objects are invisible to conventional instruments but the objects have changed the phase of the wave. Therefore, properties of some objects that can only be known from the phase change.

For example, the difference in density of a transparent cell, in comparison to its surroundings, causes a change in phase of the illuminating light. The change in the phase relates to the shape, structure and refractive index of the cell, Popescu, *et al.* [1].

There are other areas where phase information is required. For example in ophthalmology phase information is required to calculate the aberration in a patient's eye, Campbell [2], or in astronomy, when using adaptive optics to compensate for viewing through the atmosphere, Roddier [3].

Since phase cannot be measured directly, an indirect method is required. Using an indirect method either exploits some known property of the wave, for example interference, or modifies the wave in a known way, for example measuring a defocused image. From the measured intensity data, the phase can then be calculated. This requires knowledge of the specific imaging process.

This thesis examines methods for determining phase from intensity measurements through detailed theoretical analysis. It examines the limitations of standard methods and identifies opportunities for improving the data analysis and the experiment process.

1.2. General Wavefront Sensing Methods

There are three general methods for determining phase information, interferometry gradient measurements and phase diversity.

1.2.1. Interferometry

The first method is interferometry; there are two types of interferometers, common path interferometers and non-common path interferometers. This section describes each of these types of interferometer.

Interferometry is an old and well-understood technique. The principle of an interferometer is to exploit a property of the waves to produce a pattern in the measured intensity that can be used to calculate changes in the phase of the illuminating wave. Interferometers use at least two coherent beams of light, which are added together causing an interference pattern. Although more than two beams of light can be used, the discussion below is restricted to the special case of using two beams of light, as this method is not considered further in this thesis.

In general, interferometry combines the beams at an angle to produce fringes of equal inclination. A common example of this type of interferometer is a Mach-Zender interferometer, Hecht [4], as shown in Figure 1.1. This instrument uses a laser to produce a coherent source that passes through a beam splitter to give to two waves each with lower amplitudes.

Figure 1.1 shows the path of the beam of light, in red, emanating from the source and being split by the first beam splitter. These two beams then take different paths before being recombined. Differences in the waves caused by propagating along different paths cause an interference pattern, which is measured on the detector. Placing an object in the path of one of the beams causes a phase change. The phase change is caused by changing the path length of the beam. This will modify the interference pattern from which the phase can be calculated.

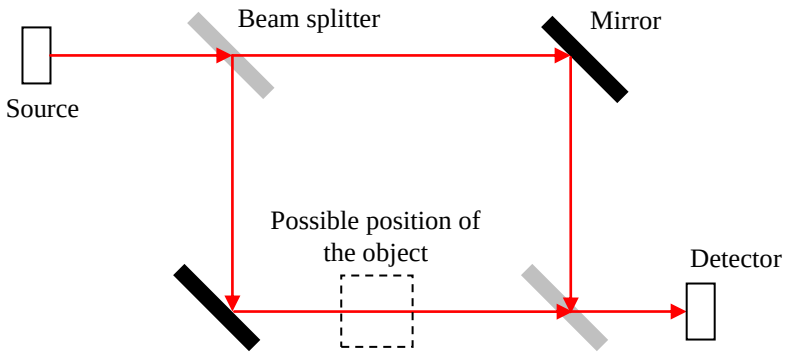


Figure 1.1 Schematic of a Mach-Zender interferometer

In general interferometry has the advantage of high resolution but is experimentally difficult as it requires a well defined reference beam against which the phase can be calculated, Lago, *et al.* [5]. Interferometry is also sensitive to vibration induced errors between the two optical paths. This increases the need for environmental control and thus reduced the applicability of this method in, for example, microscopy.

1.2.2. Common Path Interferometry

A common path interferometer is a variation on the previously described interferometer, where the beam of light is separated. In this imaging system shown in Figure 1.2, a single beam of light passes through the imaging system.

If there is no object then the light is focused in a small area of plane C. This area is defined as the reference wave. This light is passing directly through the instrument. When an object is introduced, the wavefront is modified so the light no longer passes directly through the instrument. Therefore, some of the light passes outside the area of the reference wave. The light that passes outside the reference area is called the scattered wave. This wave relates to the wavefront that is to be determined.

The modified wavefront can be due to the interaction of the light with an object, for example a cell, which modifies the wavefront as shown in plane A. The complex amplitude propagates from plane A, through the lens in plane B and is focused onto plane C.

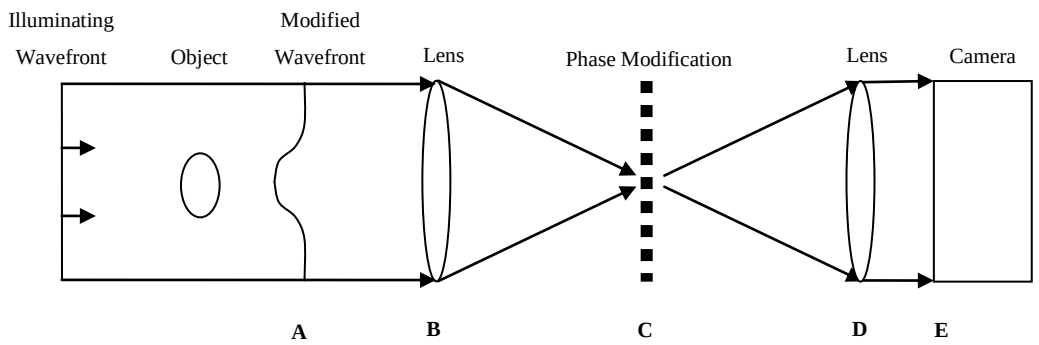


Figure 1.2 Schematic of a general common path interferometer. The Black lines represent the path of the light when there is no object.

To make the phase visible a phase change is applied to the reference wave in plane C represented by the dashed line. The modified wave then passes through a second lens, plane D, to the camera, plane E, the two parts of the wave interfere, producing an interference pattern. The modification of the reference wave relative to the scattered wave causes modulations in the intensity that relate to the phase difference between the reference wave and the wave modified by the object.

A common example, of this type of instrument, is phase contrast imaging, Zernike [6], (described in full in Chapter 2) where a constant phase change of $\pm\pi/2$ radians is applied to the reference wave. The phase shift can be applied, for example, using a uniformly thick piece of glass except where the reference wave passes through. In this area, the glass is a different thickness causing a relative phase change, Born, *et al.* [7]. The application of the phase change causes modulation in the measured intensity that relate directly to the phase object.

Figure 1.3 a) shows the set weak phase. Figure 1.3 b) shows the images with no phase shift in this image there are no modulations relating to the object. Figure 1.3 c) shows the phase contrast image with a phase shift of $\pi/2$ radians. In this image, there is a direct relation between the modulations in the intensity and the object.

Other examples of this general principle include various types of point diffraction interferometry Mercer, *et al.* [8], Neal, *et al.* [9] Fourier phase microscopy, Lue, *et al.* [10] and fast Fourier phase microscopy, Popescu, *et al.* [11]. These methods all convert the phase information into visible intensity modulations.

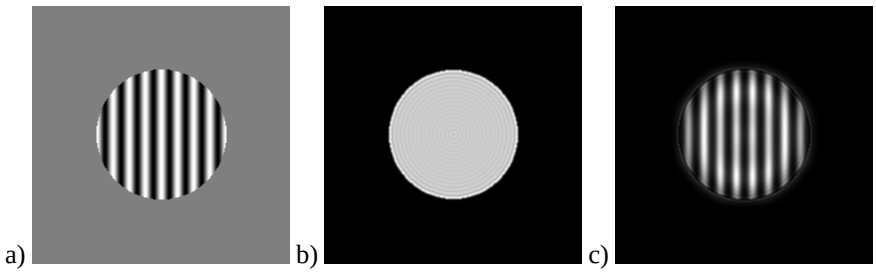


Figure 1.3 Example of phase contrast images. a) show the set phase. b) an image generated with no phase shift and c) an image generated with a phase shift of $\pi/2$ radians

The advantage of a common path interferometer is that both the scattered and the reference waves propagate along the same path, therefore, the instrument is insensitive to aberrations induced by, for example, mechanical vibration or air fluctuations Rodrigo, *et al.* [12]. The disadvantage of this process is that to calculate the phase the reference wave needs to be known. In general, neither the reference nor the scattered wave are known and they are not separable, approximations for the reference wave are introduced to the analysis causing errors in the estimated phase.

1.2.3. Gradient Measurements

The second general method to determine phase is to calculate the local gradient of the wavefront. The derivative of the wavefront is calculated with respect to the x and y axis's from which the phase is calculated. These derivative can be approximated using the measurement from, for example, a Shack Hartmann wavefront sensor, Platt [13], or a lateral shearing interferometer, Hariharan [14].

Shack Hartmann wavefront sensors are widely used to determine phase. The principle of this sensor is illustrated in Figure 1.4. To calibrate the instrument, Figure 1.4 a), the unmodified wavefront illuminates an array of lenses, marked in red, these lenses focus sections of the illuminating beam on to the measurement plane giving a set of spots, marked in red.

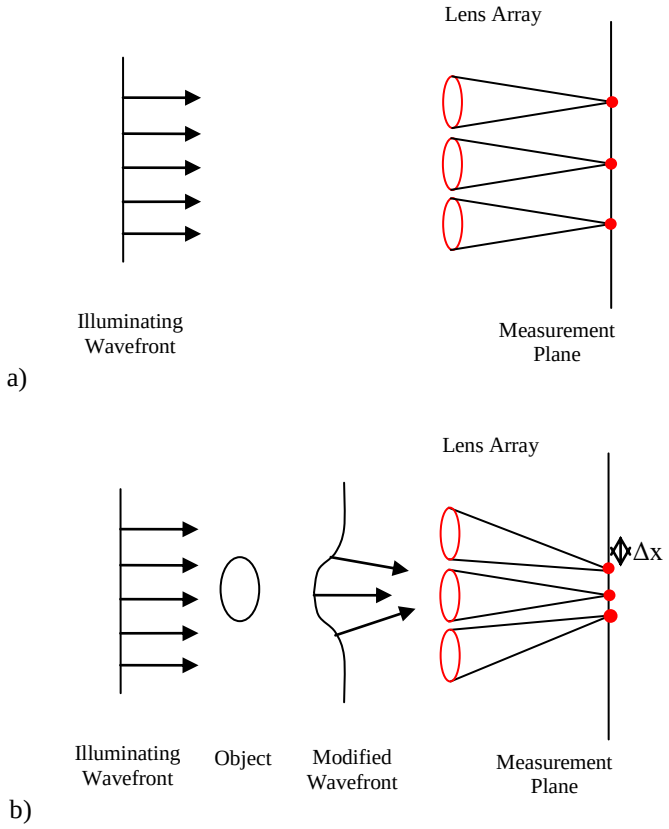


Figure 1.4 Schematic of a Shack Hartmann wavefront sensor. a) shows the instrument illuminated by a plane wave for calibration. b) shows an illuminated object which modifies the phase and the position of focal points.

After calibration, the instrument can be illuminated by a modified wavefront, Figure 1.4 b). Each part of the beam of light that illuminates each individual lens is propagating in a different direction due to the phase change caused by the object. This causes the spots to move from the initial calibrated position. This displacement combined with the distance between the measurement plane and the array of lenses can be used to give an estimation of the first derivative of the phase with respect to the x and y direction. This information can then be used to determine the phase, Zou, *et al.* [15].

Calculating phase from Shack Hartmann measurements has the advantage that the phase can have a wide dynamic range and the calculated phase is not wrapped, Tyson [16]. A disadvantage of this method is the complexity in aligning and calibrating the instrument, Rousset [17].

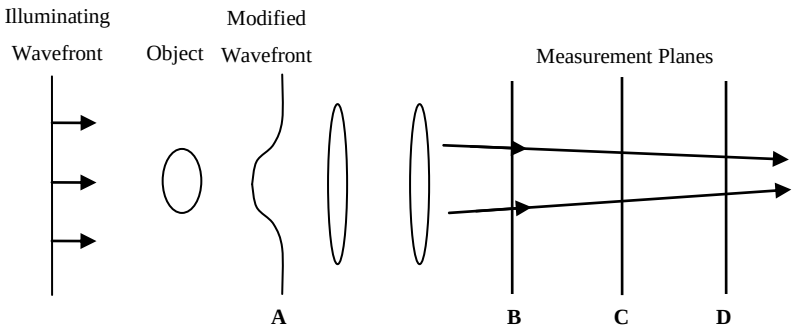


Figure 1.5 Schematic of the principle of phase diversity using defocused images.

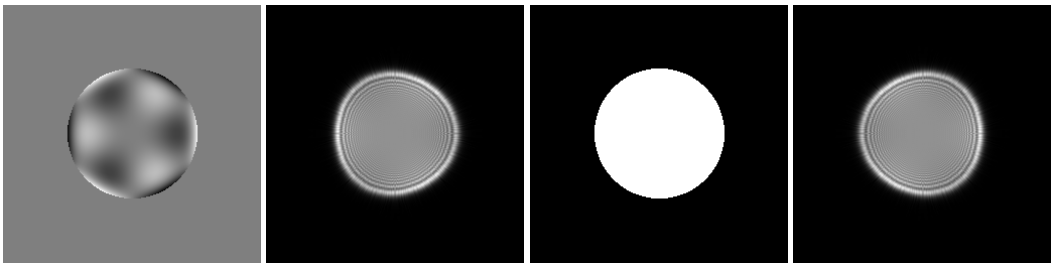
1.2.4. Phase Diversity

The final general method to determine the phase is known as phase diversity, Gonsalves [18] and Jefferies, *et al.* [19]. Originally used for application in astronomy this method uses multiple images each of which has a different, but known, phase change applied to the wavefront during the imaging process. The phase modification is applied in conjunction with reimaging through additional optics or propagation. The images are therefore linked by the application of the phase change, so from this data using a suitable inversion process, the phase can be determined.

A common phase modification applied to the wavefront is a defocus phase change. This is illustrated in Figure 1.5 for a wavefront in plane A, propagating to the right. The defocused measurements can be made, for example, by physically moving the camera to a different position, as shown in the figure, or using a beam splitter and two cameras to capture the images simultaneously. When using two images one or both of them will be out-of-focus.

The application of a defocus phase change is not a requirement as equations have also been derived using spherical aberration, Allen, *et al.* [20], and astigmatism, Petersen, *et al.* [21], to modify the intensity from which the phase can be calculated.

From the phase diverse data, the phase can be calculated using iterative processes, Misell [22]. Equations can be derived that use the difference between the two images as the input data, Campbell, *et al.* [23], from which the phase can be calculated. To link the images an equation can also be derived that relates the derivative of the intensity, with respect to the propagation direction, to the phase, Allen, *et al.* [20]. This derivative can later be estimated using a finite difference approximation.



a) b) c) d)

Figure 1.6 An example of defocused measurements. a) shows the set phase. b) shows the measurement from plane B of Figure 1.5. c) shows the measurement from plane C. d) shown the measurement in plane D.

A special case of phase diversity is curvature sensing. First introduce by Roddier [3] in 1987 for applications in astronomy, Curvature sensing has since been using in other areas, for example in ophthalmology for determining aberrations in patients eyes, Diaz-Douton, *et al.* [24].

When the intensity is constant in the plane where the phase is to be calculated, illustrated as plane C in Figure 1.5, and two defocused images are recorded, in planes B and D, equal distances from plane C, the measured data relates to the curvature of the phase, Roddier [3].

The analysis in the following chapters assumes that the intensity in plane C is constant and is defined over a sharp edged disk, this is necessary to simplify the analysis. For example, if the phase is defined as shown in Figure 1.6 a) then the measured intensity in plane C of Figure 1.5, and shown in Figure 1.6 c), does not show modulations in the intensity relating to the phase. This is similar to the phase contrast example, Figure 1.3 b), when there is no phase shift to the reference wave.

If the camera is moved to planes B and D in Figure 1.5 then images like the ones shown in Figure 1.6 b) and d) can be measured. When using phase contrast images, there was a direct relation between the measured intensity and the phase. Using defocused images this benefit is lost and post processing of the data is required to obtain a view of the phase.

When using a phase change that is equivalent to propagating a wavefront from one reference plane to another, the phase can be related to the intensity, Roddier [3], by a differential equation,

$$-\frac{k}{I_0}\frac{\partial I(\mathbf{x},z)}{\partial z} \cong P(\mathbf{x})\nabla^2\phi(\mathbf{x},z) - \delta(\mathbf{x}-\mathbf{r}_0)\hat{\mathbf{n}}\cdot\nabla\phi(\mathbf{x},z). \tag{1.1}$$

Where the intensity I_0 is a constant over the aperture with unspecified shape, $P(\mathbf{x})$. An outward pointing unit vector is denoted by $\hat{\mathbf{n}}$ which is normal to the perimeter, C of the aperture, δ denotes a Dirac delta function at position $\mathbf{r}_0$ on the perimeter of the aperture which has an area A.

This differential equation, the intensity transport equation (ITE), describes the curvature sensing setup, Roddier [3]. By solving this equation, the phase can be calculated from the intensity and known parameters of the imaging system.

There are numerous proposed solutions in the literature to solve this equation. These solutions often involve approximations, for example expressing the phase using a finite number of base functions Gureyev, *et al.* [25], Gureyev, *et al.* [26], ignoring the boundary conditions Dorrer, *et al.* [27] or numerically solving the equation Allen, *et al.* [28]. All of which introduce errors into the solution. There are two method described in the literature that can solve this equation exactly. This is discussed in full in Chapter 3.

Phase diversity is generally not used as a common method for determining phase due to the extensive post processing of the data that is require to obtain useful information Lee, *et al.* [29]. However phase diversity has the advantage of being experimentally simple Jefferies, *et al.* [19], using fewer optical components than interferometry or a Shack Hartmann wavefront sensor, this reduces potential experimental errors, for example, due to misalignment to components.

1.3. Aims and Objective of this Thesis

There are applications where knowledge of the phase can be used to determine properties of transparent or semi-transparent objects. So far, this chapter has described general methods for calculating the phase from the intensity measurements. Each of these methods has fundamental advantages and disadvantages and the choice of the individual method will depend on the specific application.

1.3.1. Comparison of the General Wavefront Sensing Methods

This thesis will consider the principle of phase diversity to calculate phase from intensity measurements. In general, phase diversity is not commonly used due to the extensive numerical post processing of the data that is required, Lee, *et al.* [29]. However, phase diversity has the advantage over other types of wavefront sensor that the measurements are experimentally simpler to make, this reduces the potential for experimental error, Jefferies, *et al.* [19].

In comparison, an interferometer has a very high resolution and the analysis of the data is simple but it is experimentally complicated as it requires a well defined reference beam, Lago, *et al.* [5].

Similarly, the commonly used Shack Hartmann wavefront sensor requires well aligned components. The resolution of the phase obtained using a Shack Hartmann wavefront sensor is lower than can be achieved using phase diversity, Diaz-Douton, *et al.* [24]**,** and as such finer detail can be determined using phase diversity. This is particularly important for biological application where high-resolution images are beneficial. For example, this allows different parts of a cell to be distinguished.

Experimentally the phase diversity method makes use of the common path design making the measurement insensitive to fluctuations in the air and mechanical vibration, Rodrigo, *et al.* [12]. Therefore, only phase changes due to the object are measured.

Therefore, this thesis looks at calculating phase using the phase diversity method because the measurement are experimentally relatively simple to make and there is potential to improve the data analysis, which is the key area restricting the use of this method. In particular, this thesis looks at calculating phase from sets of defocused images and phase contrast images. These types of images are considered as phase contrast images in a simple process to understand and there is a direct mapping between the measured intensity and the phase. The defocused images are considered as this is a commonly studied example of phase diversity and as such is well understood, but there are still potential for improvement in the data analysis.

The main aim of this thesis is to develop the mathematical analysis and data processing methods, which can be applied to data measured using standard experimental techniques

and procedures. As such, the thesis describes the theory and numerical processing of experimental data without explicit consideration of the methods for obtaining the data.

The analysis in the following chapters is not dependent on a particular method of obtaining the measurements and throughout this thesis and it is assumed that all the necessary measurements can be made. Additionally, noise is not considered in detail although general comments are made throughout this thesis. For an accurate estimation of the phase the signal to noise ratio will have to be greater than 1 to avoid amplifying the noise. It will be assumed that this can also be achieved.

1.4. Outline of the Thesis

This thesis is split into three parts. The first part, Chapter 2, looks at calculating the phase from phase contrast images, the second part, Chapters 3, 4 and 5, uses defocused images to calculate the phase. The final chapter considers the possibility of determining the structure of a 3D object from the calculated phase calculated in a single plane using the methods described in the previous chapters.

Chapter 2 follows on from the work of Rodrigo, *et al.* [12] where a framework was developed for a general common path interferometry and the phase was calculated from three phase contrast images with phase shifts, applied to the reference wave, of 0, $\pi/2$ and π radians. Chapter 2 considers similar analysis and develops a modified method for calculating the phase using up to three measurements with phase shifts of 0 and $\pm\theta$ radians. The new method developed in this chapter is more general, allowing greater flexibility in the modelled system and the potential for further improvement by including additional measurement and the possibility for further optimisation of the process.

It will be shown that there is a significant reduction in the error by increasing the number of images used to calculate the phase from one to three. The analysis is developed independent of a specific imaging technique but is considered in the context of phase contrast imaging.

The model of a general common path interferometer is developed in full to understand the benefits and the approximations that are introduced in the data inversion. It also

allows the theory used in Chapter 3 to be introduced for the specific simple case of phase contrast before the more complex process using defocused images.

Chapter 3 develops the theory of calculating the phase from defocused images. The intensity and the phase are related by the intensity transport equation (1.1), Roddier [3]. This chapter derives two analytic solutions of the ITE the first is a Green's function solution and the second uses Fourier transforms. It has been shown, Woods, *et al.* [30], that the phase can, in general, be determined using a Green's function solution to the ITE. The calculation process is implemented to return the coefficients of single modes.

In the literature, there are several different and inconsistent Green's functions stated that claim to solve this problem for a specific case of a uniformly illuminated circular aperture. Chapter 3 examines this issue, where two Green's functions are derived and shown to be consistent with two of the Green's functions stated in the literature. The chapter shows that these Green's functions satisfy all the necessary mathematical conditions and that the remaining Green's function stated in the literature is incorrect as it fails to satisfy the necessary conditions to be a valid solution.

The two correct Green's functions are found in mathematical books and are not stated in a way that can be easily translated for use with experimental data. Therefore, the numerical analysis of experimental data using these Green's functions is described in full. The Green's functions contain singularities. Therefore, a process will be described that removes the singularities.

The second method considered for solving the ITE uses Fourier analysis. Previously, this method has been partially implemented but the boundary conditions were neglected, Dorrer, *et al.* [27], or included iteratively, Diaz-Douton, *et al.* [24]. In this chapter, a full analytic solution is derived with the inclusion of the boundary conditions.

After developing the theory, it is tested with experimental data in Chapter 4. Experimentally, when measuring the data there are a number of parameters that can be optimised, for further improvement of the estimated phase. Similarly during the numerical analysis of the data there a factors that can influence the phase estimate. Chapter 5 shows the effect on the calculated phase of factors like the resolution of the data and the number of images. The effect of the number of images and the plane

spacing on the phase has been independently considered. This formed part of the work by Waller, *et al.* [31] which was published in May 2010.

Chapter 5 also shows the sensitivity to misaligning the data and incorrectly defining the position of the boundary. It will be shown which of these parameters can be used to improve the accuracy of the phase estimate and which have minimal effect.

Chapter 6 expands on these results and considers the possibility of determining the three dimensional structure of an object knowing the complex amplitude of a wave in a single plane. By focusing a microscope through an object, in some cases, it is possible to determine the structure of the object by eye, Streibl [32]. This chapter considers the same process, achieving the refocusing through numerical back propagation of the complex amplitude. This process gives access to both the intensity and the phase within the volume of back propagated data. An iterative process is then developed to determine the object from this data. It will be shown that the 3D object can be determined, in some cases, from phase contrast images and from defocused images. Areas of continuing difficulty will be highlighted.

Chapter 7 concludes the thesis, summarising the work in the previous chapters and areas for further work.

Chapter 2 – Phase Contrast Imaging

2.1. Introduction

Phase contrast imaging has been used for many years in optical microscopy e.g. Zernike [6], electron microscopy e.g. Danev, *et al.* [33] and elsewhere to visualize objects that produce little or no amplitude modulation. An example would be light passing through an almost transparent biological cell, where there is almost no measurable change in intensity but there is a phase change, as the cell is denser than its surroundings. This phase change can be exploited to provide an improved qualitative view of the object. As was described in Chapter 1, phase contrast imaging is one such method for visualising objects.

This phase information allows some of the properties of the object to be determined, such as the structure of the cell which cannot be determined from the intensity, Popescu, *et al.* [34].

Recently there have been a number of variations to the basic method of phase contrast imaging to enable phase information to be calculated from intensity measurements. It has been shown, for example, that by combining phase contrast imaging with phase shifting interferometry and using four measurements with phase shifts of 0, $\pi/2$, π and $3\pi/2$ the phase can be calculated, Popescu, *et al.* [11] and Lue, *et al.* [10].

A second variation, Popescu, *et al.* [34], splits the incoming wave into two waves each with half the amplitude. One of the waves is used to generate the reference wave, in a similar way as described in Chapter 1, by defining the area of the reference wave and blocking the remainder of the wave i.e. low pass filtering. An interference pattern is then created by interfering the two waves, this can be analysed to calculate the phase.

Both these methods are improvements to the experimental process based on a general common path interferometer. They provide an accurate method for determining the phase with a quick image acquisition time for the measurements but there is no improvement to the data analysis

The variations in the process to calculate the phase that are discussed in the literature are all changes to the experimental system without properly considering the analysis of the data. Therefore, this chapter looks at phase contrast imaging as a simple method to access the benefits that can then be achieved by improving the analysis of the data.

This chapter aims to develop a method for calculating the phase from measurements recorded using a general common path interferometer to understand the imaging system and to improve the estimation of the calculated phase without changing the experiment. The chapter will look at the specific example of phase contrast imaging.

This chapter considers the benefits of combining the data from two phase contrast images, each measured with an equal and opposite phase shift applied to the reference wave (see below), and a third image with no phase shift. The principles of combining multiple images in this way has been taken from the phase diversity method, Gonsalves [18], (described in full in Chapter 3) but is implemented in a way that maintains the direct relationship between the phase change and the spatial modulations of the measured intensity, the principle of phase contrast imaging.

Rodrigo, *et al.* [12] described a method for calculating the phase from three phase contrast images with phase shifts of 0, $\pi/2$ and π radians. The method developed in this chapter extends from this work allowing for greater flexibility in the modelled system, the potential to include more measurements and optimise the process to improve the accuracy of the calculated phase

The main benefit from this development is an improvement in the accuracy of the calculated phase, though fewer mathematical approximations. It will be shown that there is a significant improvement from this method in comparison to using a single image.

The theory is presented in a manner that is complete and self-contained. This will allow the approximations to be clearly visible and the consequences understood as well as introducing the theoretical method used in chapter 3.

This method will be used in section 6.4 in a process to determine the three dimensional structure of an object.

As was discussed in Chapter 1, there is less emphasis on the methodology of measurement as this is likely to be application specific. However, all the measurement techniques can be undertaken using existing procedures and standard commercial microscopes.

2.2. Background – Phase Contrast Imaging

This section describes the general theory that describes a common path interferometer with particular reference to phase contrast imaging. The physical imaging process will be described mathematically to produce an equation that relates the measured intensity to parameters of the imaging system and the phase. From this equation, a process will then be described to calculate the phase from a single image or from a combination of images.

2.2.1. Physical Description of General Common Path Interferometer

This section describes the theory to model a common path interferometer, with particular reference to phase contrast imaging. The imaging system in this chapter as shown in Figure 2.1 is similar to that described in Chapter 1.

The source of the phase change from the object, $\phi(\mathbf{x})$, that is to be calculated is unspecified but for example could be the result of the light passing through a cell. The experimental system is shown in Figure 2.1 a), where a plane wave illuminates a volume containing a cell that modifies the phase the light then propagates through an aperture as shown in plane A.

The complex amplitude of the wave in plane A passes through the lens to plane C. In this plane, part of the wavefront is considered as the reference wave and the remainder is considered as the scattered wave.

If there is no object then the light focused in place C is largely confined to a small area. The light that passes through this area is called the reference wave. When an object is placed in the path of the beam of light some of the light goes directly through the object, in the same way as if there was no object and some of the light passes outside the area of the reference wave. The wave that passes outside the reference area is called the scattered wave.

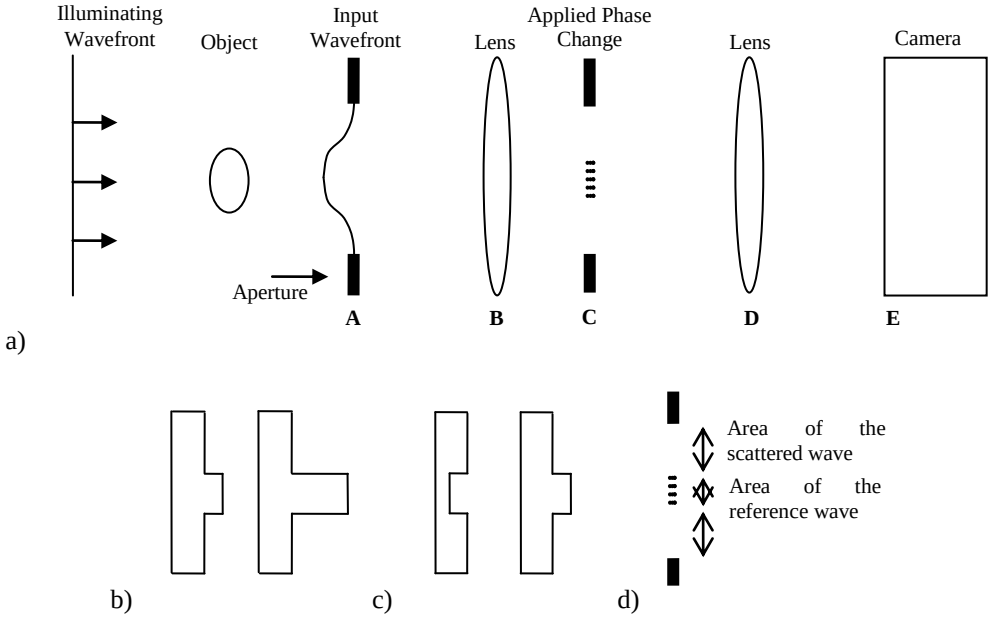


Figure 2.1 a) A schematic of a common path interferometer. b) shows the physical shape of pieces of glass that can be used to apply a phase shifts to the reference beam, in plane C. The phase shifts are in the same direction. c) shown the shape of two pieces of glass that can be used to give equal and opposite phase shifts. d) shows the division of the area of the reference wave and the scattered wave

The area and position where the reference wave passes is not specified, although drawn in the centre of plane C. The unmodified wavefront approximately passes entirely through the area defined as the reference wave. The phase change, $\phi(\mathbf{x})$, causes light to pass outside the reference area in the area considered to contain the scattered wave.

The phase shifted reference wave interferes with the scattered wave to produce an interference pattern detected by the camera in plane E after passing through the second lens in plane D.

The phase shift, γ, can be applied, using a piece of uniformly thick glass except where the reference wave is defined. In this area the glass is a different thickness causing a relative phase change Born, *et al.* [7]. This is shown in Figure 2.1 b) where the differing thickness of glass causes a different phase change.

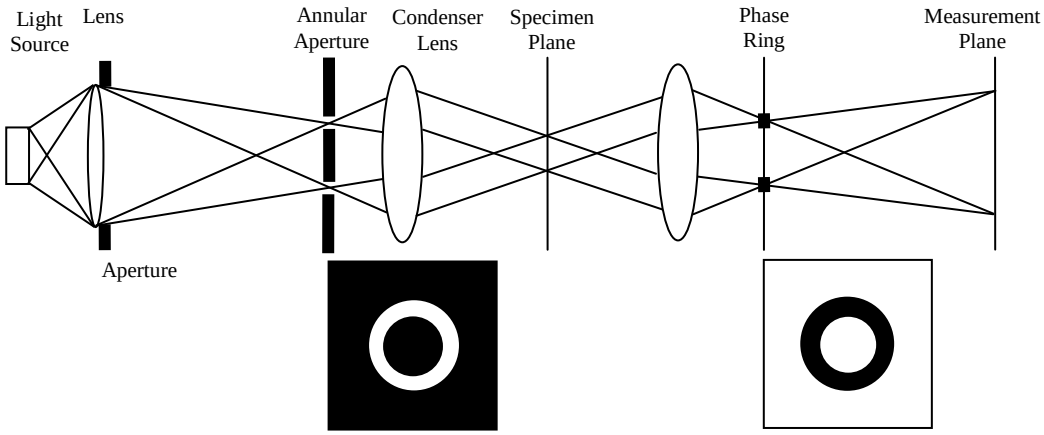


Figure 2.2 Schematic of phase contrast imaging using conical illumination. The black lines show the path of the light

In the following sections two phase contrast images with equal and opposite phase changes, are used to calculate the phase, $\phi(\mathbf{x})$. These phase changes can be applied using glass with the shape shown in Figure 2.1 c). This is not the only way to apply the phase shift as, for example, a spatial light modulator can be used, Ng, *et al.* [35].

In general, phase contrast imaging is implemented using conical illumination, Maurer, *et al.* [36]. This is shown in Figure 2.2 where the light propagates from the left to the right. The partially coherent light source illuminates a lens and the light then passes through an aperture. The light then passes through an annular aperture and a second lens and then illuminates the specimen. This process of illuminating the specimen is known as conical illumination because the light is directed along the surface of a cone.

To apply the phase shift the light passes through another lens where the phase shift is applied over an annular area. After being phase shifted, the light propagates to the camera where the measurement is made.

The benefit of conical illumination is improved spatial resolution. The disadvantage of conical illumination is that artefacts, caused by the illumination process, are introduced into the images, Maurer, *et al.* [36].

The imaging system considered in this chapter contains the essential features of a phase contrast experiment, the application of a phase change to the reference wave but uses a coherent monochromatic illuminating source instead of a partially coherent conical

illumination source. This allows the use of Fourier transforms to represent the lenses, Goodman [37] . This simplifies the analysis but this restricts the conclusions to systems with coherent illuminating sources.

2.2.2. Mathematical Representation of an Optical System

The lenses in the systems considered are represented by the application of the Fourier transform, Hecht [4], denoted by $\mathfrak{F}$. The Fourier transform of a function $\psi(\mathbf{x})$ is defined, Brigham [38], throughout this thesis as

$$\Psi(\xi) = \int \psi(\mathbf{x}) \exp(-2\pi i \mathbf{x} \cdot \xi) d\mathbf{x} \tag{2.1}$$

with the inverse defined as

$$\psi(\mathbf{x}) = \int \Psi(\xi) \exp(2\pi i \mathbf{x} \cdot \xi) d\xi. \tag{2.2}$$

The Fourier transform converts a function that is position dependent into a spatial frequency representation of it. In real space the value of the function depends on the position x, so for example if

$$h(x) = A \exp(i 2\pi f x), \tag{2.3}$$

where A is the amplitude of the wave and f is the spatial frequency of the wave. The Fourier transform of this wave is

$$H(\xi) = A\delta(\xi - f), \tag{2.4}$$

where δ is a Dirac delta function. In the Fourier domain, the position in space defines the frequency of the wave and the value at the particular point defines the amplitude of the wave.

The complex amplitude of the wave in plane C of Figure 2.1 a) is mathematically the Fourier transform of the complex amplitude of the wave in plane A of Figure 2.1 a). In plane A, the complex amplitude varies with position, whereas in plane C the position in space represent the spatial frequency of the wave and the value at a particular position represents the amplitude of wave with the specific frequency.

The frequency can be converted into angle so physically the lens applies a Fourier transform to the incoming wave and decomposes it into a set of plane waves each propagating in a different direction relative to the axis perpendicular to the plane of the lens.

Therefore, if the illuminating plane wave is now modified by an object, decomposition of this wave results in a number of plane waves with different amplitudes propagating at different angles. These can be represented in plane C as information that is away from the central point as the light has been scattered. This information can be described in terms of scattering angle. The scattering angle is the angle a plane wave has deviated from the direct path caused by the object. The concept of the object scattering the light can be used as a physical picture of the system but the analysis will describe the imaging process in terms of spatial frequency.

2.3. Theory

By mathematically describing each part of the imaging process shown in Figure 2.1 a), an equation can be derived that relates the measured intensity, in plane E, to the phase, in plane A, and parameters of the imaging system. This allows the phase to be calculated.

2.3.1. Mathematical Description of a Common Path Interferometer

The complex amplitude of the wave, $\psi(\mathbf{x})$, at position $\mathbf{x}$ in plane A of Figure 2.1 can be defined as

$$\psi(\mathbf{x}) = P(\mathbf{x})\sqrt{I(\mathbf{x})}\exp\left(-i\left(\phi(\mathbf{x})+\alpha(\mathbf{x})\right)\right). \tag{2.5}$$

The phase change caused by the object is denoted by $\phi(\mathbf{x})$ and $\alpha(\mathbf{x})$ represents any initial known phase changes independent of the object and the non-zero intensity is denoted by $I(\mathbf{x})$. The light is restricted by an aperture limiting the area where the complex amplitude is defined; the edge of the aperture is marked in plane A. The aperture is represented by $P(\mathbf{x})$ which equals one inside and zero outside it.

The propagation of the complex amplitude from plane A, through the lens in plane B to the focal plane at C can all be described by a Fourier transform. Therefore, in plane C the complex amplitude describing the wave is

$$\Psi(\boldsymbol{\xi}) = \mathfrak{F}\{\psi(\mathbf{x})\} \tag{2.6}$$

where $\boldsymbol{\xi}$ is the coordinate vector in Fourier space and represents spatial frequency.

After propagating the complex amplitude from plane A through the lens to plane C a phase modification is applied to the area defining the reference wave. The phase change is a constant value applied over a specified area. The phase change, γ, is applied to the reference wave by multiplying the complex amplitude in plane C by

$$g(\xi,\gamma)=T(\xi)\big(A+S(\xi)\big(B\exp(i\gamma)-A\big)\big), \tag{2.7}$$

where $S(\xi)$ is a binary function defining where the phase shift is applied, this specifies the area and the position of the reference wave, Glückstad, *et al.* [39]. $T(\xi)$ is a binary function that defines the band limit of the system. $T(\xi)$, limits the spatial frequencies of the propagating wave so that the measurement can unambiguously be reconstructed from the sampled data. The position of the band limit is a consequence of the spacing of the pixels on the camera. This is described in section 2.4.2.

The part of the propagating wave considered to be the scattered wave passes through the area inside the band limit $T(\xi)$ and outside the area $S(\xi)$. The transparency of the area defining the scattered wave is represented by A where $0<A\leq 1$ and the transparency of the reference wave area is represented by B where $0\leq B\leq 1$. The phase shift is defined as γ where $-\pi<\gamma\leq\pi$. If the phase shift is zero, there are no modulations in the intensity making the object invisible.

By specifying values for A, B and γ commonly used imaging systems can be defined, so for example, if $A=1, B>0$ and $\gamma=\pm\pi/2$ then Zernike phase contrast images are measured and if $A=1/\sqrt{2}$, $B=1$ and $\gamma=\pm\pi/4$ then Henning phase contrast images are measured, Glückstad, *et al.* [39]. When the reference wave is removed, so that $B=0$, then the measurements are dark field images, Hecht [4].

The phase shifted complex amplitude, in plane E, is denoted by $\psi(\mathbf{x},\gamma)$. In general, the complex amplitude with an arbitrary phase change applied to it, can be written, Allen, *et al.* [20], as

$$\psi(\mathbf{x},\gamma)=\int\Psi(\xi)\,g(\xi,\gamma)\exp(-i2\pi\xi\cdot\mathbf{x})\,d\xi. \tag{2.8}$$

Where $\Psi(\xi)$ denotes the Fourier transform of $\psi(\mathbf{x})$ representing the first lens, multiplication by $g(\xi,\gamma)$ applies the phase change and the second Fourier transform represents the second lens giving phase shifted complex amplitude in plane E.

Specifically, when the phase shift is represented by equation (2.7) the phase shifted complex amplitude is

$$\begin{aligned}\psi(\mathbf{x},\gamma) = {} & A\sqrt{I(-\mathbf{x})}P(-\mathbf{x})\exp\left(-i\left(\alpha(-\mathbf{x})+\phi(-\mathbf{x})\right)\right)\otimes\mathfrak{F}^{-1}\{T(\xi)\} \\ & -A\mathfrak{F}^{-1}\left\{S(\xi)T(\xi)\mathfrak{F}\left\{P(\mathbf{x})\sqrt{I(\mathbf{x})}\exp\left(-i\left(\alpha(\mathbf{x})+\phi(\mathbf{x})\right)\right)\right\}\right\} \\ & +B\exp(+i\gamma)\mathfrak{F}^{-1}\left\{S(\xi)T(\xi)\mathfrak{F}\left\{P(\mathbf{x})\sqrt{I(\mathbf{x})}\exp\left(-i\left(\alpha(\mathbf{x})+\phi(\mathbf{x})\right)\right)\right\}\right\},\end{aligned} \tag{2.9}$$

where $\otimes$ represents a convolution. If $S(\xi)=0$, so that there is no phase shift applied to the reference wave, then the reimaged complex amplitude is an inverted low pass filtered version of the original complex amplitude. The inversion is due to the imaging process. The two lenses cause the complex amplitude in plane E to be inverted relative to the complex amplitude in plane A, this is represented by $-\mathbf{x}$. The low pass filtering is due to $T(\xi)$. This removes all the high spatial frequencies, leaving only the low ones, so the complex amplitude in plane E may lose some fine detail.

Equation (2.9) describes the phase shifted complex amplitude with no approximations so the intensity, $I(\mathbf{x},\gamma)$, measured by the camera in plane E is

$$I(\mathbf{x},\gamma) = \left|\psi(\mathbf{x},\gamma)\right|^2. \tag{2.10}$$

This section has derived an equation that describes the measured intensity in term of the illuminating wavefront and parameters of the imaging system. This equation has been described with reference to phase contrast imaging, but it is sufficiently general to represent other types of common path interferometry. The following sections describe the process for calculating the phase from the measured images.

2.3.2. Calculating the Phase from a Single Image

The previous section derived an equation that described the measured intensity in plane E of Figure 2.1 in terms of the phase and intensity in plane A of Figure 2.1 and parameter of the imaging system. This section describes a process for calculating the phase after measuring or determining the remaining parameters using a single image.

The starting point is to simplify equation (2.9) by taking out a common factor in the last two terms to give

$$\begin{aligned}\psi(\mathbf{x},\gamma) = {} & A\sqrt{I(-\mathbf{x})}P(-\mathbf{x})\exp\left(-i\left(\alpha(-\mathbf{x})+\phi(-\mathbf{x})\right)\right)\otimes\mathfrak{F}^{-1}\left\{T(\boldsymbol{\xi})\right\} \\ & +\left(B\exp(i\gamma)-A\right)\mathfrak{F}^{-1}\left\{S(\boldsymbol{\xi})\mathfrak{F}\left\{P(\mathbf{x})\sqrt{I(\mathbf{x})}\exp\left(-i\left(\alpha(\mathbf{x})+\phi(\mathbf{x})\right)\right)\right\}\right\}.\end{aligned} \tag{2.11}$$

The first term in this summation represents the complete wave and the second term represents the part of the wave that passes through the reference area. The original part of the wave that passes through the reference area is subtracted from the complete wave and replaced with the phase-shifted version of the reference wave. Both these terms are dependent on the phase, $\phi(\mathbf{x})$, so are unknown.

To calculate the phase an approximation is applied to the reference wave. The second term can be rewritten, so that

$$\begin{aligned}\psi(\mathbf{x},\gamma) = {} & A\sqrt{I(-\mathbf{x})}P(-\mathbf{x})\exp\left(-i\left(\alpha(-\mathbf{x})+\phi(-\mathbf{x})\right)\right)\otimes\mathfrak{F}^{-1}\left\{T(\boldsymbol{\xi})\right\}+\left(B\exp(i\gamma)-A\right)\times \\ & \left(\begin{array}{l}\mathfrak{F}^{-1}\left\{S(\boldsymbol{\xi})\mathfrak{F}\left\{P(\mathbf{x})\sqrt{I(\mathbf{x})}\exp\left(-i\alpha(\mathbf{x})\right)\right\}\right\}+ \\ \mathfrak{F}^{-1}\left\{S(\boldsymbol{\xi})\mathfrak{F}\left\{P(\mathbf{x})\sqrt{I(\mathbf{x})}\exp\left(-i\alpha(\mathbf{x})\right)\left(\exp\left(-i\phi(\mathbf{x})\right)-1\right)\right\}\right\}\end{array}\right).\end{aligned} \tag{2.12}$$

Assuming that the phase is small, $\left|\phi(\mathbf{x})\right| \ll 1$, the reference wave, $R(\mathbf{x})$, can be approximated by the phase independent expression

$$R(\mathbf{x}) = \mathfrak{F}\left\{S(\boldsymbol{\xi})\mathfrak{F}\left\{P(\mathbf{x})\sqrt{I(\mathbf{x})}\exp\left(-i\alpha(\mathbf{x})\right)\right\}\right\}. \tag{2.13}$$

This is valid provided the phase change in plane A is small and the spatial frequencies that contain the phase information are predominantly confined to the region of the scattered wave, Beleggia [40]. The approximation of the reference wave and the consequences are considered in section 2.5.

The approximated reference wave cannot necessarily be evaluated to give a closed form expression but as it is independent of the phase, it can be calculated numerically when calculating the phase. Therefore, the complex amplitude in the measurement plane can be approximated as

$$\psi(-\mathbf{x},\gamma) \approx A\sqrt{I(\mathbf{x})}P(\mathbf{x})\exp\left(-i\left(\phi(\mathbf{x})+\alpha(\mathbf{x})\right)\right)\otimes\mathfrak{F}^{-1}\left\{T(\boldsymbol{\xi})\right\}+\left(B\exp(i\gamma)-A\right)R(\mathbf{x}). \tag{2.14}$$

This is the sum of the reference and scattered waves, which interfere causing modulations in the intensity. It is these modulations that relate to the phase.

To allow the phase to be calculated the convolution in the first term is removed by assuming that $T(\xi)$ is constant over the entire plane which is equivalent to stating that the information outside the band limit is negligible. Therefore, the phase shifted complex amplitude becomes

$$\psi(-\mathbf{x},\gamma) \approx A\sqrt{I(\mathbf{x})}P(\mathbf{x})\exp\left(-i\left(\phi(\mathbf{x})+\alpha(\mathbf{x})\right)\right)+\left(B\exp(i\gamma)-A\right)R(\mathbf{x}), \qquad (2.15)$$

so the expression for the intensity can be written in an algebraic form. The measured intensity is calculated by taking the magnitude squared of the phase shifted complex amplitude so that

$$I(-\mathbf{x},\gamma) \approx A^2 I(\mathbf{x})P(\mathbf{x})+\left(A^2+B^2-2AB\cos(\gamma)\right)\left|R(\mathbf{x})\right|^2+2AP(\mathbf{x})\sqrt{I(\mathbf{x})}\times \begin{Bmatrix} +\Im\{R(\mathbf{x})\}\left(\left(A-B\cos(\gamma)\right)\sin\left(\alpha(\mathbf{x})+\phi(\mathbf{x})\right)-B\sin(\gamma)\cos\left(\alpha(\mathbf{x})+\phi(\mathbf{x})\right)\right) \\ -\Re\{R(\mathbf{x})\}\left(\left(A-B\cos(\gamma)\right)\cos\left(\alpha(\mathbf{x})+\phi(\mathbf{x})\right)+B\sin(\gamma)\sin\left(\alpha(\mathbf{x})+\phi(\mathbf{x})\right)\right) \end{Bmatrix} \qquad (2.16)$$

where $\Re\{\ \}$ and $\Im\{\ \}$ represent the real and imaginary parts of a function. This equation approximates the measured intensity in plane E in terms of the parameters of the imaging system, and the phase and intensity in plane A. If all the quantities, except the phase can be known or measured then the phase can be calculated by solving this equation.

The reference wave has already been approximated by a phase independent expression, (2.13). Therefore, to yield a closed form analytic expression for the phase, a small angle approximation can be applied to the phase. By expanding the trigonometric functions in (2.16) using the identities

$$\begin{aligned} \cos(A\pm B) &= \cos A\cos B \mp \sin A\sin B \\ \sin(A\pm B) &= \sin A\cos B \pm \cos B\sin A \end{aligned} \qquad (2.17)$$

and applying the small angle approximation to the phase, $\phi(\mathbf{x})$, the intensity can be written as

$$
\begin{aligned}
I(-\mathbf{x},\gamma) \approx\ & A^2 I(\mathbf{x}) P(\mathbf{x}) + \left(A^2 + B^2 - 2AB\cos(\gamma)\right)\left|R(\mathbf{x})\right|^2 \\
& +2A^2 P(\mathbf{x})\sqrt{I(\mathbf{x})}\left(\Im\{R(\mathbf{x})\}\sin(\alpha(\mathbf{x})) - \Re\{R(\mathbf{x})\}\cos(\alpha(\mathbf{x}))\right) \\
& +2ABP(\mathbf{x})\sqrt{I(\mathbf{x})}\left(\Re\{R(\mathbf{x})\}\cos(\alpha(\mathbf{x})+\gamma) - \Im\{R(\mathbf{x})\}\sin(\alpha(\mathbf{x})+\gamma)\right) \\
& +2A^2 P(\mathbf{x})\sqrt{I(\mathbf{x})}\left(\Im\{R(\mathbf{x})\}\cos(\alpha(\mathbf{x})) + \Re\{R(\mathbf{x})\}\sin(\alpha(\mathbf{x}))\right)\phi(\mathbf{x}) \\
& -2ABP(\mathbf{x})\sqrt{I(\mathbf{x})}\left(\Re\{R(\mathbf{x})\}\sin(\alpha(\mathbf{x})+\gamma) + \Im\{R(\mathbf{x})\}\cos(\alpha(\mathbf{x})+\gamma)\right)\phi(\mathbf{x}).
\end{aligned}
\tag{2.18}
$$

For clarity, the following functions are defined

$$
\begin{aligned}
\nu(\mathbf{x}) =\ & A^2 I(\mathbf{x}) P(\mathbf{x}) + \left(A^2 + B^2 - 2AB\cos(\gamma)\right)\left|R(\mathbf{x})\right|^2 \\
& +2A^2 P(\mathbf{x})\sqrt{I(\mathbf{x})}\left(\Im\{R(\mathbf{x})\}\sin(\alpha(\mathbf{x})) - \Re\{R(\mathbf{x})\}\cos(\alpha(\mathbf{x}))\right) \\
& +2ABP(\mathbf{x})\sqrt{I(\mathbf{x})}\left(\Re\{R(\mathbf{x})\}\cos(\alpha(\mathbf{x})+\gamma) - \Im\{R(\mathbf{x})\}\sin(\alpha(\mathbf{x})+\gamma)\right) \\
\omega(\mathbf{x}) =\ & 2A^2 P(\mathbf{x})\sqrt{I(\mathbf{x})}\left(\Im\{R(\mathbf{x})\}\cos(\alpha(\mathbf{x})) + \Re\{R(\mathbf{x})\}\sin(\alpha(\mathbf{x}))\right) \\
& -2ABP(\mathbf{x})\sqrt{I(\mathbf{x})}\left(\Re\{R(\mathbf{x})\}\sin(\alpha(\mathbf{x})+\gamma) + \Im\{R(\mathbf{x})\}\cos(\alpha(\mathbf{x})+\gamma)\right),
\end{aligned}
\tag{2.19}
$$

so that the intensity can be written as

$$
I(-\mathbf{x},\gamma) \approx \nu(\mathbf{x}) + \omega(\mathbf{x})\phi(\mathbf{x}). \tag{2.20}
$$

Therefore, the phase can be calculate from the intensity using

$$
\phi(\mathbf{x}) \approx \frac{I(-\mathbf{x},\gamma) - \nu(\mathbf{x})}{\omega(\mathbf{x})}, \tag{2.21}
$$

as all the other quantities are known or can be measured and assuming that the phase is weak and that light is not scattered outside the band limit.

2.3.3. Calculating the Phase from Multiple Images

In this section, a method for calculating the phase is developed using three images, two with equal and opposite phase shifts applied during the imaging process and a third image with no phase shift. There is no need for the phase shifts to be equal and opposite as two arbitrary phase shifts can be used but the equations become more complicated.

The phase is calculated using the sum and difference of the phase contrast images. The sum and difference are used as each gives a different expression of the real and imaginary parts of the complex amplitude. The third image, $I(\mathbf{x},0)$, has no additional phase shifts applied during the imaging process and used for normalisation. It is the

same intensity distribution as plane A of Figure 2.1 assuming that there is no light scattered outside the band limit.

The intensity in plane E of Figure 2.1 is approximately described by equation (2.16). This equation was derived in the previous section and has had approximation applied to the reference wave. Using this equation the difference between the two phase shifted images with equal and opposite phase shifts is approximately described by

$$\begin{aligned}\Delta I(-\mathbf{x}) \approx & -4AB\sqrt{I(\mathbf{x})}P(\mathbf{x})\sin(\gamma) \\ & \times\left[\Re\{R(\mathbf{x})\}\sin(\phi(\mathbf{x})+\alpha(\mathbf{x}))+\Im\{R(\mathbf{x})\}\cos(\phi(\mathbf{x})+\alpha(\mathbf{x}))\right]\end{aligned} \tag{2.22}$$

and the sum of the two images is

$$\begin{aligned}\Sigma I(-\mathbf{x}) \approx & 2\left(A^2P(\mathbf{x})I(\mathbf{x})+|R(\mathbf{x})|^2\left(A^2+B^2-2AB\cos(\gamma)\right)\right) \\ & +4A\sqrt{I(\mathbf{x})}P(\mathbf{x})\left(B\cos(\gamma)-A\right) \\ & \times\left[\Re\{R(\mathbf{x})\}\cos(\phi(\mathbf{x})+\alpha(\mathbf{x}))-\Im\{R(\mathbf{x})\}\sin(\phi(\mathbf{x})+\alpha(\mathbf{x}))\right].\end{aligned} \tag{2.23}$$

For simplicity, the following functions are defined

$$\begin{aligned}\mu(\mathbf{x}) &= 2\left(A^2P(\mathbf{x})I(\mathbf{x})+|R(\mathbf{x})|^2\left(A^2+B^2-2AB\cos(\gamma)\right)\right) \\ \eta(\mathbf{x}) &= 4AP(\mathbf{x})\sqrt{I(\mathbf{x})}\left(B\cos\gamma-A\right) \\ \rho(\mathbf{x}) &= -4ABP(\mathbf{x})\sqrt{I(\mathbf{x})}\sin\gamma\end{aligned} \tag{2.24}$$

where the third intensity measurement is used here in place of the intensity in plane A of Figure 2.1.

Therefore, the sum and difference can be written as

$$\begin{aligned}\Sigma I(-\mathbf{x}) &\approx \mu(\mathbf{x})+\eta(\mathbf{x})\left(\Re\{R(\mathbf{x})\}\cos(\phi(\mathbf{x})+\alpha(\mathbf{x}))-\Im\{R(\mathbf{x})\}\sin(\phi(\mathbf{x})+\alpha(\mathbf{x}))\right) \\ \Delta I(-\mathbf{x}) &\approx \rho(\mathbf{x})\left(\Re\{R(\mathbf{x})\}\sin(\phi(\mathbf{x})+\alpha(\mathbf{x}))+\Im\{R(\mathbf{x})\}\cos(\phi(\mathbf{x})+\alpha(\mathbf{x}))\right),\end{aligned} \tag{2.25}$$

these two equations treated as simultaneous equations can be solved for the sine and the cosine of the phase so that

$$\begin{aligned}\sin(\phi(\mathbf{x})+\alpha(\mathbf{x})) &\approx \frac{1}{|R(\mathbf{x})|^2}\left(\frac{\Re\{R(\mathbf{x})\}\Delta I(-\mathbf{x})}{\rho(\mathbf{x})}-\frac{\Im\{R(\mathbf{x})\}\left(\Sigma I(-\mathbf{x})-\mu(\mathbf{x})\right)}{\eta(\mathbf{x})}\right) \\ \cos(\phi(\mathbf{x})+\alpha(\mathbf{x})) &\approx \frac{1}{|R(\mathbf{x})|^2}\left(\frac{\Im\{R(\mathbf{x})\}\Delta I(-\mathbf{x})}{\rho(\mathbf{x})}+\frac{\Re\{R(\mathbf{x})\}\left(\Sigma I(-\mathbf{x})-\mu(\mathbf{x})\right)}{\eta(\mathbf{x})}\right).\end{aligned} \tag{2.26}$$

All the quantities on the right, of each equation, can be measured or are known. Therefore, using this equation, the complex amplitude can be calculated and the phase can be calculated using

$$\phi(\mathbf{x}) = \arctan\left(\frac{\sin\left(\phi(\mathbf{x})+\alpha(\mathbf{x})\right)}{\cos\left(\phi(\mathbf{x})+\alpha(\mathbf{x})\right)}\right) - \alpha(\mathbf{x}). \tag{2.27}$$

This section has described the theory of a general common path interferometer. This gives an equation relating the measured intensity to the phase and parameters of the imaging system. Having derived an equation that describes the physical system a method has been developed that uses three phase contrast images to calculate the phase.

The phase can be calculated from two phase contrast images and a third image for normalisation. Two phase contrast images can also be used to calculate the phase where the average intensity value calculated from the two images is used for normalisation instead of the third image. Finally, as described in the previous section, a single phase contrast image can be used to calculate the phase where the normalisation factor is taken to be the average intensity value calculated from this image.

Having developed the theory, the following section describes the numerical implementation of the equations for use with sampled data and the process for evaluating the calculated phase.

2.4. Implementation and Evaluation of the Numerical Simulations

In the previous section, theory was developed that describes a general common path interferometer. The equations describing the measured data were then inverted to give the phase in terms of known parameters and measurable data. This section describes the process for numerically calculating the phase from the measured data.

2.4.1. Numerical Implementation of the Theory

This section describes the numerical process to generate simulated intensity data and to apply the equation derived in the previous chapter to evaluate this data.

To simulate the intensity in plane E, in Figure 2.1, the complex amplitude is first defined in plane A. The phase and amplitude of the wave are defined on a sampled grid, these are combined using equation (2.5) to give the complex amplitude in plane A of

Figure 2.1. Taking the Fourier transform of the complex amplitude gives the wave in plane C. To apply the phase change, the wave in phase C is multiplied by equation (2.7), with the parameters set depending on the type of images being simulated. A second Fourier transform give the phase shifted complex amplitude at the detector, plane E. The simulated intensity is the magnitude squared of the phase shifted complex amplitude.

The simulated data used to calculate the phase is generated following the process above, to give two images with equal and opposite phase shifts applied during the imaging process and a third image generated with no phase shift.

The sum and difference of the images with phase shifts are then used as the input for equation (2.26). The reference wave is approximated using equation (2.13) and the remaining functions are defined following equation (2.24). Equation (2.26) calculates the sine and cosine of the phase, from which the phase can be calculated using equation (2.27).

2.4.2. Sampling Continuous Functions

The theory developed in section 2.2 represented the complex amplitude and the optical components using continuous functions. For numerical implementation of the process to calculate the phase, the input data will be defined on a sampled gird. Therefore, this section describes sampling a function in such a way that the function can unambiguously be determined from the sampled values, and states discrete versions of the Fourier transform that will replace the Fourier transforms used in the theory.

A continuous function can be represented by an infinite set of samples so that

$$\psi(x) = T\sum_{n=-\infty}^{\infty}\psi(nT)\frac{\sin\left(2\pi f_c(x-nT)\right)}{\pi(x-nT)} \tag{2.28}$$

where T is the interval between the samples indexed by n. The sampling theorem states that if the Fourier transform of $\psi(x)$ is zero for all frequencies greater than some critical frequency, f_c, then the continuous function is uniquely determined from its sampled values provided the sample spacing is $T \le 1/2 f_c$. If the sample spacing is $T = 1/2 f_c$ then the function is sampled at Nyquist sampling rate, this is the maximum sample spacing which still allows the function to be unambiguously determined from the sampled values.

If the function is band limited and sampled at the correct rate then the discrete Fourier transform is equivalent to the continuous Fourier transform Brigham [38].

In the numerical simulations, the discrete Fourier transform will be used in place of the continuous Fourier transform. In 1D the discrete Fourier transform is defined, Brigham [38] as

$$\Psi\left(\frac{n}{NT}\right)=\frac{1}{N}\sum_{k=0}^{N-1}\psi\left(kT\right)\exp\left(\frac{-i2\pi nk}{N}\right) \tag{2.29}$$

and the inverse is defined as

$$\psi\left(kT\right)=\sum_{k=0}^{N-1}\Psi\left(\frac{n}{NT}\right)\exp\left(\frac{i2\pi nk}{N}\right). \tag{2.30}$$

The functions Ψ and ψ are sampled evenly at N points indexed by $n\in\left[0,N-1\right]$ and $k\in\left[0,N-1\right]$. This is therefore an approximation of the continuous Fourier transform with the two being equivalent in the limit as $N\to\infty$. To reduce the computational time for this calculation a fast Fourier transform can be implemented. This is an algorithm that speeds up the calculation of the discrete Fourier transform, Brigham [38].

In the following simulations the phase is defined by a function with a discontinuity, due to the apertures edge. Therefore, these functions are not band limited so cannot be sampled at the Nyquist sampling rate using a finite number of samples.

As the discontinuity cannot be represented using a finite number of spatial frequencies, the recorded images will be under sampled so any method for calculating the phase will need to work with this under sampled data implied by a low pass filter.

The number of samples representing the physical object is therefore defined from the pixelation of the measured images and the sample spacing defines an implicit band limit shown in plane C of Figure 2.1. Provided the energy outside the band limit is minimal then the sampled data should be adequate for use to determine and represent the phase function. The lost energy will cause the measured imaged to have a smooth edge instead of a sharp edge where the edge of the aperture would be.

There will be a limit on this approximation as the phase function can scatter energy outside the band limit but it has been shown that this process has been used where the

number of samples representing the disk is low, Blanchard, *et al.* [41]. The number of samples will therefore effect the calculated phase, this is considered in Chapter 5.

2.4.3. Maximum Scattering Angle

There is a limit on the spatial frequency of the waves that propagate and contribute to a measurement. Waves with a spatial frequency below $1/\lambda$ propagate, above $1/\lambda$ the waves become evanescent and decay exponentially. Therefore, the contribution of the evanescent waves to a measurement, a few wavelengths from the object, is negligible.

The change between propagating waves and evanescent waves defines the maximum spatial frequency of the propagating waves that contribute a measurement. As mentioned in section 2.2.2 spatial frequencies can be converted to scattering angles. Therefore, the change to evanescent wave defined the maximum scattering angle of the system.

In a real experiment, the resolution of the camera defines the resolution in Fourier space, which can be considered as a band limit on the data. This defines the maximum spatial frequency of the waves that can unambiguously be defined form the sampled data. This can also be considered as an angle. The band limit and the spatial frequency where the waves become evanescent can be used to calculate the maximum scattering angle of the system.

By considering the direction cosines in the angular spectrum of plane waves, Goodman [37], the scattering angle between the propagation direction, along the z axis, and an arbitrary plane wave can be defined as $\sigma = \cos^{-1}\left(\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2}\right)$ so that

$$\sin\sigma = \sqrt{(\lambda f_x)^2 + (\lambda f_y)^2} \tag{2.31}$$

where f_x and f_y are the coordinates in the x and y direction in Fourier space. The maximum scattering angle, $\sigma_{\max}$ is then calculated from the radius of the band limit and of the radius where the waves become evanescent, so that

$$\sin\sigma_{\max} = \lambda \min\{R_E, R_B\} = \lambda \min\left\{\frac{1}{\lambda}, \frac{1}{2\mathrm{T}}\right\} \tag{2.32}$$

where the radius of the evanescent filter is R_E and equals $1/\lambda$ and the radius of the band limit, R_B, for Nyquist sampling is $1/2T$ where T is the sample spacing.

2.4.4. Evaluation of the Calculated Phase

After calculating the phase, a method is required for evaluating it in comparison to the set phase. This section describes the error metrics used throughout this thesis to evaluate the calculated phase.

The first metric used to evaluate the phase is the root mean square (RMS) error between the set and the calculated phase. This gives a measure of the overall error in the calculated phase. The RMS error metric is defined, Fienup [42], as

$$RMS = \left(\frac{1}{N} \sum_{n=1}^{N} \left(\phi_{Output}\left(\mathbf{x}_n\right) - \phi_{Input}\left(\mathbf{x}_n\right) \right)^2 \right)^{\frac{1}{2}}. \tag{2.33}$$

Where the N samples within the area of the aperture at position $\mathbf{x}_n$ are indexed by n. The set and calculated phases are represented by $\phi_{Input}\left(\mathbf{x}_n\right)$ and $\phi_{Output}\left(\mathbf{x}_n\right)$. Before applying this metric, the mean value is subtracted from both the set and calculated phase so that the correct phase corresponds to zero error. Removing the mean value is valid as only relative phase changes can be determined.

This method to evaluate the calculated phase gives a measure of the overall error. As an alternative the calculated phase can be decomposed into a set base functions, this gives a measure of the accuracy of individual modes. This also allows the phase to be represented by a set of number instead of a picture. This method is used in ophthalmology to look at the quantity of common aberrations in the phase, He, *et al.* [43].

Two sets of orthonormal base functions are used throughout this thesis to define the phase as an input for the simulations. In this chapter, trigonometric functions are used and in Chapter 3 and Chapter 5 Zernike polynomials are used. The set of Zernike polynomials are used as they represent common optical aberrations and trigonometric functions are used as the frequency of the wave can be defined so that the phase information is predominantly contained in the scattered wave.

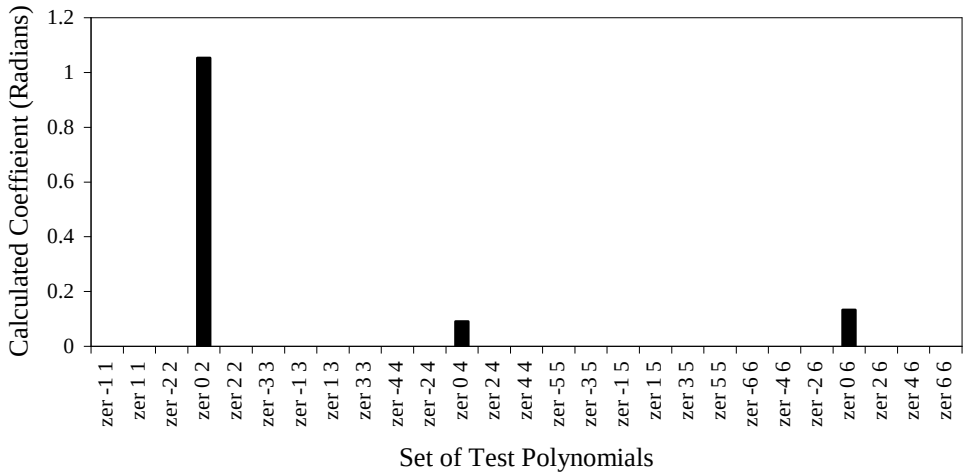


Figure 2.3 Projection of Z_2^0 on to a subset of the Zernike polynomials specified along the *x* axis

Throughout this thesis, the phase is expressed as a linear combination of orthogonal base functions or modes, $\phi_n(\mathbf{x})$, such that

$$\phi(\mathbf{x}) = \sum_{n=1}^{N} c_n \phi_n(\mathbf{x}), \tag{2.34}$$

with corresponding coefficients c_n. The phase is defined in this way as decomposition of a function naturally results in a set of base functions and a set of corresponding coefficients. To evaluate the calculated phase, the coefficient of the input mode can be compared to the coefficient of the same mode calculated from the estimated phase.

The coefficients of the functions could be calculated by applying the orthogonality property to decompose the phase. This method is not used as Zernike polynomials are only orthogonal as continuous functions and these properties do not translate directly to sampled functions, Schwiegerling, *et al.* [44].

As an example of the lack of orthogonality, if the phase is set as $\phi(\mathbf{x}) = Z_2^0$ it can be decomposed into a set of Zernike polynomials as shown in Figure 2.3. It shows the base function along the x axis and the calculated coefficient along the y axis.

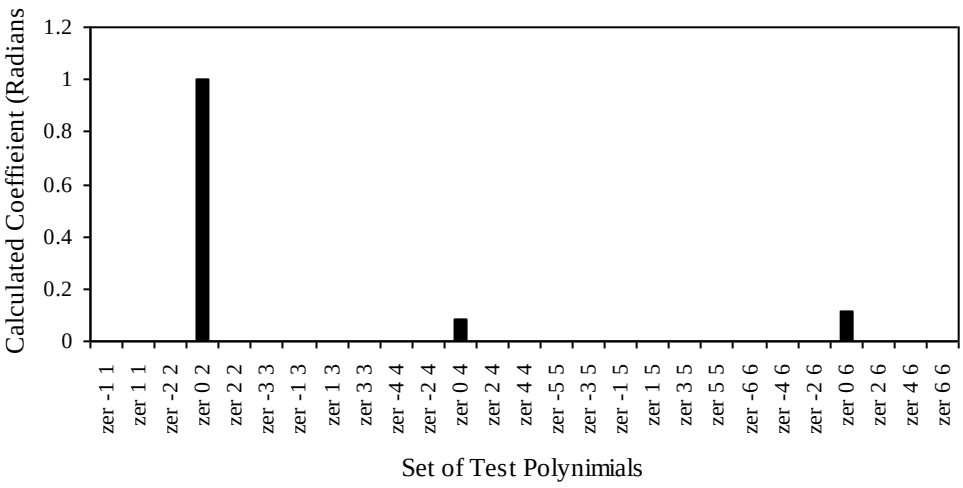


Figure 2.4 Fit of a set of test functions defined along the x axis to Z_2^0.

The coefficient of a particular mode, c_n, can be calculated from

$$c_n = w_n \int_A \phi(\mathbf{x}) \phi_n(\mathbf{x}) d\mathbf{x}. \tag{2.35}$$

The phase, $\phi(\mathbf{x})$, and the mode, $\phi_n(\mathbf{x})$, are multiplied together and integrated over the defined area of the function, in this case a disk of area A. The scale factor w_n is required so the modes are orthonormal instead of orthogonal.

Figure 2.3 shows an overestimation of the Z_2^0 coefficient where it is estimated as 1.05 radians instead of 1 radian. There are also non zero coefficients of other modes, Z_4^0 and Z_6^0, that were not present in the set phase. These coefficients should be zero as the set phase is a single mode.

The lack of orthonormality is caused by a finite number of samples, as the integral in (2.35) is approximated by a Riemann sum. Therefore, increasing the number of samples will reduce the error but there can only ever be a finite number of samples in experimental data so there will always be some error in the estimated coefficient.

To decompose the phase correctly, the set of base functions should be orthogonalised on the discrete grid. The phase can then be projected onto this set of functions and if required the coefficients in this set can be mapped back to the original set, but this is computationally expensive due the large size of the matrices.

As an alternative method for estimating the phase coefficients, the base functions can be fit numerically to the phase. The can be achieved by varying the coefficient of the base function to minimise the mean square error between the phase and the base function.

Using the same set phase as the previous figure, Figure 2.4 shows the fit of a set of Zernike polynomials defined along the x axis to the set phase. Figure 2.4 shows that there are still extra modes incorrectly calculated but the correct coefficient of the set mode is calculated.

If the phase is defined as a single mode and the calculated phase is correct, this method will return the correct coefficient. All the examples in this thesis except Chapter 6 use a single mode to define the phase. Therefore, any deviation from the correct value of the coefficient is an indication of error in the calculated phase.

2.5. Benefits and Limitations

So far, this chapter has introduced the theory describing a general common path interferometer and the process for calculating the phase from the measured images. The following sections look at the benefits of using three images to calculate the phase in comparison to a single image, and the limitations due to the approximation in the model.

2.5.1. Benefits of the 3 Image Method

This section considered the improvement in accuracy of the phase calculated from three images in comparison to one or two images.

In the examples the phase is defined as $\phi(\mathbf{x}) = A\cos(2\pi fx)$. Single frequency cosine waves are used so that the frequency of the wave can be specified. Therefore, as far as possible the phase information is confined to known regions of the scattered wave. Throughout this chapter, the aperture has a radius r_0 and the frequency, f, of the cosine wave has units of cycles per radius. In Fourier space, the coordinate ξ has units $\lambda/2r_0$ so all the sizes of the objects are appropriately scaled and non-dimensionalised.

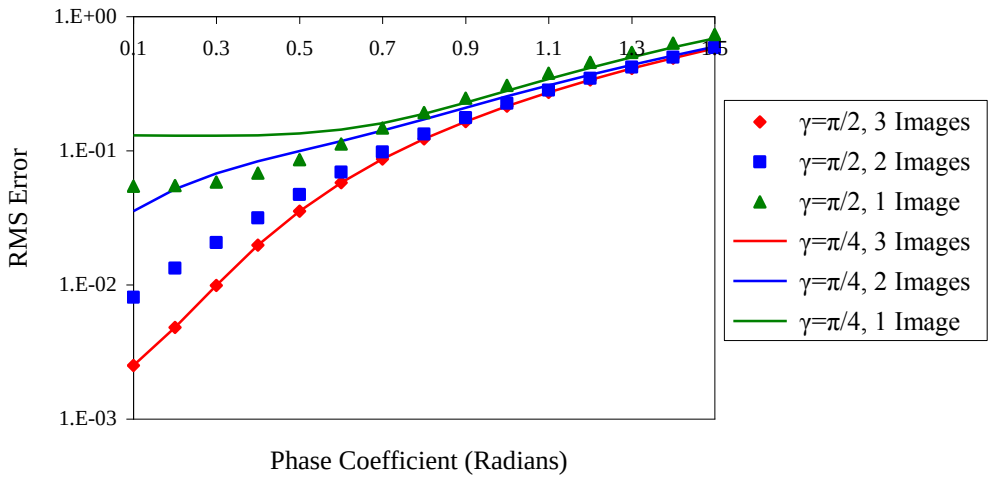


Figure 2.5 RMS error plotted as a function of the input phase coefficient for difference phase shifts and combinations of images indicated in the legend.

In this example the frequency of the wave representing the phase is 5 cycles per radius and the coefficient of the phase A is varied, the area of the reference wave is defined as a circle with a radius of $\lambda/4r_0$. The aperture radius is defined as r_0 metres represented by 64 samples in a grid of 256×256 samples.

The intensity data is generated applying different phase shifts to the reference wave and the data is then analysed using the three different methods, the results are shown in Figure 2.5. It shows the RMS phase error as a function of the input phase coefficient. For each combination of images, the figure shows that as the phase coefficient increases, the phase error increases. This is due to the approximation of the reference wave, which is common to each method.

The difference between the individual methods is due to extra approximations. The difference between using three images and two images is due to normalisation, caused by approximating the third image. The difference between using two images and a single image is due to the additional small angle approximation applied to the scattered wave.

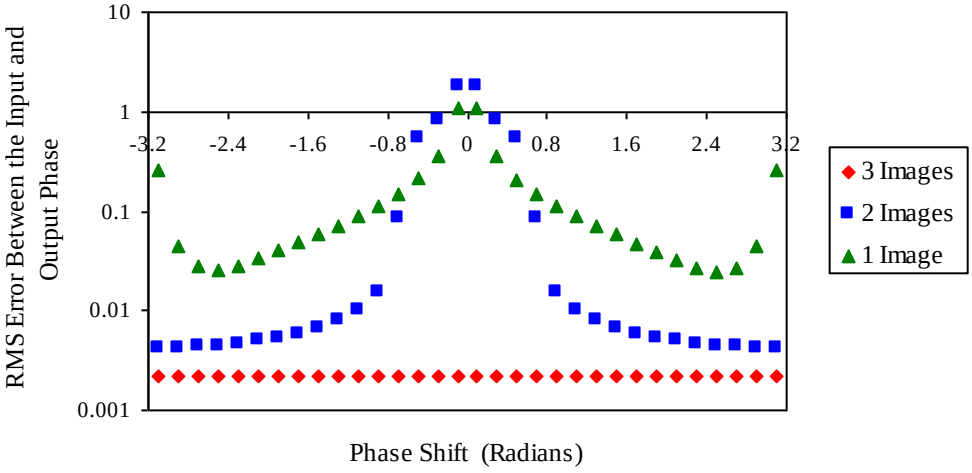


Figure 2.6 RMS error between the set and calculated phase as a function of the phase shift, γ. The phase is calculated from one, two and three images. The phase coefficient is set at 0.1 radians.

For phase coefficients larger than approximately 0.5 radians all three methods give poor approximations of the phase as expected from the approximation of the reference wave, but within the limitations of the equations there is a significant reduction in the error using three images in comparison to using a single image. In this example, there is a reduction in the error metric by a factor of between 2 and 20 depending on phase coefficient.

Figure 2.5 and other tests show that using three measurements will result in a phase estimate with the lowest error, due to fewer approximations, and suggests that the calculated phase is independent of the phase shift, γ, as the error is the same when the phase shift is $\pi/2$ or $\pi/4$ radians.

To test the dependence of the calculated phase, $\phi(\mathbf{x})$, on the phase shift, γ, simulations were run using a set input phase and varying the phase shift to generate different input data which was then used to calculate the phase.

Figure 2.6 shows the RMS error between the set phase and the calculated phase as a function of the applied phase shift. The phase is calculated using the 1, 2 and 3 image methods. The figure shows that when using one or two images as the input data, the error depends on the applied phase shift. Only when using three images is the error constant and independent of the phase shift. There is a significant improvement in the

error by a factor of at least 10 using three images in comparison to one. The constant error in the calculated phase using three images is due to the estimation of the reference wave.

As the phase error is independent of the phase shift, in an experiment the value of the phase shift only needs to be calibrated once the experiment has been assembled and not specified when designing the experiment. Anderson [45] has shown that when there is noise present and high contrast is required then the value of the applied phase shift becomes important.

Using the same parameters as Figure 2.5, with a phase coefficient of 0.8 radians, Figure 2.7 shows the contrast in the difference between the two phase contrast images as a function of the phase shift.

The contrast is defined as $\max\left\{I\left(\mathbf{x},\gamma\right)-I\left(\mathbf{x},-\gamma\right)\right\}-\min\left\{I\left(\mathbf{x},\gamma\right)-I\left(\mathbf{x},-\gamma\right)\right\}$, where $\max\{\ \}$ and $\min\{\ \}$ are the largest and smallest values of the function in the brackets. The maximum points in Figure 2.7 mark the highest contrast so in this particular case the optimum phase shift, to give the highest contrast is approximately $\pm\pi/2$ radians. The symmetry of the graph is a result of applying positive and negative phase shifts together to generate the data.

This section has shown that the contrast depends on the phase shift and the accuracy of estimated phase is independent of the phase, the phase shift is a free parameter that can be used to increase the contrast without affecting the accuracy of the estimated phase.

This section has looked at the benefits of using three images to calculate the phase with the equations derived in section 2.2. It has shown that there is a significant reduction in the error between the set and calculated phase when using three images compared to using a single image.

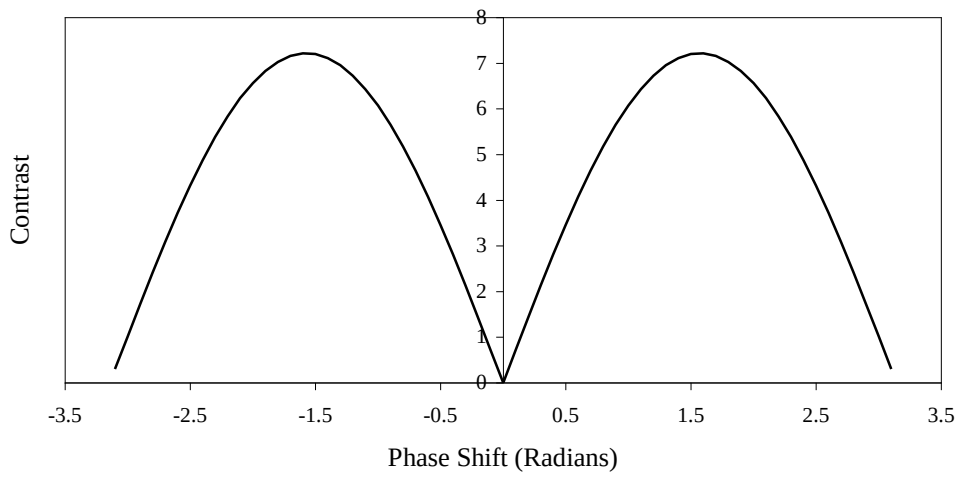


Figure 2.7 Contrast of the difference data plotted as a function of phase shift, γ.

The phase, $\phi(\mathbf{x})$, calculated from three images was also shown to be independent of the phase shift, γ, unlike the phase calculated from a single image. This removes another source of error from the calculated phase. Additionally, this is particularly useful experimentally as the phase shift can be calibrated instead of being specified and the phase shift can be varied to give the required high contrast without affecting the accuracy of the estimated phase.

2.6. Iterative Refinement of the Phase Solution

When calculating the phase there are always errors due to the approximations, therefore the section considers using an iterative process to reduce these errors and to extending the use of these equation beyond the small angle approximation.

To reduce the error in the calculated phase an iterative scheme can be constructed by defining the phase error, $\phi_{error}(\mathbf{x})$, using

$$\exp\left(-i\phi_{error}(\mathbf{x})\right) = 1 + \exp\left(-i\left(\phi_{true}(\mathbf{x}) + \alpha(\mathbf{x})\right)\right) - \exp\left(-i\left(\phi_n(\mathbf{x}) + \alpha(\mathbf{x})\right)\right) \qquad (2.36)$$

where $\phi_{true}(\mathbf{x})$ and $\phi_n(\mathbf{x})$ represent the true phase and the nth phase estimate. The constant one is included so that the estimated complex amplitude equals the real complex amplitude when the phase error is zero. Converting this into trigonometric functions and substituting in a rearranged version of equation (2.26), the complex exponential of the phase error can be written as

$$\exp\left(-i\left(\phi_{error}\left(\mathbf{x}\right)\right)\right) \approx 1 + \frac{1}{\left|R\left(\mathbf{x}\right)\right|^2}\left(\frac{\Im\left\{R\left(\mathbf{x}\right)\right\}\left(\Delta I_M - \Delta I_{est}\right)}{\rho} + \frac{\Re\left\{R\left(\mathbf{x}\right)\right\}\left(\Sigma I_M - \Sigma I_{est}\right)}{\eta}\right)$$
$$- \frac{1}{\left|R\left(\mathbf{x}\right)\right|^2}\left(\frac{\Re\left\{R\left(\mathbf{x}\right)\right\}\left(\Delta I_M - \Delta I_{est}\right)}{\rho} - \frac{\Im\left\{R\left(\mathbf{x}\right)\right\}\left(\Sigma I_M - \Sigma I_{est}\right)}{\eta}\right). \tag{2.37}$$

The differences between the measured and nth estimate of the phase contrast images are denoted by ΔI_M and ΔI_n. Similarly, the sum of the measured phase contrast images is denoted by ΣI_M and the nth estimate of the phase contrast images is denoted by ΣI_n. The phase error therefore depends on the difference between the measured and estimated images.

The estimated data is calculated from

$$\psi\left(\mathbf{x}\right) = \sqrt{I\left(\mathbf{x}\right)}\exp\left(-i\left(\phi_n\left(\mathbf{x}\right) + \alpha\left(\mathbf{x}\right)\right)\right) \tag{2.38}$$

using the same process for generating the data, as described in 2.4.1, and the reference wave is defined as

$$R\left(\mathbf{x}\right) = \mathfrak{F}\left\{S\left(\xi\right)\mathfrak{F}\left\{P\left(\mathbf{x}\right)\sqrt{I\left(\mathbf{x}\right)}\exp\left(-i\left(\alpha\left(\mathbf{x}\right) + \phi_n\left(\mathbf{x}\right)\right)\right)\right\}\right\}. \tag{2.39}$$

The addition of the nth phase estimate, $\phi_n\left(\mathbf{x}\right)$ when estimating the reference wave starts to account for the missing terms ignored in the initial approximation of the reference wave.

The phase error is calculated from equation (2.37), this is then added to the phase estimate to give a new phase estimate,

$$\phi_{n+1}\left(\mathbf{x}\right) = \phi_n\left(\mathbf{x}\right) + \phi_{error}\left(\mathbf{x}\right) \tag{2.40}$$

The estimated data and the reference wave are then recalculated and the process is repeated. Recalculating the reference wave starts to account for the neglected terms in (2.12) which were assumed to be zero.

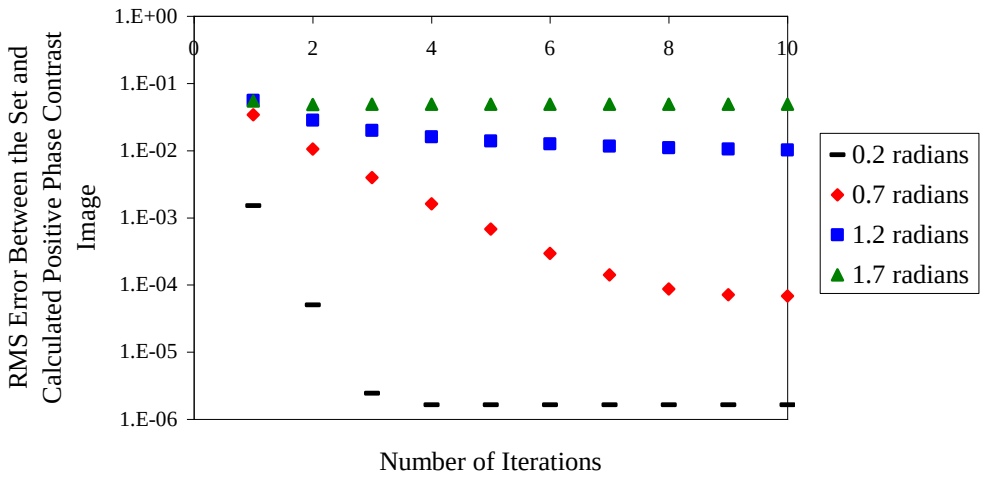
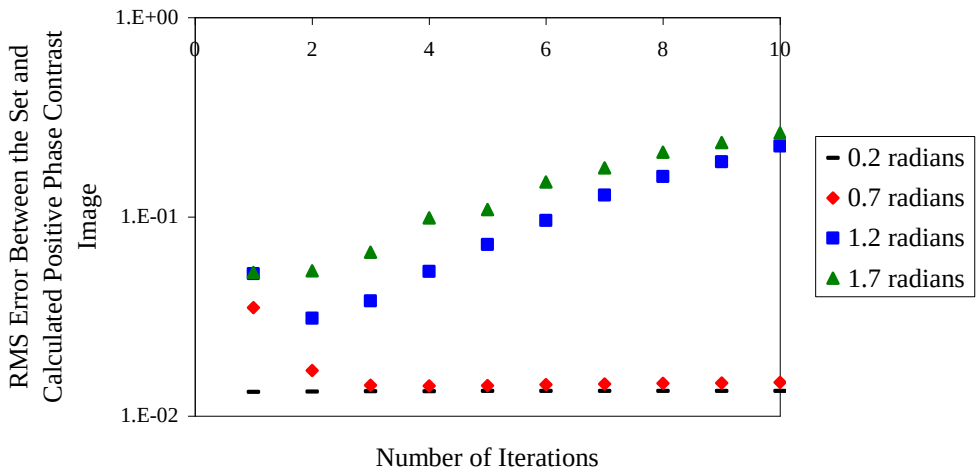


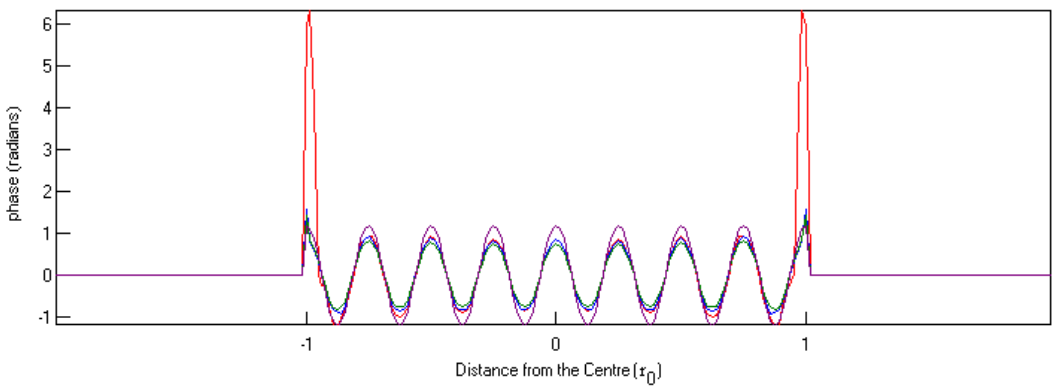
Figure 2.8 RMS Error plotted as a function of the number of iterations. The error is between the measured and the recalculated positive phase shifted intensity for various input phase coefficients (radians) defined in the legend.

To assess the benefits of this iterative scheme, simulations were constructed with a set phase and an increasing phase coefficient. The initial estimate of the phase is from the analytic solution, described in section 2.3.3, therefore increasing the phase coefficient results in a poorer initial estimate of the phase. In the following examples, the RMS error is between the "measured" and estimated data since in a real experiment, the phase is unknown and only the intensities can be measured. An error of zero indicates that the estimated phase can be used to generate data that equals the measured data.

Figure 2.8 shows the RMS error as a function of the number of iteration, where a finite sized phase object is placed on an infinitely extended background, similar to a cell suspended in a fluid. Although there is a discontinuity in the phase, the amplitude is continuous. The figure shows that provided the phase is small, so that the equation is valid, after only a few iterations there is a considerable reduction in the error.



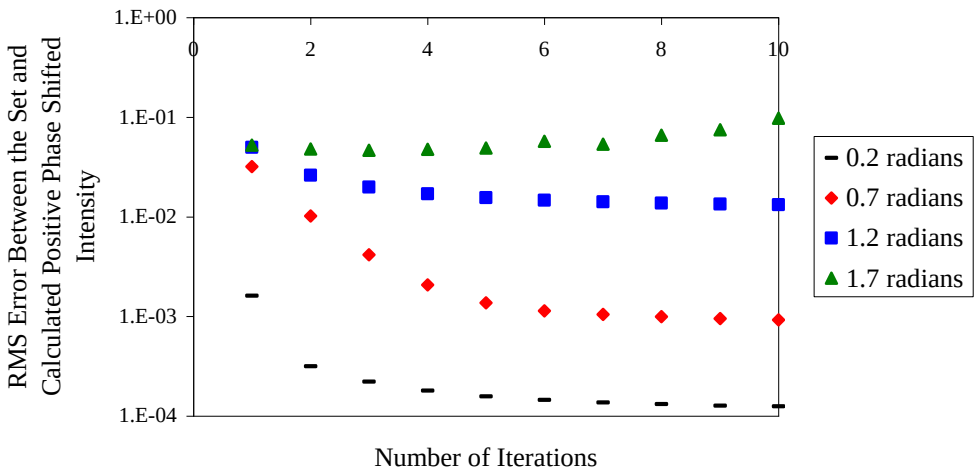
a)



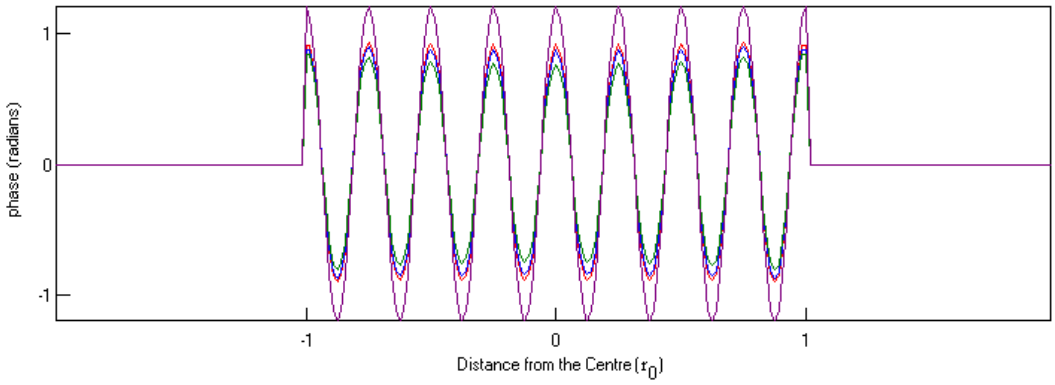
b)

Figure 2.9 Results of the iterative scheme applied to a system where there is a discontinuity in the phase and the amplitude. a) shows the RMS error plotted as a function of the number of iterations for different input coefficients. b) shows cross sections of the phase during the iterative process. Purple represents the set phase, green represents the phase after 1 iteration, blue represents the phase after 3 iterations and red represents the phase after 5 iterations.

In the following example, the phase and amplitude are defined over a finite area, as opposed to an infinitely extended background, so there is a discontinuity in the phase and amplitude. Figure 2.9 a) shows that the iterative scheme does not appear to tend towards zero, within the limited number of iterations, even for the smallest phase coefficients. Figure 2.9 b) shows a cross section of the phase during the iterative process, where the set phase coefficient is 1.2 radians. As the iterative process progresses, the phase estimate becomes worse. The problem starts from the position of the amplitude discontinuity where the phase error does not improve the phase estimate causing an increased error in the estimated data. This may cause the phase error not to converge to zero or diverge.



a)



b)

Figure 2.10 Results of the iterative scheme applied to a system where there is a discontinuity in the phase and the amplitude. The third image is replaced with the average intensity calculated from this image. a) shows the RMS error plotted as a function of the number of iterations for different input coefficients. b) shows cross sections of the phase during the iterative process. Purple represents the set phase, green represents the phase after 1 iteration, blue represents the phase after 3 iterations and red represents the phase after 5 iterations.

To avoid this problem normalisation using the third image is replaced with the average intensity value calculated from the third image. Figure 2.10 a) shows that as the number of iterations increases the iterative scheme appears to reduce the error, except for the example when the phase coefficient is set to be very large at 1.7 radians. Cross sections of the phase during the iterative process, for the 1.2 radian case are shown in Figure 2.10 b), as the iterative scheme progresses there are no adverse effects at the amplitude discontinuity and scheme appears to be converging, as, after each iteration, the cross section of the phase is closer to the set phase.

Convergence of this iterative scheme has not been proven so this process is not guaranteed to converge to the set phase. Additionally, the process may stagnate, where successive iterations only produce a minimal change to the estimated phase even though the estimation is not correct. The issue of stagnation is a problem with some of the iterative process as described by Fienup [42]. Although these are issues when implementing the scheme for a specific example, this thesis considers the general process for determining phase so these issues are out with the scope of the thesis.

This section has stated an iterative scheme to refine the calculated phase. As expected, the iterative process is limited as it is based on equations that are restricted to use with weak phase objects. Within the limitations of this scheme, there is a significant reduction in the error by approximately an order of magnitude within only 10 iterations and the magnitude of the phase coefficient that can be determined has been approximately doubled to one radian.

If this were to be used with noisy experimental data, the data may need to be band pass filtered so that iteration of the phase can potentially produce a zero-error phase estimate consistent with the data.

2.7. Further Development of Phase Contrast Imaging

It has been shown in the previous sections that increasing the number of images has reduced the error between the calculated phase and the set phase. As a generalisation of the method described in section 2.3.3, the number of images used to calculate the phase can be increased by considering the derivative of the intensity with respect to the phase change in a similar way to, Allen, *et al.* [20]. Allen, *et al.* [20] considered the derivative of the complex amplitude instead of the intensity that is considered here and the analysis did not apply to phase contrast imaging.

The phase shifted complex amplitude was defined in equation (2.8). The intensity is the magnitude squared of the complex amplitude, which can also be written as the complex amplitude multiplied by its complex conjugate so that

$$I(\mathbf{x},\gamma)=\iint \Psi(\xi)\Psi^{*}(\xi')g(\xi,\gamma)g^{*}(\xi',\gamma)\exp\left(-i2\pi\mathbf{x}\cdot(\xi-\xi')\right)d\xi d\xi' \qquad (2.41)$$

where $*$ denotes complex conjugation and ξ' is a position vector in Fourier space. This defines the intensity with the application of an arbitrary phase shift during the imaging process.

The derivative of $I(\mathbf{x},\gamma)$ with respect to γ gives an expression for the change in intensity with respect to the applied phase coefficient, so that

$$\frac{\partial I(\mathbf{x},\gamma)}{\partial\gamma} = \iint \Psi(\xi)\Psi^*(\xi')\exp\left(-i2\pi\mathbf{x}.(\xi-\xi')\right) \times\left\{\frac{\partial g(\xi,\gamma)}{\partial\gamma} g^*(\xi',\gamma) + g(\xi,\gamma)\frac{\partial g^*(\xi',\gamma)}{\partial\gamma}\right\} d\xi d\xi'. \tag{2.42}$$

From equation (2.7) the derivative of the phase shifting function is defined as

$$\frac{\partial g(\xi,\gamma)}{\partial\gamma} = iBS(\xi)\exp(i\gamma). \tag{2.43}$$

So the derivative of the intensity can be written as

$$\frac{\partial I(\mathbf{x},\gamma)}{\partial\gamma} = \iint \Psi(\xi)\Psi^*(\xi')\exp\left(-i2\pi x.(\xi-\xi')\right) \times\left\{iBS(\xi)\exp(i\gamma)g^*(\xi',\gamma) - g(\xi,\gamma)iBS(\xi')\exp(-i\gamma)\right\} d\xi d\xi'. \tag{2.44}$$

Expanding out the integrals gives

$$\frac{\partial I(\mathbf{x},\gamma)}{\partial\gamma} = iB\exp(i\gamma)\int\Psi(\xi)S(\xi)\exp(-i2\pi\mathbf{x}.\xi)d\xi\int\Psi^*(\xi')g^*(\xi',\gamma)\exp(i2\pi\mathbf{x}.\xi')d\xi' -iB\exp(-i\gamma)\int\Psi^*(\xi')S(\xi')\exp(i2\pi\mathbf{x}.\xi')d\xi'\int\Psi(\xi)g(\xi,\gamma)\exp(-i2\pi\mathbf{x}.\xi)d\xi. \tag{2.45}$$

In the same way as before the reference wave $R(\mathbf{x})$ can be defined assuming that $\phi(\mathbf{x}) \ll 1$ and that the phase information is confined to the area outside $S(\xi)$ and inside $T(\xi)$ so that the first integral on each line can be approximated by the expression for the reference wave, (2.13). Making this approximation and substituting for $g(\xi,\gamma)$, the derivative of the intensity can be written as

$$\frac{\partial I(\mathbf{x},\gamma)}{\partial\gamma} \cong iB\exp(i\gamma)R(\mathbf{x})\int\Psi^*(\xi')T(\xi')\left(A+S(\xi')\left(B\exp(-i\gamma)-A\right)\right)\exp(i2\pi\mathbf{x}.\xi')d\xi' -iB\exp(-i\gamma)R^*(\mathbf{x})\int\Psi(\xi)T(\xi)\left(A+S(\xi)\left(B\exp(i\gamma)-A\right)\right)\exp(-i2\pi\mathbf{x}.\xi)d\xi. \tag{2.46}$$

Expanding the expressions in the integrals and evaluating the terms gives

$$\begin{aligned}\frac{\partial I(\mathbf{x},\gamma)}{\partial\gamma} \approx\ & iAB\exp(i\gamma)R(\mathbf{x})\int\Psi^*(\xi')T(\xi')\exp(i2\pi\mathbf{x}.\xi')d\xi' \\ & -iAB\exp(-i\gamma)R^*(\mathbf{x})\int\Psi(\xi)T(\xi)\exp(-i2\pi\mathbf{x}.\xi)d\xi \\ & +iB(B-A\exp(i\gamma))|R(\mathbf{x})|^2 \\ & -iB(B-A\exp(-i\gamma))|R(\mathbf{x})|^2,\end{aligned} \tag{2.47}$$

which simplifies to

$$\begin{aligned}\frac{\partial I(\mathbf{x},\gamma)}{\partial\gamma} \approx\ & iAB\exp(i\gamma)R(\mathbf{x})\int\Psi^*(\xi')T(\xi')\exp(i2\pi\mathbf{x}.\xi')d\xi' \\ & -iAB\exp(-i\gamma)R^*(\mathbf{x})\int\Psi(\xi)T(\xi)\exp(-i2\pi\mathbf{x}.\xi)d\xi \\ & +iBA|R(\mathbf{x})|^2(\exp(i\gamma)-\exp(-i\gamma)).\end{aligned} \tag{2.48}$$

As before, the scattered wave is assuming to pass only inside $T(\xi)$ or equivalently the information outside $T(\xi)$ is negligible, so that the remaining integrals can be evaluated to give the unmodified complex amplitude, which can then be written in terms of intensity and phase. Therefore, the derivative can be written as

$$\begin{aligned}\frac{\partial I(\mathbf{x},\gamma)}{\partial\gamma} \approx\ & iABP(\mathbf{x})\sqrt{I(\mathbf{x})}\begin{pmatrix}+\Re\{R(\mathbf{x})\}\left(\exp\left(i\left(\gamma+\phi(\mathbf{x})\right)\right)-\exp\left(-i\left(\gamma+\phi(\mathbf{x})\right)\right)\right) \\ +i\Im\{R(\mathbf{x})\}\left(\exp\left(i\left(\gamma+\phi(\mathbf{x})\right)\right)+\exp\left(-i\left(\gamma+\phi(\mathbf{x})\right)\right)\right)\end{pmatrix} \\ & +2AB|R(\mathbf{x})|^2\sin\gamma.\end{aligned} \tag{2.49}$$

Converting the exponentials to trigonometric functions, then

$$\begin{aligned}\frac{\partial I(\mathbf{x},\gamma)}{\partial\gamma} \approx\ & -2ABP(\mathbf{x})\sqrt{I(\mathbf{x})}\left(\Re\{R(\mathbf{x})\}\sin\left(\gamma+\phi(\mathbf{x})\right)+\Im\{R(\mathbf{x})\}\cos\left(\gamma+\phi(\mathbf{x})\right)\right) \\ & +2AB|R(\mathbf{x})|^2\sin\gamma.\end{aligned} \tag{2.50}$$

The derivative needs to be estimated using a finite difference approximation, as it cannot be measured directly. The first order approximation is the difference between two images each with an equal and opposite phase shift. Therefore, a higher order approximation of the derivative gives a method for including more images.

To calculate the phase a small angle approximation could be applied to the phase, $\phi(\mathbf{x})$ then the equation can be solved to give the phase. Alternatively, the sum of the images could be used to give a second expression for the sine and cosine of the phase then,

following a similar process as section 2.3.3, the two equations can be solved simultaneously to give the phase.

Although not developed any further, with the significant improvement in the phase error calculated using three images, this gives a potential method for future development of this process, to improve the calculated phase, using an arbitrary number of images.

2.8. Conclusions

This chapter developed a method for calculating quantitative phase information from multiple phase contrast measurements. To derive an equation relating the phase and the intensity, a small angle approximation was introduced that restricts the application of this method to weak phase objects. The following chapter describes a method for determining stronger phase objects.

An experimental method, using three images, was introduced. In comparison to the method using a single phase contrast image there is a considerable improvement using three images, the examples show the phase error is significantly reduced by a factor of between 2 and 20. This is achieved in an idealised noise free situation so this will be an upper limit on the improvement. With experimental data and the presence of noise, the conditioning of the data will become an issue and this improvement may be reduced.

An additional benefit of the new method is the independence of the calculated phase from the applied phase shift. This allows the phase shift to be varied to increase the contrast of the images without affecting the accuracy of the calculated phase.

Iteration was also considered for refinement of the phase by calculating a phase correction from the difference between the measured and estimated data. Provided the phase coefficient was less than one radian, in the tested examples, there was a reduction in the error by approximately one order of magnitude in only a few iterations. This considerably extends the range.

It should be noted that convergence of this scheme has not been proven and so it is not guaranteed that this scheme will converge even for small set phase coefficients.

As was discussed a further development from three images could use more measurements, within the framework described in the final section, to potentially provide a further reduction in the phase error.

Chapter 3 – Theory and Implementation of Defocus Phase Diversity

3.1. Introduction

Phase diversity is a method for calculating phase from multiple intensity measurements where a known phase change is applied to each measurement, as discussed in Chapter 1.

In Chapter 2, phase contrast images were used to produce a set of phase diverse data, which can be analysed to determine the phase. Phase contrast imaging is a simple system that produces an image which relates directly to the phase. The calculation of the phase is also simple to understand from the point of view of two interfering waves.

In this chapter, defocused images are used as the input data, from which the phase will be calculated. When imaging in this way, the direct visualisation of the object is lost and the mathematical analysis is more complicated but it builds on the principles of the previous chapter and is a method that is not restricted to small phase angles.

In the previous chapter, phase contrast imaging was developed using three images to determine the phase. Compared to using a single image there was a significant reduction in the phase error but the equations were still limited due a small angle approximation in the analysis.

The principle used in this chapter is to make the phase visible in the intensity by defocusing the images. As the light propagates the intensity changes, which can be related to the phase. This was introduced in Chapter 1. This approach removes the small angle approximation requirement of the phase but replaces it with different approximations, requiring a small scattering angle. This method will be used in Chapter 6 in a process to determine the structure of a 3D phase object.

3.1.1. Aims and Objectives

This chapter aims to improve upon existing methods for determining the phase from defocused images using the intensity transport equation for a more accurate estimation of the phase.

As mentioned in Chapter 1 there are a number of ways of calculating the phase from a set of defocused images. This chapter considers the uses of the intensity transport equation (ITE),

$$-\frac{k}{I_0}\frac{\partial I(\mathbf{x},z)}{\partial z} \approx P(\mathbf{x})\nabla^2\phi(\mathbf{x},z) - \delta(\mathbf{x}-\mathbf{r}_0)\hat{\mathbf{n}}\cdot\nabla\phi(\mathbf{x},z), \tag{3.1}$$

to calculate the phase, $\phi(\mathbf{x})$, from the measured intensity. Where the intensity I_0 is a constant over the aperture, with unspecified shape, $P(\mathbf{x})$, $\hat{\mathbf{n}}$ is an outward pointing unit vector, normal to the perimeter, C, of the aperture, δ denotes a Dirac delta function at position $\mathbf{r}_0$ on the perimeter of the aperture which has an area A. This will be described in detail in section 3.2.1.

In the literature, there are two methods suggested that can be used to exactly solve this equation. The first method uses a Green's function and the second method uses Fourier transforms. As will be explained, neither of these methods are complete. Therefore, full solutions to the ITE will be derived in this chapter and the numerical implementation with experimental data.

It has been shown, in general, that the ITE can be solved exactly using a Green's function solution, Woods, *et al.* [30], and that the phase, $\phi(\mathbf{x})$ can be calculated from

$$\phi(\mathbf{x}) = \int G(\mathbf{x},\mathbf{x}')\,Data(\mathbf{x}')\,d^2\mathbf{x}', \tag{3.2}$$

where the Green's function is represented by $G(\mathbf{x},\mathbf{x}')$ and the input data by $Data(\mathbf{x}')$.

In the published literature there are three different Green's functions stated without derivation that claim to be the correct Green's function to solve the ITE for the specific case where $P(\mathbf{x})$ represents a disk of radius r_0. The first stated by Stakgold [46] is

$$\begin{aligned} G(r,\theta,R,\Theta) = {} & \frac{1}{4\pi}\log\frac{\left(r^2+R^2-2rR\cos(\theta-\Theta)\right)\left(r_0^4+r^2R^2-2rRr_0^2\cos(\theta-\Theta)\right)}{r_0^4} \\ & -\frac{r^2}{4\pi r_0^2} + a_0(R,\Theta). \end{aligned} \tag{3.3}$$

the second stated Trim [47] is

$$G(r,\theta,R,\Theta)=\frac{1}{4\pi}\log\frac{\left(r^2+R^2-2rR\cos(\theta-\Theta)\right)\left(r_0^4+r^2R^2-2rRr_0^2\cos(\theta-\Theta)\right)}{r_0^4R^2}$$
$$-\frac{r^2}{4\pi r_0^2}+\frac{F(R,\Theta)}{2\pi}. \tag{3.4}$$

The final Green's function according to Zhu, *et al.* [48] is

$$G(r,\theta,R,\Theta)=-\frac{1}{4\pi}\log\frac{\left(r^2+R^2-2rR\cos(\theta-\Theta)\right)\left(r_0^4+r^2R^2-2rRr_0^2\cos(\theta-\Theta)\right)}{r_0^4}, \tag{3.5}$$

where (r,θ) and (R,Θ) are coordinates within the aperture, and $F(R,\Theta)$ and $a_0(R,\Theta)$ are arbitrary functions. The Green's functions are all slightly different and cannot all be correct. It will be shown that the first two Green's functions calculate equivalent phases and that the third is wrong, as it fails to satisfy the necessary conditions to be a valid Green's function. This chapter derives the first two Green's functions in full to show that there are no additional unknown assumptions or approximations and shows that both satisfy the mathematical conditions to be a valid solution. Therefore, either of the first two Green's functions, equations (3.3) or (3.4), can be used to solve the ITE in this specific case.

The two equivalent Green's functions are stated in mathematical textbooks and are largely unknown to the optics community. In addition, they are written in a way that does not translate directly to the application of calculating phase from sampled data. Therefore, the process for calculating the phase from the sampled data is described.

It can be seen from examination of the Green’s function that there are singularities. To use the Green's function numerically these points can be removed by setting them to zero or some other arbitrary value, this is the approach taken by Teague [49]. This is unsatisfactory as it is an arbitrary process and potentially introduces systematic errors into the calculation therefore, a process will be described that gives well defined values at each point.

Even with these improvements to the process of calculating the phase, using a Green's function, there is still a fundamental limitation placed on the phase that requires it to be continuous. However, for many cases, such as a cell suspended in a fluid, there is a discontinuity in the phase at the boundary of the cell due to the change in density.

To remove this condition a second solution to the ITE will be described, that uses Fourier transforms. This method has been partly implemented by Dorrer, *et al.* [27] and Allen, *et al.* [28] but the boundary conditions were not included in their solutions. Therefore, the calculated phase will potentially miss the contribution from the entire set of harmonic functions. Harmonic functions produce no data inside the area of the aperture and therefore they can only be calculated from the boundary data.

The contribution from the boundary data has been included iteratively by Diaz-Douton, *et al.* [24] but there has been no complete analytic solution to the ITE using Fourier transforms that include the boundary conditions.

The full solution to the ITE is therefore derived here with the inclusion of the boundary data. The implementation of this process with sampled data is also described, as the boundary data is defined on a circle and experimental data is sampled on a square grid a method is required to extract the correct data for input into the analysis.

The specific objectives of this chapter are to determine which Green's functions are correct. Derive a full solution to the ITE using Fourier transforms. The process for numerically implementing these solutions will be described along with a method for removing the singularities. These processes solve the ITE exactly giving an improved estimation of the phase.

3.2. Theory

This chapter considers using defocused images to calculate the phase. This section starts by describing the theory of applying an arbitrary known phase modification during the imaging process. The specific example of a defocus phase change will be considered at a later stage.

3.2.1. Derivation of the Intensity Transport Equation

This section introduces the theory that describes applying an arbitrary known phase change to a beam of light to modify the measured intensity.

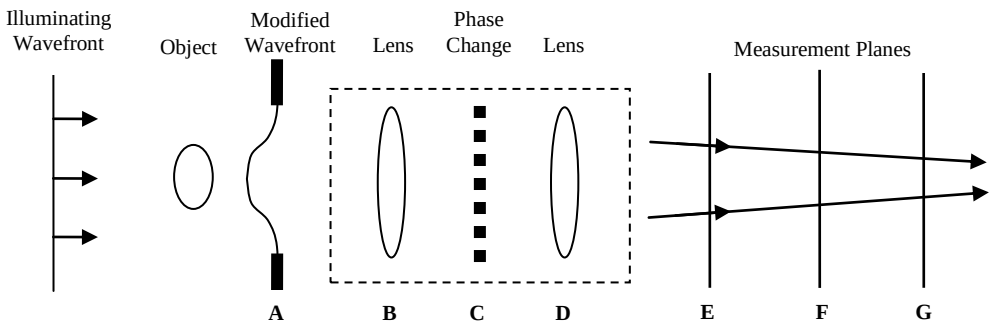


Figure 3.1 Schematic of the optical system for recording defocused images

In a same way as the previous chapter, a modified wavefront in plane A of Figure 3.1 is focused by the lens in plane B onto plane C. In plane C a known phase change is applied to the wave before it propagates through the second lens and so focused on the camera in plane F.

In general, the complex amplitude of a wave in plane F of Figure 3.1 with an arbitrary phase change applied to it, can be written as, Allen, *et al.* [20]

$$\psi(\mathbf{x},\tau)=\int\Psi(\xi)g(\xi,\tau)\exp(-i2\pi\xi\cdot\mathbf{x})d\xi \tag{3.6}$$

where $\Psi(\xi)$ is the Fourier transform of the complex amplitude of the wave, $\psi(\mathbf{x})$, in plane A of Figure 3.1. The position vector in real space is represented by $\mathbf{x}$ and ξ represents the position vector in Fourier space. The Fourier transform represents the action of the first lens in plane B of Figure 3.1. The known phase modification, in plane C, will be modelled by $g(\xi,\tau)=\exp(i\tau\gamma(\xi))$, where the phase change is denoted by $\gamma(\xi)$ with a coefficient τ. The definition of $g(\xi,\tau)$ can be changed depending on the applied phase modification, for example the definition was different when phase contrast images were considered in Chapter 2. The modified complex amplitude in plane F is represented by $\psi(\mathbf{x},\tau)$.

The intensity in plane F of Figure 3.1 is therefore

$$I(\mathbf{x},\tau)=\iint\Psi(\xi)\Psi^*(\xi')g(\xi,\tau)g^*(\xi',\tau)\exp(-i2\pi\mathbf{x}\cdot(\xi-\xi'))d\xi d\xi' . \tag{3.7}$$

Following the process in section 2.7 but generalising it for an arbitrary phase modification, the derivative of $I(\mathbf{x},\tau)$ with respect to τ gives an expression for the change in intensity with respect to the phase coefficient,

$$\frac{\partial I(\mathbf{x},\tau)}{\partial \tau} = \iint \Psi(\xi)\Psi^*(\xi')\exp\left(-i2\pi\mathbf{x}\cdot(\xi-\xi')\right) \times\left\{\frac{\partial g(\xi,\tau)}{\partial \tau} g^*(\xi',\tau) + g(\xi,\tau)\frac{\partial g^*(\xi',\tau)}{\partial \tau}\right\} d\xi d\xi'. \tag{3.8}$$

Substituting for the derivative of $g(\xi,\tau)$, it follows directly that

$$\frac{\partial I(\mathbf{x},\tau)}{\partial \tau} = \iint \Psi(\xi)\Psi^*(\xi')\exp\left(-i2\pi\mathbf{x}\cdot(\xi-\xi')\right) \times\left\{i\gamma(\xi)g(\xi,\tau)g^*(\xi',\tau) - i\gamma(\xi')g(\xi,\tau)g^*(\xi',\tau)\right\} d\xi d\xi'. \tag{3.9}$$

This can be expanded and the integrals evaluated to give

$$\frac{\partial I(\mathbf{x},\tau)}{\partial \tau} = i\psi^*(\mathbf{x},\tau)\int \Psi(\xi)\exp\left(-i2\pi\mathbf{x}\cdot\xi\right)g(\xi,\tau)\gamma(\xi)d\xi - i\psi(\mathbf{x},\tau)\int \Psi^*(\xi')\exp\left(i2\pi\mathbf{x}\cdot\xi'\right)g^*(\xi',\tau)\gamma(\xi')d\xi' . \tag{3.10}$$

The right side is i times the difference of complex conjugates so is real valued, therefore the derivative is real and it can be applied to real data. By writing $\Psi(\xi)$ and $\psi(\mathbf{x},\tau)$ in terms of $\psi(\mathbf{x})$, which can be replaced with the intensity and phase, then this equation relates the change in intensity with respect to the phase modification to the phase and other known or measurable quantities.

This chapter looks at the specific phase modification of defocus. The application of a defocus phase change in plane C of Figure 3.1 is equivalent to moving the camera and making a measurement in, for example, plane E or G of Figure 3.1.

Throughout this thesis the propagation of the complex amplitude between the planes perpendicular to the propagation direction is modelled using the angular spectrum of plane waves, Goodman [37]. The propagation is modelled by taking a Fourier transform of the complex amplitude in some plane, which can be considered as a set of plane waves propagating in different directions. Each plane wave travels in a different direction so there will be a relative phase change to each plane wave relating to the propagation distance. The phase change due to propagation is equivalent to a defocus phase change, which can be written as

$$\gamma(\xi) = k\sqrt{1-|\xi|^2\lambda^2}. \tag{3.11}$$

The application of the phase change, with coefficient $\tau = z$, is modelled as

$$g(\xi,z) = \exp\left(ikz\sqrt{1-|\xi|^2\lambda^2}\right). \tag{3.12}$$

The plane waves are then recombined using an inverse Fourier transform to give the propagated complex amplitude.

Therefore, substituting (3.11) into (3.10) gives

$$\begin{aligned}\frac{\partial I(\mathbf{x},z)}{\partial z} &= i\psi^*(\mathbf{x},z)\int \Psi(\xi)\, g(\xi,z)\, k\sqrt{1-|\xi|^2\lambda^2}\exp(-i2\pi\mathbf{x}\cdot\xi)\,d\xi \\ &\quad - i\psi(\mathbf{x},z)\int \Psi^*(\xi')\, g^*(\xi',z)\, k\sqrt{1-|\xi'|^2\lambda^2}\exp(i2\pi\mathbf{x}\cdot\xi')\,d\xi' .\end{aligned} \tag{3.13}$$

When $1<\lambda^2|\xi|^2$ the waves are evanescent then $g(\xi,z)$ represents an exponential decay and the derivative is complex, Goodman [37]. When this happens, the equation is no longer valid, as the derivative is required to be real so it can be applied to real data.

It is beyond the scope of this thesis to determine if this equation can be solved analytically, in a form that is useful. Therefore, to simplify this equation the square root term can be written as

$$\sqrt{1-\lambda^2|\xi|^2} = 1-\frac{\lambda^2|\xi|^2}{2}-\sum_{n=2}^{\infty}\frac{(-1)^n(2n)!\lambda^{2n}|\xi|^{2n}}{(2n-1)(n!)^2 4^n} \tag{3.14}$$

and can be approximated by the first two terms in the series provided $\left||\xi|^2\lambda^2\right| \ll 1$. Under this condition, the equation is only valid for small scattering angles where there are no evanescent waves.

The approximation is not applied to $g(\xi,z)$, only to the square root explicitly written in (3.13). Therefore, approximating the square root by the first two terms in the series, the derivative of the intensity can be approximated as

$$\begin{aligned}-i\frac{\partial I(\mathbf{x},z)}{\partial z} &\approx -k\frac{\lambda^2}{2}\psi^*(\mathbf{x},z)\int |\xi|^2\,\Psi(\xi)\, g(\xi,z)\exp(-i2\pi\mathbf{x}\cdot\xi)\,d\xi \\ &\quad + k\frac{\lambda^2}{2}\psi(\mathbf{x},z)\int |\xi'|^2\,\Psi^*(\xi')\, g^*(\xi',z)\exp(i2\pi\mathbf{x}\cdot\xi')\,d\xi' .\end{aligned} \tag{3.15}$$

A second order approximation is considered in Appendix A where it is shown that the contribution from higher order terms becomes increasingly small.

Consider a differentiable function $f(x)$ that has a Fourier transform $F(\xi)$. Then from the derivative property of the Fourier transform, Bracewell [50], the derivative of $f(x)$ can written as

$$\frac{d^n f(x)}{dx^n} = \int (-i2\pi\xi)^n F(\xi) \exp(-i2\pi x\xi) d\xi \tag{3.16}$$

Applying this definition to (3.15) it follows, noting $\Psi(\xi, z) = \Psi(\xi) g(\xi, z)$, that

$$i\frac{\partial I(\mathbf{x}, z)}{\partial z} \approx \frac{1}{2k}\left(\psi(\mathbf{x}, z)\nabla^2\psi^*(\mathbf{x}, z) - \psi^*(\mathbf{x}, z)\nabla^2\psi(\mathbf{x}, z)\right). \tag{3.17}$$

Making the substitution $\psi(\mathbf{x}, z) = \sqrt{I(\mathbf{x}, z)}\exp(-i\phi(\mathbf{x}, z))$ and evaluating the terms on the right separately then

$$\nabla\cdot\psi(\mathbf{x}, z) = \frac{\exp(-i\phi(\mathbf{x}, z))}{2\sqrt{I(\mathbf{x}, z)}}\nabla\cdot I(\mathbf{x}, z) - i\sqrt{I(\mathbf{x}, z)}\exp(-i\phi(\mathbf{x}, z))\nabla\cdot\phi(\mathbf{x}, z). \tag{3.18}$$

Therefore,

$$\begin{aligned}\nabla^2\psi(\mathbf{x}, z) = &-\frac{\exp(-i\phi(\mathbf{x}, z))}{4I(\mathbf{x}, z)^{3/2}}\left(\nabla\cdot I(\mathbf{x}, z)\right)^2 - i\frac{\exp(-i\phi(\mathbf{x}, z))}{\sqrt{I(\mathbf{x}, z)}}\nabla\cdot\phi(\mathbf{x}, z)\nabla\cdot I(\mathbf{x}, z)\\ &+\frac{\exp(-i\phi(\mathbf{x}, z))}{2\sqrt{I(\mathbf{x}, z)}}\nabla^2 I(\mathbf{x}, z) - i\sqrt{I(\mathbf{x}, z)}\exp(-i\phi(\mathbf{x}, z))\nabla^2\phi(\mathbf{x}, z)\\ &-\exp(-i\phi(\mathbf{x}, z))\sqrt{I(\mathbf{x}, z)}\left(\nabla\cdot\phi(\mathbf{x}, z)\right)^2.\end{aligned} \tag{3.19}$$

Upon multiplication of this expression by the complex conjugate of the complex amplitude, $\psi(\mathbf{x}, z)$, the exponential terms will all cancel leaving

$$\begin{aligned}\psi^*(\mathbf{x}, z)\nabla^2\psi(\mathbf{x}, z) = &-\frac{\left(\nabla\cdot I(\mathbf{x}, z)\right)^2}{4I(\mathbf{x}, z)} - i\nabla\cdot\phi(\mathbf{x}, z)\nabla\cdot I(\mathbf{x}, z)\\ &+\frac{\nabla^2 I(\mathbf{x}, z)}{2} - iI(\mathbf{x}, z)\nabla^2\phi(\mathbf{x}, z) - I(\mathbf{x}, z)\left(\nabla\cdot\phi(\mathbf{x}, z)\right)^2.\end{aligned} \tag{3.20}$$

A similar expression can be derived for the other term in (3.17). Substituting these expressions into (3.17), the real parts will cancel to give the intensity transport equation (ITE), Roddier [3],

$$-k\frac{\partial I(\mathbf{x}, z)}{\partial z} \approx I(\mathbf{x}, z)\nabla^2\phi(\mathbf{x}, z) + \nabla\phi(\mathbf{x}, z)\bullet\nabla I(\mathbf{x}, z). \tag{3.21}$$

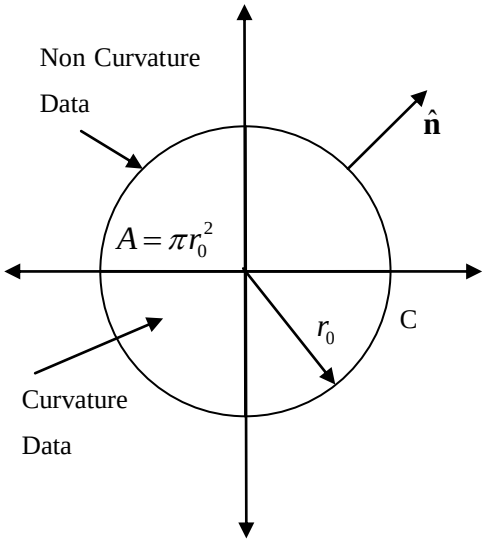


Figure 3.2 Schematic of the division of the data over the aperture according to the ITE

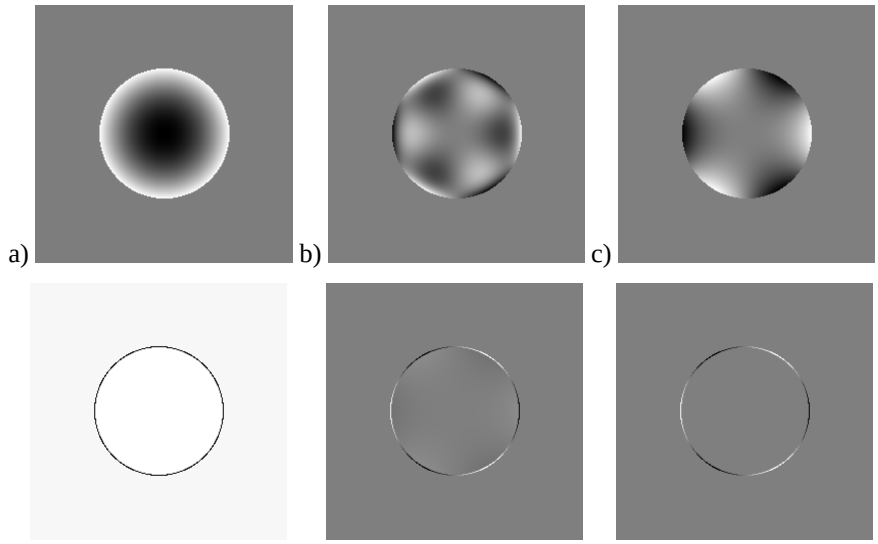


Figure 3.3 Examples of the way the phase is encoded in to the intensity derivative. The top line shows the phase a) shows defocus Z_2^0 b) shows Z_5^3 and c) shows Z_3^3. The bottom line shows the corresponding derivative calculated using the ITE.

For the special case of constant intensity $I(\mathbf{x}) = I_0 P(\mathbf{x})$ where I_0 is a constant and $P(\mathbf{x})$ is a binary function defining the aperture, with unspecified shape, then

$$-\frac{k}{I_0}\frac{\partial I(\mathbf{x},z)}{\partial z} \approx P(\mathbf{x})\nabla^2\phi(\mathbf{x},z) - \delta(\mathbf{x}-\mathbf{r}_0)\hat{\mathbf{n}}\cdot\nabla\phi(\mathbf{x},z), \tag{3.22}$$

where $\hat{\mathbf{n}}$ is an outward pointing unit vector, normal to the perimeter, δ denotes a delta function at position $\mathbf{r}_0$ on the perimeter C of the illuminated area A. The delta function on the boundary is a consequence of the discontinuity in $P(\mathbf{x})$ at the edge of the pupil. An example, for a circular aperture is shown in Figure 3.2.

The left side of equation (3.22) is composed of measured or known quantities that form the input data. The derivative will be estimated using a finite difference approximation. After defining the input data the phase can then be calculated by solving this equation.

The input data, (3.22), is considered to be composed of two parts, consisting of phase curvature and the derivative of the phase normal to the perimeter, as shown in Figure 3.2. The phase is considered to have curvature if the Laplacian, the vector operator ∇^2, of the phase inside aperture is non-zero. Phase distributions that have no curvature result in a zero Laplacian inside the disk and as a result, these functions can only be determined from the boundary data. The distinction between curvature and non-curvature is a result of the ITE and is independent of the method used to solve the equation.

As described in section 2.4.4 the set of Zernike polynomials will be used in this chapter to define the set phase. Within this set, there are functions that can be calculated without considering the boundary conditions, these are the cylindrically symmetric functions Z_m^0 that have no recoverable information encoded into the boundary. There is also a subset of functions that are harmonic or have zero curvature, $Z_{|m|}^m$, which according to the ITE, are calculated only from the boundary conditions, the rest of the functions in the set require the solution to the differential equation and consideration of the boundary conditions.

Examples of these three types of functions are shown in Figure 3.3. This figure shows the set phase on the top line and the corresponding approximation of the derivative of the intensity on the bottom line, calculated using the ITE, (3.22). For cylindrically symmetric functions, a), there is only recoverable information inside the disk. The constant value on the edge provided no additional information. For harmonic functions, c), there is no information inside the disk, it is all concentrated at the edge where the frequency of the oscillations around the edge is equal to the azimuthal frequency of the

set mode. For the remaining modes, full consideration of the differential equation and the boundary conditions are required as there is data inside the disk and on the boundary.

3.2.2. Green's Function Solution

In the previous section, the ITE was derived relating the intensity to the phase. By solving the equation, the phase can be expressed in terms of measurable and known parameters. As discussed in the introduction to this chapter, the ITE can be solved exactly using a Green's function. This section restates the work of Woods, *et al.* [30] to define all the conditions necessary to derive a Green's function. The derivation of the Green's function starts at the beginning so that the analysis is complete and consistent.

Green's second theorem states that

$$\int_A \phi(\mathbf{x}')\nabla^2 G(\mathbf{x},\mathbf{x}')d^2\mathbf{x}' = \int_A G(\mathbf{x},\mathbf{x}')\nabla^2\phi(\mathbf{x}')d^2\mathbf{x}' + \oint_C \phi(\mathbf{x}')\nabla G(\mathbf{x},\mathbf{x}')\cdot d\hat{\mathbf{n}}' - \oint_C G(\mathbf{x},\mathbf{x}')\nabla\phi(\mathbf{x}')\cdot d\hat{\mathbf{n}}'. \tag{3.23}$$

This requires that the phase, $\phi(\mathbf{x}')$, is continuous with bounded derivatives, in this case of order one, at every point in the area A, with a boundary C, Woods, *et al.* [30]. As the phase is restricted to continuous functions, finite discontinuities cannot be reconstructed.

The Green's function is defined so that

$$\nabla^2 G(\mathbf{x},\mathbf{x}') = \delta(\mathbf{x}-\mathbf{x}'), \tag{3.24}$$

where δ is a Dirac delta function. Using this definition the left side of equation (3.23) evaluates to leave the phase, $\phi(\mathbf{x})$. The boundary data in the ITE, (3.22), can be defined as

$$\hat{\mathbf{n}}\cdot\nabla\phi(\mathbf{x}) = h(\mathbf{x}) \quad \forall\mathbf{x}\in C \tag{3.25}$$

where $h(\mathbf{x})$ is a known measured value on the boundary.

The boundary conditions applied to the Green's function are chosen to simplify the problem and are dictated by the available phase information. Therefore, because the derivative of the phase is known and nothing is known about the phase, the first contour

integral in (3.23) can be set to zero by imposing boundary conditions on $G(\mathbf{x},\mathbf{x}')$, such that

$$\nabla G(\mathbf{x},\mathbf{x}')\cdot\hat{\mathbf{n}}=0 \quad \forall \mathbf{x}\in C\,. \tag{3.26}$$

The final condition on the Green's function is that it must satisfy the divergence theorem,

$$\int_A \nabla^2 G(\mathbf{x},\mathbf{x}')d^2\mathbf{x}' = \oint_C \nabla G(\mathbf{x},\mathbf{x}')\cdot d\hat{\mathbf{n}}'. \tag{3.27}$$

Satisfying this equation ensures that the Green's function is consistent with the two definitions, (3.24) and (3.26), and that a solution exists. Using the defined conditions for the Greens function the left side of (3.27) evaluates to 1 and the right to 0. To resolve this difference there is a choice between modifying the differential equation or the boundary conditions. This makes no significant difference therefore the differential equation defining the Green's function is modified and the conditions specifying the Green's function are

$$\begin{aligned}\nabla^2 G(\mathbf{x},\mathbf{x}') &= \delta(\mathbf{x}-\mathbf{x}')-\frac{1}{A}\\ \nabla G(\mathbf{r_0},\mathbf{x}')\cdot\hat{\mathbf{n}} &= 0.\end{aligned} \tag{3.28}$$

By changing the differential equation, all the necessary conditions are now satisfied and the phase can still be calculated from the ITE. The phase can now be written in terms of an area integral and a contour integral so that

$$\phi(\mathbf{x})-\frac{1}{A}\int_A \phi(\mathbf{x}')d^2\mathbf{x}' = \int_A G(\mathbf{x},\mathbf{x}')\nabla^2\phi(\mathbf{x}')d^2\mathbf{x}' - \oint_C G(\mathbf{x},\mathbf{x}')\nabla\phi(\mathbf{x}')\cdot d\hat{\mathbf{n}}'. \tag{3.29}$$

The modification introduces an extra integral, on the left, that subtracts the mean value of the phase. As only relative changes can be determined this constant can be ignored, Woods, *et al.* [30]. By changing the limits of the integral to cover the entire plane, and imposing the limits for the area and the boundary by specific additional functions, then

$$\phi(\mathbf{x}) = \int G(\mathbf{x},\mathbf{x}')\left(P(\mathbf{x}')\nabla^2\phi(\mathbf{x}')-\delta_C\nabla\phi(\mathbf{x}')\cdot\hat{\mathbf{n}}'\right)d^2\mathbf{x}'. \tag{3.30}$$

Substituting from equation (3.22), the phase can be written in terms of the derivative of the intensity so that

$$\phi(\mathbf{x}) = -\frac{k}{I_0}\int G(\mathbf{x},\mathbf{x}')\frac{\partial I(\mathbf{x}',z)}{\partial z}d^2\mathbf{x}'. \tag{3.31}$$

Therefore, using this equation the phase at a specific point, $\mathbf{x}$, can be calculated from the derivative of the intensity multiplied by a Green's function and integrated over the area of the aperture, Woods, *et al.* [30].

This section has shown that the ITE can be solved using a Green's function and in the process defined the conditions to derive a specific Green's function.

3.2.3. Derivation of the First Green's Function

Having shown that the ITE can be solved using a Green's function, this section derives the first Green's function stated in the literature that can be used to solve the ITE. As stated in the previous section the requirements for the Green's function are

$$\begin{aligned} &\nabla^2 G(\mathbf{x},\mathbf{x}') = \delta(\mathbf{x}-\mathbf{x}') - \frac{1}{\pi r_0^2} \\ &\nabla G(\mathbf{x},\mathbf{x}')\cdot\hat{\mathbf{n}} = 0 \quad \forall \mathbf{x} \in C. \end{aligned} \tag{3.32}$$

To derive a Green's function the area of the aperture and its boundary need to be defined, in this case $P(\mathbf{x})$ is defined as a disk with radius r_0 and area πr_0^2 as shown in Figure 3.2.

The Green's function is constructed in three parts, so that $G(\mathbf{x},\mathbf{x}') = U(\mathbf{x},\mathbf{x}') + g(\mathbf{x},\mathbf{x}') + \kappa(\mathbf{x},\mathbf{x}')$. $U(\mathbf{x},\mathbf{x}')$ solves the differential equation containing the delta function, $g(\mathbf{x},\mathbf{x}')$ enforces the boundary conditions and $\kappa(\mathbf{x},\mathbf{x}')$ ensures the Green's function satisfy's the divergence theorem.

The free space Green's function $U(\mathbf{x},\mathbf{x}')$ is the solution, without the application of boundary conditions, to

$$\nabla^2 U(\mathbf{x},\mathbf{x}') = \delta(\mathbf{x}-\mathbf{x}'). \tag{3.33}$$

The solution, in polar coordinates is defined, Trim [47], as

$$U(r,\theta,R,\Theta) = \frac{1}{4\pi}\log\left(r^2 + R^2 - 2rR\cos(\theta-\Theta)\right) \tag{3.34}$$

where $\mathbf{x} = (r,\theta)$ and $\mathbf{x}' = (R,\Theta)$. By substituting this solution into the differential equation it can be shown, Strichartz [51], that this is the correct solution where $U(\mathbf{x},\mathbf{x}')$ is considered as a distribution. Using the following identity

$$\sum_{n=1}^{\infty}\alpha^n\frac{\cos n\theta}{n}=-\frac{1}{2}\log\left(1+\alpha^2-2\alpha\cos\theta\right)\quad |\alpha|<1 \tag{3.35}$$

$U\left(r,\theta,R,\Theta\right)$ can be written as an infinite series, Trim [47], as

$$U\left(r,\theta,R,\Theta\right)=\frac{1}{4\pi}\log r^2-\frac{1}{2\pi}\sum_{n=1}^{\infty}\left(\frac{R}{r}\right)^n\frac{\cos\left(n\left(\theta-\Theta\right)\right)}{n}\quad R<r. \tag{3.36}$$

This expression is used later to apply the boundary condition.

The contribution from $\kappa\left(r,\theta,R,\Theta\right)$ ensures the Green's function satisfies the divergence theorem. $\kappa\left(r,\theta,R,\Theta\right)$ without loss of generality is assumed to be cylindrically symmetric therefore writing the Laplacian in polar coordinates

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\kappa\left(r,\theta,R,\Theta\right)}{\partial r}\right)=-\frac{1}{\pi r_0^2}. \tag{3.37}$$

By direct integration the solution to this equation is

$$\kappa\left(r,\theta,R,\Theta\right)=-\frac{r^2}{4\pi r_0^2}+\varepsilon\left(R,\Theta\right)\log r+\beta\left(R,\Theta\right). \tag{3.38}$$

So that $\kappa\left(r,\theta,R,\Theta\right)$ is bound at $r=0$, $\varepsilon\left(R,\Theta\right)$ is set to zero. Now to impose the boundary conditions $g\left(\mathbf{x},\mathbf{x}'\right)$ is required to satisfy

$$\nabla^2 g(\mathbf{x},\mathbf{x}')=0 \tag{3.39}$$

while enforcing the boundary conditions in (3.32). The set of solutions that satisfy this first requirement is the set of harmonic functions, a complete set of orthogonal functions, defined as

$$g(r,\theta,R,\Theta)=a_0\left(R,\Theta\right)+\sum_{n=1}^{\infty}r^n\left(a_n\left(R,\Theta\right)\cos\left(n\theta\right)+b_n\left(R,\Theta\right)\sin\left(n\theta\right)\right). \tag{3.40}$$

The remaining functions, $a_n\left(R,\Theta\right)$ and $b_n\left(R,\Theta\right)$ are specified by applying the boundary conditions. From (3.32) as the aperture is a disk then $\nabla G\cdot\hat{\mathbf{n}}=\partial G/\partial r$, so that

$$\frac{\partial G\left(r_0,\theta,R,\Theta\right)}{\partial r}=\frac{\partial U\left(r_0,\theta,R,\Theta\right)}{\partial r}+\frac{\partial g\left(r_0,\theta,R,\Theta\right)}{\partial r}+\frac{\partial\kappa\left(r_0,\theta,R,\Theta\right)}{\partial r}=0. \tag{3.41}$$

Making substitutions for $U\left(r,\theta,R,\Theta\right)$, $g\left(r,\theta,R,\Theta\right)$ and $\kappa\left(r,\theta,R,\Theta\right)$ and evaluating the equation at $r=r_0$ then

$$\sum_{n=1}^{\infty}nr_0^{n-1}\left(a_n\left(R,\theta\right)\cos n\theta+b_n\left(R,\theta\right)\sin n\theta\right)=-\frac{1}{2\pi}\sum_{n=1}^{\infty}\frac{R^n}{r_0^{n+1}}\cos\left(n\left(\theta-\Theta\right)\right). \tag{3.42}$$

By expanding the cosine term, on the right, using $\cos(\alpha-\beta)=\cos(\alpha)\cos(\beta)+\sin(\alpha)\sin(\beta)$ and equating the coefficients of the trigonometric functions, the functions $a_n(R,\Theta)$ and $b_n(R,\Theta)$ can be defined as

$$\begin{aligned} a_n(R,\Theta) &= -\frac{R^n}{2\pi n r_0^{2n}}\cos n\Theta \\ b_n(R,\Theta) &= -\frac{R^n}{2\pi n r_0^{2n}}\sin n\Theta. \end{aligned} \tag{3.43}$$

The function $a_0(R,\Theta)$ is unspecified due to the type of boundary conditions. It will be shown in the following section that $a_0(R,\Theta)$ only adds a constant to the phase so can be set to zero.

By combining the three parts of the Green's function it can written as

$$\begin{aligned} G(r,\theta,R,\Theta) &= \frac{1}{4\pi}\log\left(r^2+R^2-2rR\cos(\theta-\Theta)\right)-\frac{r^2}{4\pi r_0^2} \\ &\quad +a_0(R,\Theta)+\sum_{n=1}^{\infty} r^n\left(-\frac{R^n}{2\pi n r_0^{2n}}\cos n\Theta\cos n\theta-\frac{R^n}{2\pi n r_0^{2n}}\sin n\Theta\sin n\theta\right). \end{aligned} \tag{3.44}$$

Rewriting the summation using $\cos(n\theta)\cos(n\Theta)+\sin(n\theta)\sin(n\Theta)=\cos\left(n(\theta-\Theta)\right)$ and applying identity (3.35), the Green's function simplifies to

$$\begin{aligned} G(r,\theta,R,\Theta) &= \frac{1}{4\pi}\log\frac{\left(r^2+R^2-2rR\cos(\theta-\Theta)\right)\left(r_0^4+r^2R^2-2rRr_0^2\cos(\theta-\Theta)\right)}{r_0^4} \\ &\quad -\frac{r^2}{4\pi r_0^2}+a_0(R,\Theta). \end{aligned} \tag{3.45}$$

This matches the Green's function stated by Stakgold [46]. The validity of this solution and the term $a_0(R,\Theta)$ are discussed in the following sections.

3.2.4. Derivation of the Second Green's Function

As mentioned in the introduction, there is more than one Green’s function stated in the literature. This section derives a second Green's function and shows that it and the Green's function from the previous section can be used to calculate equivalent phases.

As an alternative to the previous section where the Green's function was split into parts, the differential equation can be solved directly to give the Green's function. This is

achieved by decomposing the Green's function into a Fourier series with respect to θ as the Green's function is 2π periodic and the Green's function and its derivative are continuous so $G(r,\pi,R,\Theta)=G(r,-\pi,R,\Theta)$ and $\partial G(r,\pi,R,\Theta)/\partial\theta=\partial G(r,-\pi,R,\Theta)/\partial\theta$, Stakgold [46]. The conditions for the Green's function, in polar coordinates, are defined as

$$\begin{aligned}&\nabla^2 G(r,\theta,R,\Theta)=\frac{\delta(r-R)\delta(\theta-\Theta)}{r}-\frac{1}{\pi r_0^2}\\&\frac{\partial G(r_0,\theta,R,\Theta)}{\partial r}=0.\end{aligned}\tag{3.46}$$

The differential equation can be written in polar coordinates as

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G(r,\theta,R,\Theta)}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2 G(r,\theta,R,\Theta)}{\partial\theta^2}=\frac{\delta(r-R)\delta(\theta-\Theta)}{r}-\frac{1}{\pi r_0^2}.\tag{3.47}$$

Multiplying this equation by $\exp(-in\theta)/2\pi$ and integrating it with respect to θ then

$$\begin{aligned}&\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)\int_{-\pi}^{\pi}G(r,\theta,R,\Theta)\frac{\exp(-in\theta)}{2\pi}\partial\theta+\int_{-\pi}^{\pi}\frac{1}{r^2}\frac{\partial^2 G(r,\theta,R,\Theta)}{\partial\theta^2}\frac{\exp(-in\theta)}{2\pi}\partial\theta\\&=\frac{\delta(r-R)}{r}\frac{\exp(-in\Theta)}{2\pi}-\frac{\sin(n\pi)}{\pi^2 r_0^2 n}.\end{aligned}\tag{3.48}$$

The second term on the left can be evaluated using integration by parts to give

$$\begin{aligned}&\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)\int_{-\pi}^{\pi}G(r,\theta,R,\Theta)\frac{\exp(-in\theta)}{2\pi}\partial\theta-\frac{n^2}{r^2}\int_{-\pi}^{\pi}G(r,\theta,R,\Theta)\frac{\exp(-in\theta)}{2\pi}\partial\theta\\&=\frac{\delta(r-R)}{r}\frac{\exp(-in\Theta)}{2\pi}-\frac{\sin(n\pi)}{\pi^2 r_0^2 n}.\end{aligned}\tag{3.49}$$

The integrals now represent the coefficients of a Fourier series such that

$$g_n(r,R,\Theta)=\int_{-\pi}^{\pi}G(r,\theta,R,\Theta)\frac{\exp(-in\theta)}{2\pi}\partial\theta.\tag{3.50}$$

Therefore, solving the differential equations

$$\left.\begin{aligned}&\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial g_n(r,R,\Theta)}{\partial r}\right)-\frac{n^2}{r^2}g_n(r,R,\Theta)=\frac{\exp(-in\Theta)\delta(r-R)}{2\pi r}\\&\left.\frac{\partial g_n}{\partial r}\right|_{r_0}=0\end{aligned}\right\}\quad n\neq 0$$

$$\left.\begin{aligned}&\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial g_0(r,R,\Theta)}{\partial r}\right)=\frac{\delta(r-R)}{2\pi r}-\frac{1}{\pi r_0^2}\\&\left.\frac{\partial g_0}{\partial r}\right|_{r_0}=0\end{aligned}\right\}\quad n=0\tag{3.51}$$

for each value of n gives a set of function which when combined with the corresponding complex exponentials fully defines the Green's function, so that

$$G(r,\theta,R,\Theta)=\sum_{n=-\infty}^{\infty} g_n(r,R,\Theta)\exp(in\theta). \tag{3.52}$$

The differential equations are Sturm-Liouville type equations of the form

$$\frac{d}{dx}\left(a(x)\frac{dh(x,X)}{dx}\right)+b(x)h(x,X)=F(x). \tag{3.53}$$

The solution, when $n\neq 0$, is considered for $n>0$ and $n<0$ separately as this makes the summation to derive the full Green's function easier to evaluate. For convenience let $g'_{n>0}(r,R,\Theta)=2\pi\exp(in\Theta)g_{n>0}(r,R,\Theta)$ then the solution to the differential equation for $n>0$ is

$$g'_{n>0}(r,R,\Theta)=\begin{cases}\left(A(R,\Theta)r^n+B(R,\Theta)r^{-n}\right) & r<R\\ \left(C(R,\Theta)r^n+D(R,\Theta)r^{-n}\right) & R<r.\end{cases} \tag{3.54}$$

This is the general solution to the homogeneous equation avoiding the delta function at $r=R$, Roach [52]. To determine the remaining functions there are four conditions that need to be applied. The first is that the solution is bound at $r=0$ so $B(R,\Theta)=0$. The second condition is the application of the boundary condition, (3.51), at $r=r_0$ which results in the expression

$$D(R,\Theta)=r_0^{2n}C(R,\Theta). \tag{3.55}$$

There are two additional conditions for this type of differential equation from the Schwartz theory of distributions, Stakgold [46], Trim [47]. These were not explicitly applied during the derivation of the Green's function in the previous section. The first is that $g_n(r,R,\Theta)$ is continuous over the interval $[0,r_0]$ so the solutions defined over the two intervals are equal at $r=R$ therefore

$$A(R,\Theta)R^n+B(R,\Theta)R^{-n}=C(R,\Theta)R^n+D(R,\Theta)R^{-n}. \tag{3.56}$$

The final condition for solving the general equation (3.53) states that the derivative of $h(x,X)$ is continuous except at $x=X$ where there is a finite discontinuity in $dh(x,X)/dx$ of magnitude $1/a(X)$ so that

$$\lim_{x\to X^+}\frac{dh(x,X)}{dx}-\lim_{x\to X^-}\frac{dh(x,X)}{dx}=\frac{1}{a(X)} \tag{3.57}$$

The discontinuity, when solving (3.51), has a magnitude of $1/R$ therefore

$$\frac{1}{R}=\left(nC\left(R,\Theta\right)R^{n-1}-nD\left(R,\Theta\right)R^{-n-1}-nA\left(R,\Theta\right)R^{n-1}+nB\left(R,\Theta\right)R^{-n-1}\right). \quad (3.58)$$

With these four equations the four functions can be defined as

$$\begin{aligned} A\left(R,\Theta\right)&=-\frac{1}{2n}\left(\frac{R^n}{r_0^{2n}}+\frac{1}{R^n}\right), \quad B\left(R,\Theta\right)=0, \\ C\left(R,\Theta\right)&=-\frac{R^n}{2nr_0^{2n}}, \qquad D\left(R,\Theta\right)=-\frac{R^n}{2n}. \end{aligned} \quad (3.59)$$

By substituting these functions back into (3.54) then

$$g_{n>0}\left(r,R,\Theta\right)=\begin{cases} -\dfrac{r^n}{2n}\left(\dfrac{R^n}{r_0^{2n}}+\dfrac{1}{R^n}\right)\dfrac{\exp\left(-in\Theta\right)}{2\pi} & r\le R \\ -\dfrac{R^n}{2n}\left(\dfrac{r^n}{r_0^{2n}}+\dfrac{1}{r^n}\right)\dfrac{\exp\left(-in\Theta\right)}{2\pi} & R\le r. \end{cases} \quad (3.60)$$

By the same argument the solution for $g_{n<0}$ can be expressed as

$$g_{n<0}\left(r,R,\Theta\right)=\begin{cases} -\dfrac{r^{|n|}}{2|n|}\left(\dfrac{1}{R^{|n|}}+\dfrac{R^{|n|}}{r_0^{2|n|}}\right)\dfrac{\exp\left(i|n|\Theta\right)}{2\pi} & r\le R \\ -\dfrac{R^{|n|}}{2|n|}\left(\dfrac{1}{r^{|n|}}+\dfrac{r^{|n|}}{r_0^{2|n|}}\right)\dfrac{\exp\left(i|n|\Theta\right)}{2\pi} & R\le r. \end{cases} \quad (3.61)$$

The solution is expressed in terms of the absolute value of n so that the negative part of the number is explicitly contained in the solution. This makes the summation, for the expression of the Green's function, easier in the following steps.

Let $g_0'=2\pi g_0$, then the general solution, when $n=0$ to the differential equation, (3.51), is

$$g_0'=\begin{cases} -\dfrac{r^2}{2r_0^2}+E\left(R,\Theta\right)\log\left(r\right)+F\left(R,\Theta\right) & r<R \\ -\dfrac{r^2}{2r_0^2}+H\left(R,\Theta\right)\log\left(r\right)+J\left(R,\Theta\right) & R<r. \end{cases} \quad (3.62)$$

Again, the restriction on the intervals is to avoid the delta function at $r=R$. Applying the same conditions as before

$$H\left(R,\Theta\right)=1. \quad (3.63)$$

Due to the boundary conditions at $r=r_0$. At $r=0$ the function needs to be bound so $E\left(R,\Theta\right)=0$. The Green's function is required to be continuous at $r=R$, therefore

$$E\left(R,\Theta\right)\log\left(R\right)+F\left(R,\Theta\right)=H\left(R,\Theta\right)\log\left(R\right)+J\left(R,\Theta\right). \quad (3.64)$$

Finally, the discontinuity condition requires that

$$\frac{H(R,\Theta)}{R}-\frac{E(R,\Theta)}{R}=\frac{1}{R}. \tag{3.65}$$

From these conditions

$$\begin{aligned} &H(R,\Theta)=1, \quad E(R,\Theta)=0, \\ &J(R,\Theta)=F(R,\Theta)-\log R. \end{aligned} \tag{3.66}$$

So $g_0(r,R,\Theta)$ can now be determined up to an additive function $F(R,\Theta)$ as

$$g_0(r,R,\Theta)=\begin{cases} -\dfrac{r^2}{4\pi r_0^2}+\dfrac{F(R,\Theta)}{2\pi} & r\le R \\ -\dfrac{r^2}{4\pi r_0^2}+\dfrac{1}{2\pi}\log(r)+\dfrac{F(R,\Theta)}{2\pi}-\dfrac{1}{2\pi}\log(R) & R\le r. \end{cases} \tag{3.67}$$

Therefore, the full Green's function can be written as

$$G(r,\theta,R,\Theta)=\begin{cases} -\dfrac{r^2}{4\pi r_0^2}+\dfrac{F(R,\Theta)}{2\pi}- \\ \displaystyle\sum_{n=-\infty}^{-1}\frac{r^{|n|}}{2|n|}\left(\frac{R^{|n|}}{r_0^{2|n|}}+\frac{1}{R^{|n|}}\right)\frac{\exp(i|n|\Theta)}{2\pi}\exp(in\theta)- \\ \displaystyle\sum_{n=1}^{\infty}\frac{r^n}{2n}\left(\frac{R^n}{r_0^{2n}}+\frac{1}{R^n}\right)\frac{\exp(-in\Theta)}{2\pi}\exp(in\theta) & r\le R \\ \\ -\dfrac{r^2}{4\pi r_0^2}+\dfrac{1}{2\pi}\log\left(\dfrac{r}{R}\right)+\dfrac{F(R,\Theta)}{2\pi}+ \\ \displaystyle\sum_{n=-\infty}^{-1}-\frac{R^{|n|}}{2|n|}\left(\frac{r^{|n|}}{r_0^{2|n|}}+\frac{1}{r^{|n|}}\right)\frac{\exp(i|n|\Theta)}{2\pi}+ \\ \displaystyle\sum_{n=1}^{\infty}-\frac{R^n}{2n}\left(\frac{r^n}{r_0^{2n}}+\frac{1}{r^n}\right)\frac{\exp(-in\Theta)}{2\pi}\exp(in\theta) & R\le r \end{cases} \tag{3.68}$$

The two sums can be directly written as a single sum from one to infinity, this is where the separation between $n<0$ and $n>0$ becomes useful. The Green's function can be simplified to

$$G(r,\theta,R,\Theta)=\begin{cases} -\dfrac{r^2}{4\pi r_0^2}+\dfrac{F(R,\Theta)}{2\pi} \\ -\dfrac{1}{2\pi}\displaystyle\sum_{n=1}^{n=\infty}\frac{r^n}{n}\left(\frac{R^n}{r_0^{2n}}+\frac{1}{R^n}\right)\cos(n(\theta-\Theta)) & r\le R \\ -\dfrac{r^2}{4\pi r_0^2}+\dfrac{1}{4\pi}\log\left(\dfrac{r}{R}\right)+\dfrac{F(R,\Theta)}{2\pi} \\ -\dfrac{1}{2\pi}\displaystyle\sum_{n=1}^{n=\infty}\frac{1}{n}\left(\frac{R^n}{r^n}+\frac{r^nR^n}{r_0^{2n}}\right)\cos(n(\theta-\Theta)) & R\le r \end{cases}. \tag{3.69}$$

Using the identity (3.35) the Green's function can be written in closed form as

$$G(r,\theta,R,\Theta) = -\frac{r^2}{4\pi r_0^2} + \frac{F(R,\Theta)}{2\pi} + \frac{1}{4\pi}\log\left(\frac{\left(r^2+R^2-2rR\cos(\theta-\Theta)\right)\left(r_0^4+r^2R^2-2rRr_0^2\cos(\theta-\Theta)\right)}{r_0^4R^2}\right). \tag{3.70}$$

This matches the Green's function stated by Trim [47], and the previous derivation, (3.45) provided that

$$a_0(R,\Theta) = \frac{F(R,\Theta)}{2\pi} - \frac{1}{4\pi}\log R^2. \tag{3.71}$$

If the functions in (3.66) were specified so that $F(R,\Theta) = J(R,\Theta) + \log R$ then the Green's function can be defined as

$$G(r,\theta,R,\Theta) = -\frac{r^2}{4\pi r_0^2} + \frac{J(R,\Theta)}{2\pi} + \frac{1}{4\pi}\log\left(\frac{\left(r^2+R^2-2rR\cos(\theta-\Theta)\right)\left(r_0^4+r^2R^2-2rRr_0^2\cos(\theta-\Theta)\right)}{r_0^4}\right), \tag{3.72}$$

now both derived Green's function, (3.45) and (3.72) are equal up to the addition of a function dependent on (R,Θ).

Two Green's functions have now been derived which match the published literature. These functions differ by functions dependent on (R,Θ). The remainder of this section shows that the two derived Green's functions result in equivalent phases.

For simplicity Green's function (3.72) can be written as

$$G(r,\theta,R,\Theta) = \frac{J(R,\Theta)}{2\pi} + G_A(r,\theta,R,\Theta), \tag{3.73}$$

where $G_A(r,\theta,R,\Theta)$ represents the rest of the terms in the Green's function. Therefore from equation (3.31), the phase is defined as

$$\phi(\mathbf{x}) = -\frac{k}{I_0}\int \frac{\partial I(\mathbf{x}',z)}{\partial z} G_A(\mathbf{x},\mathbf{x}')\,d^2\mathbf{x}' - \frac{k}{I_0}\int \frac{\partial I(\mathbf{x}',z)}{\partial z}\frac{J(\mathbf{x}')}{2\pi}\,d^2\mathbf{x}'. \tag{3.74}$$

The first integral evaluates to a function dependent on (r,θ) and the second integral evaluates to give a constant. Therefore, the addition to the Green's function of a function depending only on (R,Θ) will only add a constant to the phase.

This is consistent with the phase only being defined up to the addition of a constant due to the Neumann boundary conditions, Trim [47] and is consistent with the definition of the Green's function (3.32) which is only dependent on (r,θ).

This argument is the same for each derived Green's function; the addition of a function depending on (R,Θ) will only add a constant to the phase. Therefore, two of the solutions stated in the literature, Stakgold [46], Trim [47], result in equivalent phases.

3.2.5. Validity of the Green's Functions

As described in the introduction, there are three Green's functions stated in the literature that claim to solve the ITE. The derived Green's function (3.70) matches Trim [47] and, (3.72) and (3.45) match Stakgold [46]. It has also been stated by Zhu, *et al.* [48] that the Green's function that solves the ITE is

$$G(r,\theta,R,\Theta) = -\frac{1}{4\pi}\log\frac{\left(r^2+R^2-2rR\cos(\theta-\Theta)\right)\left(r_0^4+r^2R^2-2rRr_0^2\cos(\theta-\Theta)\right)}{r_0^4}, \tag{3.75}$$

The validity of each of the Green's functions can be tested by substituting it back into the differential equation and the boundary condition (3.46), and by showing that the divergence theorem is satisfied (3.27).

Each of the derived Green's functions contain a term for the free space Green's function, Trim [47],

$$U(r,\theta,R,\Theta) = \frac{1}{4\pi}\log\left(r^2+R^2-2rR\cos(\theta-\Theta)\right). \tag{3.76}$$

For the Green's functions to be valid, substitution of this expression back into the differential equation, (3.24), must result in a delta function at $(r,\theta)=(R,\Theta)$ and the remaining terms in the Green's function evaluate to $-1/2\pi r_0$. It is shown by Strichartz [51] that $U(r,\theta,R,\Theta)$ satisfies the differential equation producing a delta function at

$(r,\theta)=(R,\Theta)$ as required. By directly evaluating the Laplacian of the remaining terms, it is straightforward to show that

$$\nabla^2 G(\mathbf{x},\mathbf{x}')=\delta(\mathbf{x}-\mathbf{x}')-\frac{1}{2\pi r_0}. \tag{3.77}$$

This is true for each for each of the derived Green's functions.

For a valid solution the Green's function must also satisfy the boundary conditions (3.32). The derivative of Green's function (3.70) with respect to r is

$$\begin{aligned}\frac{\partial G(r,\theta,R,\Theta)}{\partial r}&=\frac{\partial}{\partial r}\left(-\frac{r^2}{4\pi r_0^2}+\frac{B(R,\Theta)}{2\pi}\right)\\&\quad+\frac{1}{4\pi}\frac{\partial}{\partial r}\left(\log\left(\frac{\left(r^2+R^2-2rR\cos(\theta-\Theta)\right)\left(r_0^4+r^2R^2-2rRr_0^2\cos(\theta-\Theta)\right)}{r_0^4R^2}\right)\right).\end{aligned} \tag{3.78}$$

The derivatives can be evaluated to give

$$\begin{aligned}\frac{\partial G(r,\theta,R,\Theta)}{\partial r}&=-\frac{r}{2\pi r_0^2}+\frac{1}{4\pi}\frac{2r-2R\cos(\theta-\Theta)}{\left(r^2+R^2-2rR\cos(\theta-\Theta)\right)}\\&\quad+\frac{1}{4\pi}\frac{\left(2rR^2-2Rr_0^2\cos(\theta-\Theta)\right)}{\left(r_0^4+r^2R^2-2rRr_0^2\cos(\theta-\Theta)\right)}.\end{aligned} \tag{3.79}$$

Setting $r=r_0$ and writing the second two terms with a common denominator then gives

$$\frac{\partial G(r_0,\theta,R,\Theta)}{\partial r}=-\frac{1}{2\pi r_0}+\frac{1}{4\pi}\frac{2r_0\left(r_0-R\cos(\theta-\Theta)\right)+2R\left(R-r_0\cos(\theta-\Theta)\right)}{r_0\left(r_0^2+R^2-2r_0R\cos(\theta-\Theta)\right)}. \tag{3.80}$$

Upon further simplification

$$\frac{\partial G(r_0,\theta,R,\Theta)}{\partial r}=0 \tag{3.81}$$

which satisfies the boundary condition. Similarly, Green's function (3.72) can also be shown to satisfy the boundary condition. The final condition is to satisfy the divergence theorem, substituting (3.77) and (3.81) into (3.27) then

$$\int_A \delta(\mathbf{x},\mathbf{x}')-\frac{1}{2\pi r_0}\,d^2A=\oint_C 0\,dC. \tag{3.82}$$

Both sides of this equation evaluate to zero, therefore the divergence theorem is satisfied. This is true for both derived Green's functions.

The Green's function stated by Zhu, *et al.* [48] contains the term $-U(r,\theta,R,\Theta)$ which gives a negative delta function at $(r,\theta)=(R,\Theta)$. The remaining terms evaluate to zero so the Laplacian of the full Green's function is

$$\nabla^2 G(\mathbf{x},\mathbf{x}') = -\delta(\mathbf{x}-\mathbf{x}'). \tag{3.83}$$

This does not necessarily mean the Green's function is wrong, as the conditions are not specified in the literature and the differential equation defining the Green's function could be different as there was a choice in (3.28) to modify the differential equation or the boundary conditions of the Green's function.

By evaluating the derivative of the Green's function with respect to r, and evaluating it on the boundary then $\nabla \cdot G(r,\theta,R,\Theta)=0$. Substituting these expressions into the divergence theorem then gives

$$\int_A \delta(\mathbf{x},\mathbf{x}')d^2A = \oint_C 0dC. \tag{3.84}$$

The left hand side of this equation evaluates to one and the right to zero therefore this condition is not satisfied and the Green's function stated by Zhu, *et al.* [48] is incorrect.

This section has shown that the Green's functions derived in this chapter are valid solutions as they satisfy all the necessary conditions. It has also shown that Green's function (3.75), which is the only one stated in the literature with a direct relation to calculating phase, is wrong and does not correctly solve the ITE.

As all the Green's functions derived in this chapter are equivalent, throughout the remainder of this thesis Green's function (3.70) will be used in the simulations and in all the following discussion.

3.2.6. Fourier Transform Solution to the Intensity Transport Equation

In the previous sections Green's functions were derived that can be used to solve the ITE. As an alternative to using a Green's function, the ITE can also be solved using Fourier transforms. This is a different approach to solving the same problem and so analytically, the final solution will be the same. When implemented numerically it will be shown that there are differences between the two solutions to the ITE.

The use of Fourier analysis has the advantage that it is naturally suited to analysing sampled data and the analysis may be easier to extend to arbitrary aperture shapes. The solution to the differential equation is not specific to the shape of the aperture but the boundary conditions depend on the shape, unlike the Green's function, which depends specifically on the geometry.

Additionally, there is no condition requiring the phase to be continuous. Therefore, this approach should be able to deal with a finite number of discontinuities in the phase.

The Fourier transform solution is in two parts. First the differential equation,

$$-\frac{k}{I_0}\frac{\partial I(\mathbf{x},z)}{\partial z} \approx P(\mathbf{x})\nabla^2\phi_C(\mathbf{x}), \tag{3.85}$$

is solved giving the curvature component of the phase, $\phi_C(\mathbf{x})$, then the boundary condition is applied to calculate the non-curvature parts of the phase, $\phi_N(\mathbf{x})$. The final solution is then the sum of these two parts. The definition of curvature and non curvature was discussed in section 3.2.1.

Using the derivative property of the Fourier transform, as stated in equation (3.16), equation (3.85) can be Fourier transformed and solved as an algebraic equation and then transformed back to real space, Dorrer, *et al.* [27], Allen, *et al.* [28]. Therefore, the curvature component of the phase is

$$P(\mathbf{x})\phi_C(\mathbf{x}) = \frac{k}{I_0}\frac{\partial I(\mathbf{x},z)}{\partial z} \otimes \mathfrak{F}^{-1}\left\{-\frac{1}{4\pi^2|\xi|^2}\right\}. \tag{3.86}$$

The inverse Fourier transform is represented by $\mathfrak{F}^{-1}$ and $\otimes$ denotes a convolution. At $\xi = 0$ there is a singularity. The value in Fourier space at $\xi = 0$ is therefore set to zero to remove this undefined value. This spatial frequency corresponds to the mean phase value. Setting it to zero is consistent with an inability to calculate absolute phase.

The solution, neglecting the boundary conditions, is not dependent on a specific shape of the aperture although the shape needs to be defined. This allows for more complicated shapes than the Green's function where the solution was derived specifically for a circular aperture.

It can be shown, Appendix B, that this part of the solution is identical to using the free space Green's function, (3.34), to solve the ITE, without boundary conditions.

The full solution, $\phi(\mathbf{x})$, to the differential equation is made up of two parts so that

$$\phi(\mathbf{x}) = \phi_C(\mathbf{x}) + \phi_N(\mathbf{x}). \tag{3.87}$$

The curvature part $\phi_C(\mathbf{x})$ corresponds to data from the open area and the non-curvature part $\phi_N(\mathbf{x})$ to the boundary data, Figure 3.2. This is the step that Allen, *et al.* [28] and Dorrer, *et al.* [27] omitted. Therefore, there is no contribution to the phase from the harmonic functions in their solutions. The non-curvature component of the phase is zero inside the aperture so that

$$\nabla^2 \phi_N(\mathbf{x}) = 0. \tag{3.88}$$

The boundary conditions depend on the shape of the aperture. In the same way as with the Green's function, $P(\mathbf{x})$ is defined as a disk with a radius r_0 and area A, as shown in Figure 3.2.

The solution to equation (3.88), over the area of a disk, is the set of harmonic functions, which in polar coordinates are defined as

$$\phi_N(r,\theta) = \sum_{n=0} \frac{r^n}{r_0^{\,n}} \left(a_n \cos(n\theta) + b_n \sin(n\theta) \right). \tag{3.89}$$

The coefficients of the nth order harmonic functions are a_n and b_n. Again, the boundary conditions for the ITE are defined as

$$\frac{\partial \phi(r_0,\theta)}{\partial r} = h(r_0,\theta), \tag{3.90}$$

where $h(r_0,\theta)$ is a known function on the boundary. Therefore, by substituting in equations (3.87) and (3.89), and evaluating at $r = r_0$

$$h(r_0,\theta) = \frac{\partial \phi_C(r_0,\theta)}{\partial r} + \sum_{n=0} \frac{n}{r_0} \left(a_n \cos(n\theta) + b_n \sin(n\theta) \right). \tag{3.91}$$

The derivative of the phase with curvature, $\phi_C(r,\theta)$, has only been included symbolically in the analysis as it is phase dependent and when using real data it will be defined on a discrete array, so it will have to be estimated numerically.

All the functions are 2π periodic in θ, therefore, by using a Fourier transform with respect to θ, denoted by $\mathfrak{F}_\theta$, along the boundary at $r = r_0$

$$\mathfrak{F}_\theta\left\{h(r_0,\theta)-\frac{\partial\phi_C(r_0,\theta)}{\partial r}\right\}=\sum_{n=1}\frac{n}{2r_0}\begin{pmatrix} a_n\left(\delta\left(\chi-\frac{n}{2\pi}\right)+\delta\left(\chi+\frac{n}{2\pi}\right)\right)+ \\ ib_n\left(\delta\left(\chi+\frac{n}{2\pi}\right)-\delta\left(\chi-\frac{n}{2\pi}\right)\right)\end{pmatrix} \tag{3.92}$$

where χ is the corresponding coordinate in the Fourier domain. This produces a spectrum of values, which encode the coefficients of the harmonic modes. The constant coefficient a_0 is undefined, which is consistent with an inability to measure absolute phase and is set to zero.

The real part of the spectrum corresponds to the coefficients a_n of the harmonic functions containing cosine functions and the imaginary part of the spectrum corresponds to coefficients b_n of the harmonic functions containing sine functions. The position within the spectrum relates to the order of the function and the value at each position relates to the coefficient of the harmonic function.

The final solution is then the curvature component (3.86) plus the non-curvature component (3.89) with appropriate values for a_n and b_n so that

$$\phi(\mathbf{x})=\frac{k}{I_0}\frac{\partial I(\mathbf{x},z)}{\partial z}\otimes\mathfrak{F}^{-1}\left\{-\frac{1}{4\pi^2|\xi|^2}\right\}+\sum_{n=0}\frac{r^n}{r_0^{\,n}}\left(a_n\cos(n\theta)+b_n\sin(n\theta)\right). \tag{3.93}$$

Having derived exact solution to the ITE the following section describes the process for numerically implementing them for use with experimental data.

3.3. Numerical Implementation of the Theory

So far, solutions have been derived that exactly solve the ITE, the following section describes the process for numerically simulating data and for analysing the simulated or measured data to calculate the phase.

The data acquired in planes E and G in Figure 3.1 are treated as functions a spatial variables irrespective of whether they are conjugate to a pupil plane or an image plane in the optical system.

3.3.1. Generation of the Simulated Data

This section describes the process to generate the simulated intensity measurements that are used as the input to calculate the phase. The intensity measurements are simulated in three parallel planes, as shown in Figure 3.1. The central plane is where the phase will be calculated. To simulate the data, the phase is defined on a grid and the complex amplitude of the wave in the centre plane is calculated using equation (2.5), where the intensity is constant over the area of a disk.

The propagation of the complex amplitude between planes is modelled using the angular spectrum of plane waves, Goodman [37]. The propagation is modelled by Fourier transforming the complex amplitude, which represents a set of plane waves propagating in different directions. These waves are then multiplied by a function that represents a phase change due to the propagation distance and an inverse Fourier transform recombines the plane waves to give the propagated complex amplitude.

This phase change is modelled by multiplying the plane wave spectrum by $\exp\left(ikz\sqrt{1-\lambda^2|\xi|^2}\right)$. In the simulations this is defined as

$$\exp\left(ikz\sqrt{1-\lambda^2 S^2\left(u^2+v^2\right)}\right). \tag{3.94}$$

Where u and v are the number of samples counted from a point defined to be the centre of the grid and S is a scale factor to convert the sample spacing into physical units,

$$S=\frac{\text{disk radius(samples)}}{\text{grid size}\times\text{disk radius(meters)}}. \tag{3.95}$$

The grid size is the number of points in the array in the horizontal or vertical direction. In this case, the array is square so both are equal.

The intensity is the magnitude squared of the propagated complex amplitude. Using this process the intensity can be simulated in the required planes for use as the input for the phase calculation process.

3.3.2. Description of the Inversion Process using Fourier Analysis

This section describes the process to calculate the phase from a set of defocused images using the Fourier transform method. The analysis is split into three parts; the first part describes the process to get the data into the correct form before applying the analysis. The second part calculates the curvature component of the phase and the final part calculates the non-curvature component of the phase. This is illustrated in Figure 3.4, which is a process map that was initially created to describe the data analysis process to AMO/Wavefront Sciences Inc where this analysis was applied in a commercial application. This is discussed further in section 4.4.

The first part of the calculation process, shown in the top left of Figure 3.4, is to define the data for the calculation. The axial derivative in plane F of Figure 3.1 is approximated from the difference between the intensity in planes E and G to give

$$Data(\mathbf{x}) = -\frac{k\left(I(\mathbf{x},\Delta z) - I(\mathbf{x},-\Delta z)\right)}{2\Delta z I_0} \tag{3.96}$$

where $I(\mathbf{x},\pm\Delta z)$ are the measurements in planes E and G of Figure 3.1 and the separation between these two planes is $2\Delta z$. This approximates the left side of (3.22) as the difference between the intensity measurements is a finite difference approximation to the derivative. Normalisation by the constant intensity can be achieved using the measurement in plane F, or using an average intensity value calculated from the two defocused images in planes E and G.

After processing the measurements to produce the correct data for the calculation process, the data is then split into data that is on the boundary and data that is within the disk. The definition of the boundary data is described in section 3.3.3.

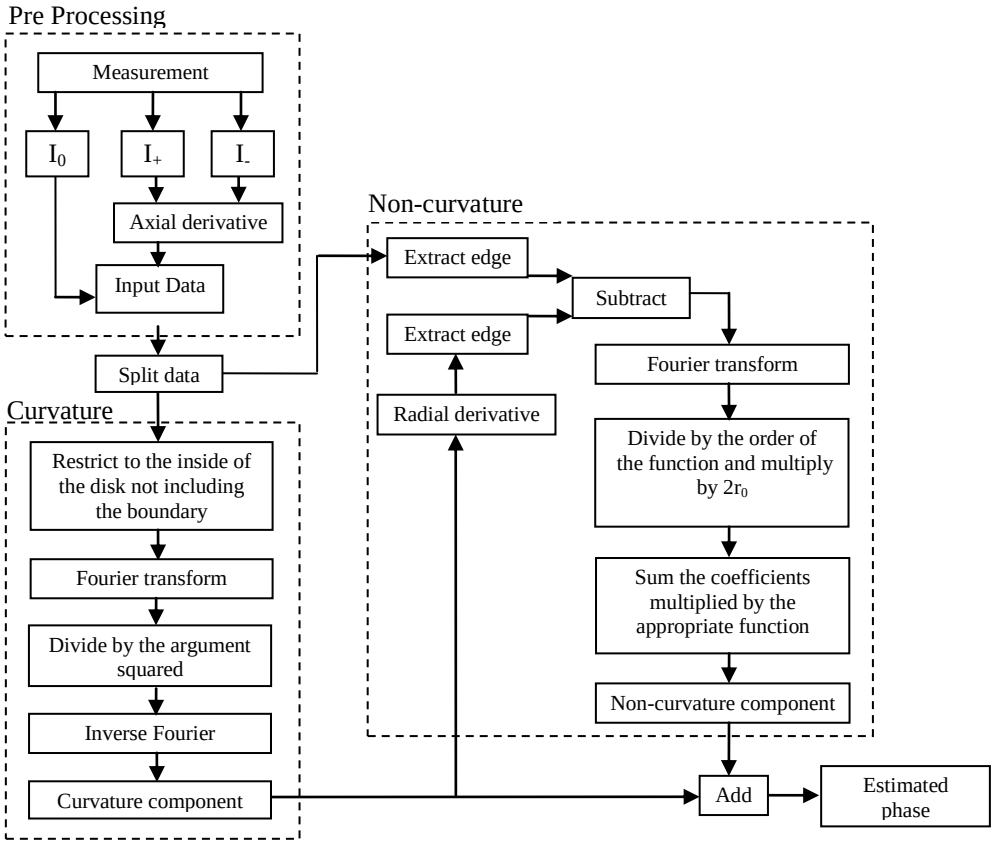


Figure 3.4 Flow diagram of the procedure for inverting the data using the Fourier method.

The first part of the calculation returns the curvature component of the phase, by taking the inner part of the data excluding the boundary. This is then processed as shown in the lower left of Figure 3.4. The curvature component is calculated using equation (3.86), the data is Fourier transformed and then divided by

$$-4\pi^2 S^2 \left(u^2 + v^2\right). \tag{3.97}$$

The number of samples counted from a point defined as the centre of the grid, in the x and y directions, are denoted by u and v . The scale factor S converts the sample spacing into physical units as defined in (3.95). The point where $u = v = 0$ is set to zero to remove the singularity. This has the effect of only allowing relative phase to be measured, as any constant information has been lost. Finally, applying an inverse Fourier transform gives the curvature component of the phase, defined over the area of the aperture.

The last step in the phase calculation is to calculate the non-curvature component of the phase, by analysing the data on the boundary of both the measured data and the radial derivative of the curvature component of the phase.

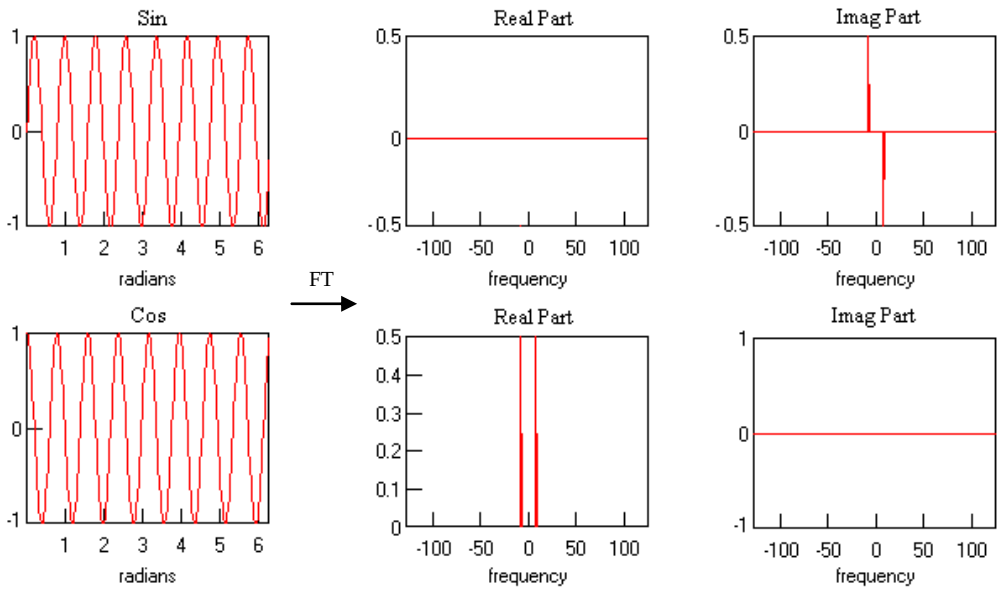


Figure 3.5 Example of the way the coefficients of the harmonic modes are encoded into the boundary data. The left figures represent the data along the circular boundary so the data is by definition 2π periodic.

The data on the boundary is a combination of the radial derivative of the phase with curvature and without curvature, (3.90). So to obtain information about the non-curvature part of the phase the radial derivative of the phase with curvature, is subtracted from the boundary data.

To calculate the radial derivative at the edge a backwards finite difference method is used with appropriate scaling for the sample spacing. Summing the derivatives along the Cartesian grid using the adjacent pixel within the disk in the appropriate direction and with the appropriate weighting gives the radial derivative. In the simulations this is calculated using

$$\begin{aligned}\frac{\partial f(r,\theta)}{\partial r} \approx & \frac{S|u|}{u^2+v^2}\left(f(u,v)-f(u-\operatorname{sgn}(u),v)\right) \\ & +\frac{S|v|}{u^2+v^2}\left(f(u,v)-f(u,v-\operatorname{sgn}(v))\right)\end{aligned} \tag{3.98}$$

where f is an arbitrary function, the scale factor S is defined as $S = \text{disk radius(samples)}/\text{disk radius(meters)}$ and

$$\operatorname{sgn}(x) = \begin{cases} 1 \forall x > 0 \\ 0 \forall x = 0 \\ -1 \forall x < 0. \end{cases} \tag{3.99}$$

A one sided approximation of the radial derivative is used as the phase is only defined inside the disk. The singular point at the centre, when $u = v = 0$, is not important as only the derivative on the boundary is used.

The boundary data and the radial derivative of the phase that has curvature are the input for equation (3.92), which calculates a spectrum of coefficients corresponding to the missing modes. This is illustrated in Figure 3.5, where the sine and cosine waves represent the boundary data, on the left, with the Fourier transform shown on the right. The real and imaginary parts of the spectrum differentiate between the sine and cosine function of the same order. The position in the spectrum defines the order of the missing harmonic function and the value at that position defines the coefficient of the particular mode.

This section has described the pre processing of the measurements, which is required before using any solution to the ITE to calculate the phase. The section has also described the numerical implementation of the Fourier transform method. The definition of the boundary has not been specified this is discussed in the following section

3.3.3. Definitions of the Boundary Data

In the previous section, the numerical implementation of the Fourier transform method was described, but the definition of the boundary data was left unspecified. This section considered three different methods for calculating the boundary data from the available data.

Three methods are considered, the first and simplest is a nearest neighbour method. It approximates the data on the boundary by scanning the array to assess which sample points are within half of the sample spacing from the specified radius (i.e. within half a pixel from the edge of the pupil). These are then sorted into an array by increasing angle. Edge data extracted in this way results in samples that are not uniformly spaced and many samples are at an incorrect radius.

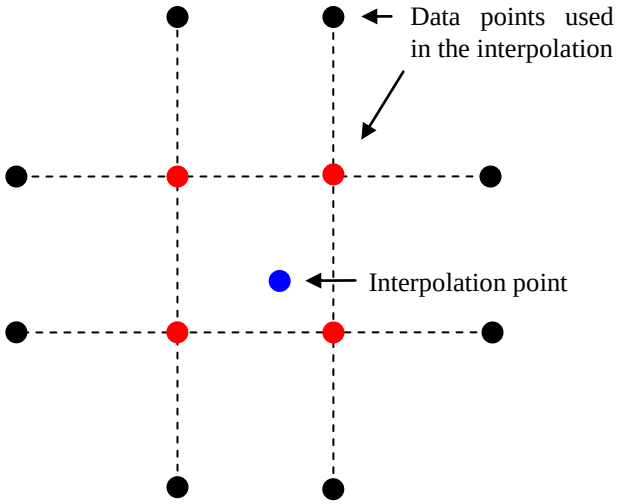


Figure 3.6 Schematic of the data points used in the bicubic interpolation

The second method is a bicubic interpolation. It uses the 12 data points, surrounding the interpolated point as shown in Figure 3.6. To calculate the value at the interpolation point, in blue, finite difference approximations for the derivatives $\partial/\partial x$, $\partial/\partial y$ and $\partial/\partial x\partial y$ are calculated at the four surrounding points marked in red. Using the values of the function and its derivatives at these points the value at the interpolation point can be approximated, Press, *et al.* [53]. Some of the points will be outside the boundary but the data will be smooth so these points can still be used.

The third method is a sinc interpolation. The value at the interpolated point, $f\left(u_{itp}, v_{itp}\right)$, is calculated using

$$f\left(u_{itp}, v_{itp}\right) = \sum_{n=1}^{N} f\left(u_n, v_n\right) \frac{\sin\left(\pi\left(u_n - u_{itp}\right)\right)}{\pi\left(u_n - u_{itp}\right)} \frac{\sin\left(\pi\left(v_n - v_{itp}\right)\right)}{\pi\left(v_n - v_{itp}\right)}. \tag{3.100}$$

Where $\left(u_{itp}, v_{itp}\right)$ is the interpolation point and $\left(u_n, v_n\right)$ is one of the N sampled points indexed by n within the disk.

The advantage of the nearest neighbour method is speed although this may be at the expense of accuracy as the sample points are not at the correct radius and are unevenly spaced. However, if the sampling is sufficiently fine then the separation of the data point will be approximately equal and the effects of the non-uniform spacing will be small.

The bicubic interpolation uses more of the data to interpolate the values at the correct radius and evenly sampled positions. This should improve the estimation of the boundary data. The sinc interpolation should be a further improvement in terms of accuracy as all the data within the disk is used to estimate each point on the boundary but the addition of more data points in the calculation will increase the computation time. This process defines a band-limited function where the band-limit is that of the data which is generally under sampled.

Only in the nearest neighbour method is the number of sample points defined by the data and therefore sample spacing. Therefore, for consistency in the examples below each interpolation method uses the same number of samples defined from the nearest neighbour method, as the interpolation methods cannot add more information that is useful.

Given a defined sample spacing, measured in arc length, the critical frequency, f_c, for Nyquist sampling can be defined, Brigham [38], so that

$$f_c = \frac{Number\,Of\;Samples}{4\pi} \tag{3.101}$$

and therefore the number of determinable coefficients is defined.

If only a limited number of modes are of interest, for example in ophthalmology, Campbell [2], then the number of samples can be determined from the azimuthal frequency of the highest order mode that is of interest. The highest frequency can then be used to specify the required sample spacing.

To illustrate the differences between these methods all the following figures in this section use the same parameters, except where specified. The input phase is defined as Z_5^3 with a coefficient of 1 radian, the radius of the disk is set at 1 mm represented by 64 samples in an array with 256×256 elements. The wavelength is 633 nm and the plane spacing between planes E, F and G in Figure 3.1 is $5\,mm$.

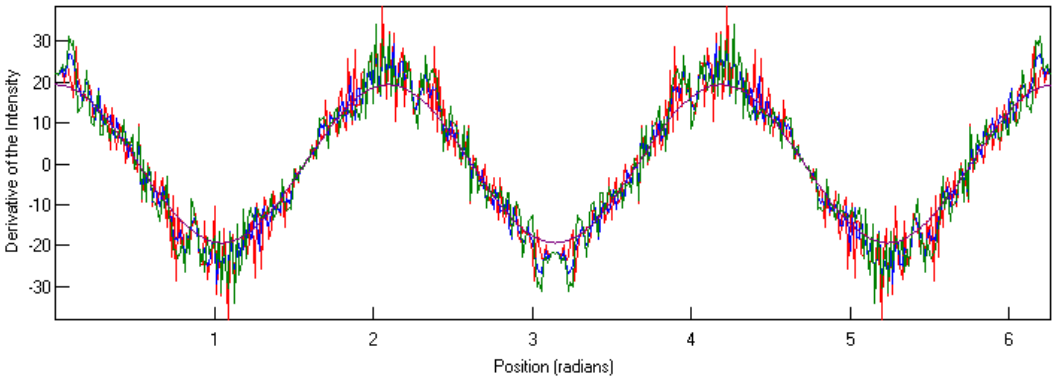
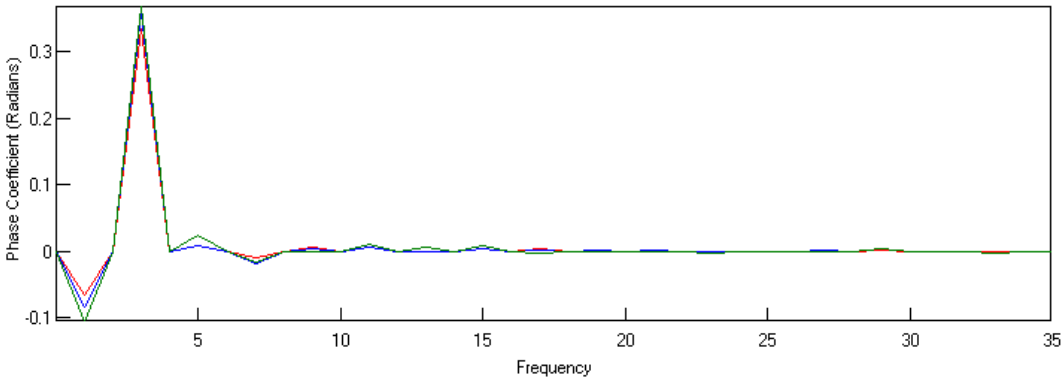
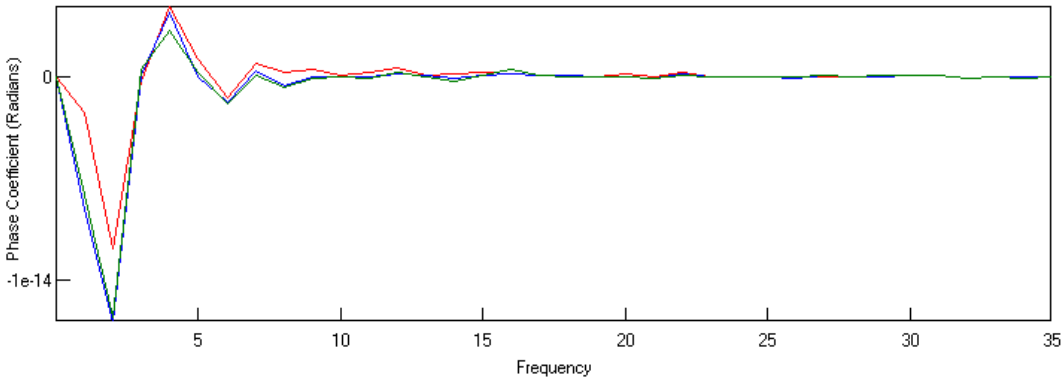


Figure 3.7 Boundary data used to calculate the coefficients of the harmonic functions. Red represents the data calculated using the nearest neighbour method, blue represents the bicubic interpolated data, green represents the sinc interpolated data and purple represents the boundary data calculated using the ITE..



a)



b)

Figure 3.8 Calculated spectrum of coefficients. a) shows the real part of the spectrum. b) shows the imaginary part of the spectrum. In both graphs, the coefficients calculated from the nearest neighbour data are shown in red, blue represents the coefficients calculated from the bicubic interpolated data and green represents the coefficients from the sinc interpolated data

Figure 3.7 shows the boundary data calculated using each of the interpolation methods. In each case, the data shows a distinct cosine wave of frequency 3 cycles over the length of the circumference and on top of that there are higher frequency waves of lower

amplitude. Each of the methods gives a similar estimation of the boundary data and all are three methods are close to the boundary data calculated using the ITE.

Using this data the set of coefficient corresponding to the harmonic modes can be calculated. Figure 3.8 a) shows the coefficient relating to the harmonic functions containing cosine functions and Figure 3.8 b) shows the coefficients of the modes containing sine functions and is at the level of machine precision.

The coefficients calculated using each method are almost identical as the input data is very similar. The calculated phase is missing a contribution from the mode Z_3^3 this has a frequency of 3 and should be encoded on the real part of the spectrum, this is correctly shown in Figure 3.8. a). This mode should have a coefficients of 4 radians but has been incorrectly estimated as 0.35 radians. There are also small contributions from other modes that have been incorrectly calculated all of which should have a value of 0 radians.

These errors are due to other approximation like the finite difference approximation of the intensity derivative, the estimation of part of the phase that has curvature and the finite difference approximation of the radial derivative. These errors can be reduced but improving the estimation of the derivative of the intensity and increasing the number of samples.

As an example of the phase calculated using the three interpolation methods, Figure 3.9 shows the output phase coefficient plotted as a function of the input phase coefficient. The figure shows that the phase calculated using each of the methods is very similar.

Each method returns very similar results in all tested cases, there can still be substantial errors due to other approximations in the model. The reduction of these errors is considered in Chapter 5. As all the methods give similar results, only one method will be used throughout the remainder of this thesis, so the nearest neighbour method will be used, as it is the simplest.

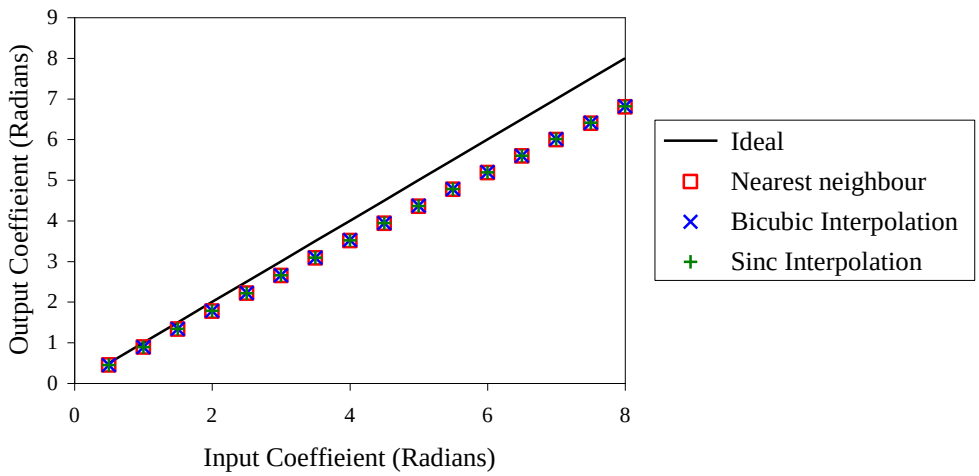


Figure 3.9 Output phase coefficient as a function of the input phase coefficient. The boundary data has been approximated using the nearest neighbour method, bicubic interpolation and sinc interpolation.

3.3.4. Implementing the Green's Function and Removing the Singularities

In the previous section, the numerical process for calculating the phase from the intensity data, using the Fourier transform method, was described. This section describes the numerical process to calculate the phase using a Green's function.

After making the measurements, the input data of the phase calculation is defined following the process described in the previous section. The phase can be calculated using one of the correct Green's functions from section 3.2.5. The phase at a point $\mathbf{x}$ can be calculated by approximating the integral in equation (3.31) by a Riemann sum so that

$$\phi(\mathbf{x}) = \sum_{n=1}^{N} G(\mathbf{x}, \mathbf{x}'_n)\, Data(\mathbf{x}'_n) A_n, \tag{3.102}$$

where N is the number of samples in the disk, the area of the disk is split into N areas A_n and the samples at position $\mathbf{x}'_n$ are indexed by n.

The numerical integration requires well defined values at each point but as can be seen from the Green's functions there are singularities at some combinations of $\mathbf{x}$ and $\mathbf{x}'_n$ giving undefined values. This section looks at defining values that can be used in the numerical process and the effect of the singularities on the calculated phase

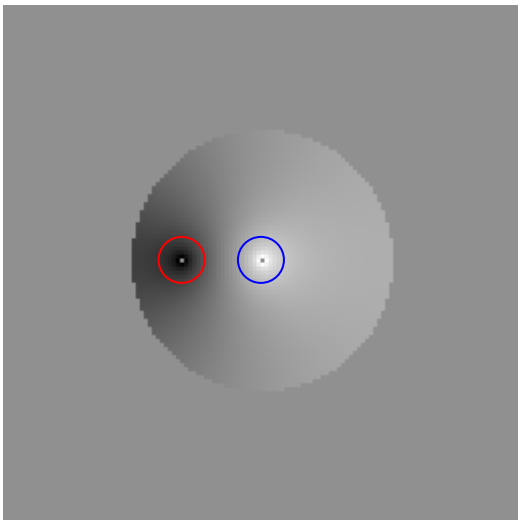


Figure 3.10 An example of Green's function (3.70) plotted for a specific value of (r,θ). The variable in the graph are (R,Θ). The singularity at (r,θ) is circled in red and the singularity at $R=0$ circled in blue. The singularities have been set to zero for display purposes.

The Green's functions have integrable logarithmic singularities. These points have no defined value but when the function is integrated, the singularities are removed giving a well defined value. As an example of the singularities in the Green's function, Figure 3.10 shows a plot of Green's function (3.70) for specific values of (r,θ) varies (R,Θ).

There are always two singularities. The singularity in Green's function (3.70) at $(r,\theta)=(R,\Theta)$, away from the boundary, is due to the term

$$\frac{1}{4\pi}\log\left(r^2+R^2-2rR\cos(\theta-\Theta)\right), \tag{3.103}$$

and is circled in red in Figure 3.10. The singularity at $R=0$, circled in blue is due to the term

$$-\frac{1}{4\pi}\log\left(r_0^4R^2\right). \tag{3.104}$$

On the boundary when $(r,\theta)=(R,\Theta)$ the singularity is caused by two terms, (3.103) and

$$\frac{1}{4\pi}\log\left(r_0^4+r^2R^2-2rRr_0^2\cos(\theta-\Theta)\right). \tag{3.105}$$

The values at both the singular points in Figure 3.10 have been set to zero for display purposes.

As the Green's function is a distribution, for a given point there is not necessarily a well defined value for the Green's function but when it is analytically multiplied by the data and integrated this will give a well defined value, Strichartz [51].

One method to remove the singularities is to define a modal projector function, Woods, *et al.* [30]. This function is calculated by multiplying the Green's function with an orthogonal base function and integrating the product over the area of the aperture. Therefore, the projector function has no singularities. However, when the data is analysed, by integrating the product of the data and the projector function over the area of the aperture, only the coefficient of the orthogonal base function is calculated.

This method has applications in for example ophthalmology, Campbell [2], where only a few coefficients are of interest. However, for the purposes of determining a 3D phase object in Chapter 6 the full phase estimate is required and an alternative method is required.

When used experimentally, the integration of the Green's function with the data is approximated by a Riemann sum. For numerical integration, the approximation of the Riemann sum requires well-defined values at each sampled point of the Green's function.

Teague [49] set the singularities to zero without any justification. Therefore, to remove this arbitrary process the remainder of this section will describe a method that correctly evaluates the singularities giving well-defined values for use in the phase calculation.

The phase can be calculated using a Green's function and equation (3.31). If the total area of the integral is now split into M smaller contiguous sub areas, A_m, then

$$\phi(\mathbf{x}) = \sum_{m=1}^{M} \int_{A_m} Data(\mathbf{x}') G(\mathbf{x},\mathbf{x}')\, d\mathbf{x}'. \tag{3.106}$$

Experimental data is sampled on a discrete grid so over the area of a pixel it is a constant. Therefore, the data can be taken out of the integral, when A_m represents the area of a pixel, so that

$$\phi(\mathbf{x}) = \sum_{m=1}^{M} Data(\mathbf{x}') \int_{A_m} G(\mathbf{x},\mathbf{x}')\, d\mathbf{x}'. \tag{3.107}$$

Where m indexes the M samples each with an associated area A_m. This leaves an integral of the Green's function over each sub area, which is independent of the data. The sub areas could be different sizes and shapes so that the total area exactly covers the disk. In the process shown here integral is over the area of a pixel.

By integrating the Green's function, the singularities can be removed. This changes the numerical integration from using a midpoint approximation to using the Green's function analytically integrated over each sub area.

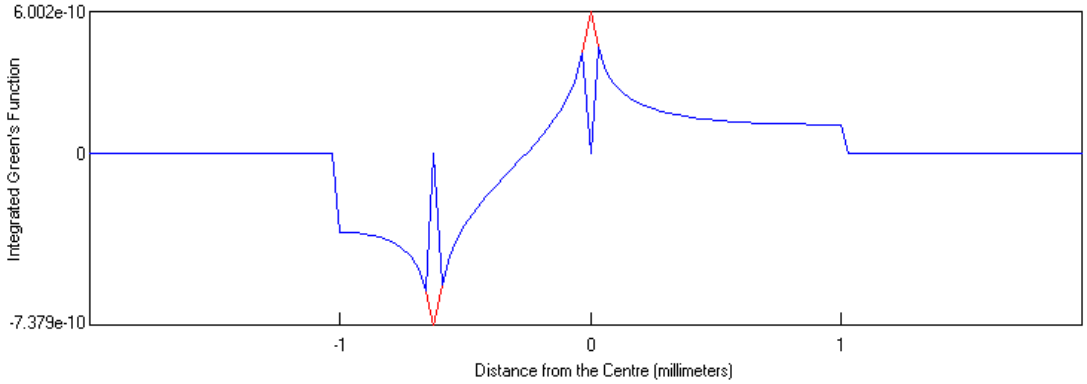
Integrating the full Green's function, over each sample area, can be defined analytically for a specific point (x, y) the Green's function is integrated over a square pixel with side length $2a$, at position (p, q) so that

$$\int_{A_m} G(\mathbf{x}, \mathbf{x}')\, d\mathbf{x}' = \int_{p-a}^{p+a} \int_{q-a}^{q+a} G(x, y, X, Y)\, dX dY. \tag{3.108}$$

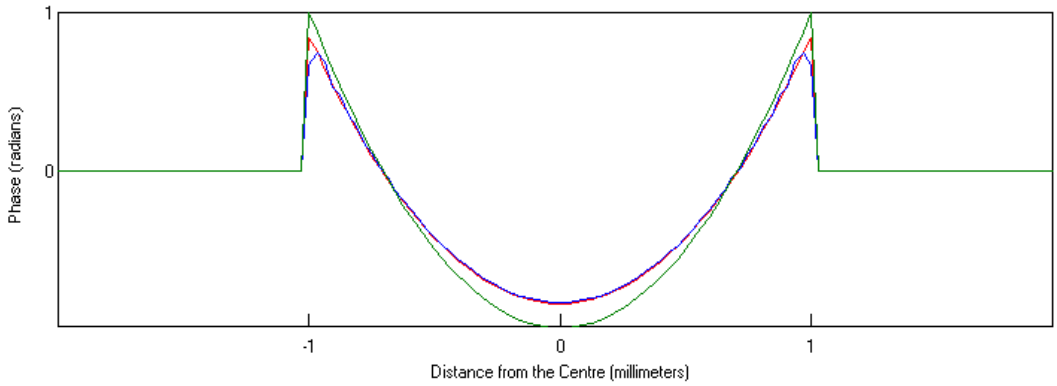
The resulting expression is not written here as it is lengthy but it can be calculated using a standard computer algebra package.

To show the difference between this integrated Green's function and the numerical integration, Figure 3.11 a) shows a cross section through the singularities of the numerically integrated Green's function, shown in Figure 3.10, and the equivalent cross section of the analytically integrated Green's function. This figure shows that the only difference between the two methods is at the singular points where there is now a defined value in place of arbitrarily setting the value to zero. At all other point the two methods are identical.

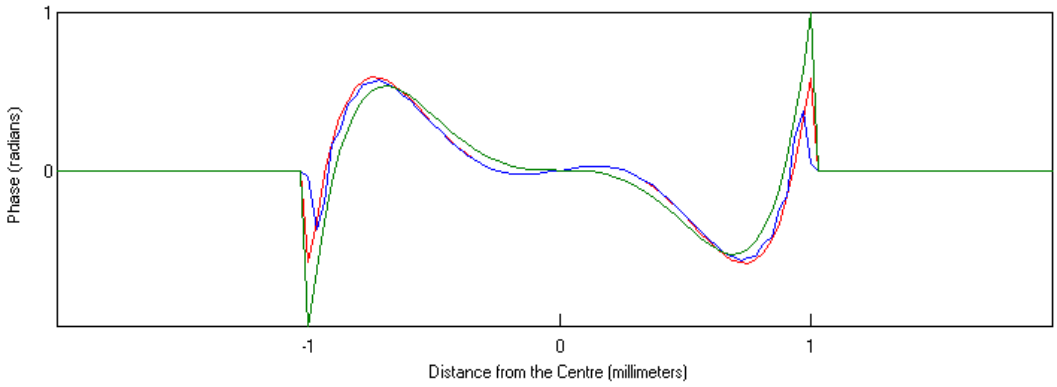
To show the difference between the phases, calculated using the two methods, Figure 3.11 b) and c) show examples of the calculated phase using each method in comparison to the set phase. In both examples, the phase calculated where the singularities are set to zero is very similar to the phase calculated using the more rigorous treatment of the singularities. However, the improvement is small and does not normally justify the additional computational expense.



a)



b)



c)

Figure 3.11 a) shows a cross section of the analytically integrated Green's functions in red and the numerically integrated Green's function, excluding the singular points, in blue. Examples of the calculated phase compared to the set phase are shown in b) where the set phase is Z_2^0 and in c) where the set phase is Z_5^3. The set phase is shown in green, the phase calculated from the numerically integrated Green's function is shown in blue and the phase calculated using the analytically integrated Green's function is shown in red.

Due to the computational expense and the minimal difference between the two methods the Green's function, at the singular points will be set to zero, following Teague [49] throughout the remainder of the thesis. This will introduce a small error into the solution but there are only two singularities for each reconstruction point.

3.4. Conclusions

This chapter described methods for calculating the phase from multiple defocused images using the ITE. This equation can be solved exactly for the special case of a wavefront defined over a circular aperture with constant intensity. Two solutions were derived; a Green's function solution and a Fourier transform solution.

Two different Green's functions were derived and shown to be equivalent resulting in phases differing by a constant. These Green’s functions are consistent with the Green's functions stated in the literature without derivation, Stakgold [46], Trim [47]. A third Green's function, also stated in the literature, Zhu, *et al.* [48], was shown to be wrong, as it failed to satisfy the necessary mathematical conditions for a valid solution.

A Fourier transform solution to the ITE was also derived. This method has been partially implemented by Dorrer, *et al.* [27] and Allen, *et al.* [28] but in each case the boundary condition is not included, this misses the contribution from the harmonic functions. It has been shown that the phase can be calculated using the full analytic solution to the ITE with the inclusion of the boundary conditions.

The numerical implementation of both these solutions was described. The analysis using Fourier transforms requires separation of the boundary data. It was shown that the calculated phase was unchanged using three different methods to define the boundary data.

The use of the Green's function required the evaluation of singular points. It was shown that the singularities could be removed by integrating the Green's function over the area of each pixel, but there is a large computational expense for only a minimal improvement. In this form, the method is not beneficial but the two methods could be combined so that the numerical integration is used for the non singular points and the analytically integrated Green's function is used for the singular points. This would give improved accuracy for a minimal increase in calculation time.

In comparisons the Green's function and the Fourier transform methods are both exact solutions to the ITE and are therefore identical except the Fourier transform method can solve a larger class of problem as there is no restriction to continuous functions.

Chapter 4 – Experimental Testing of the Theory

4.1. Introduction

In the previous chapter, theory was developed for calculating the phase from defocused intensity measurements and the numerical implementation for applying this to experimental data was described. In this chapter, the theory is tested with experimental data before considering potential improvements to the experimental method in the following chapter.

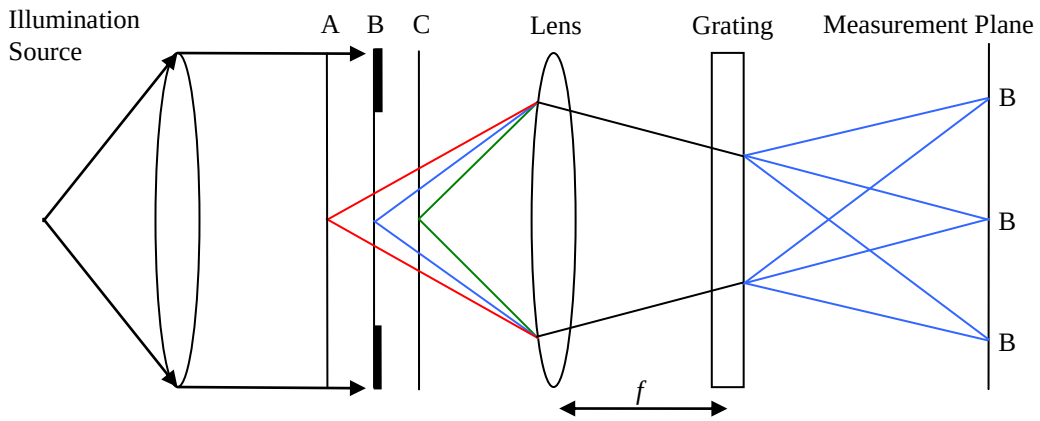
Two experiments were carried out, the first in the laboratory and the second was with AMO / Wavefront Sciences Inc.

4.2. Background – Simultaneous Multi-Plane Imaging

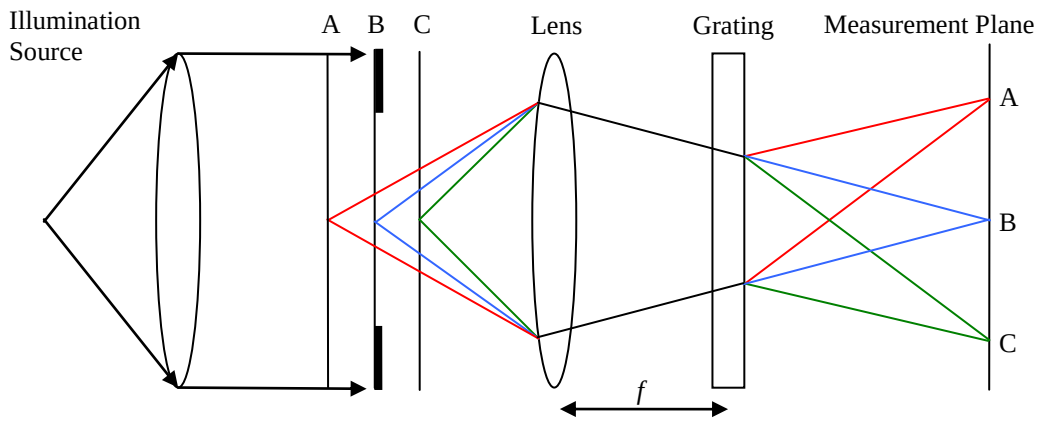
The experimental method involves measuring three images from which the phase change applied to the beam of light can be calculated. The experiment is initially setup so that the beam of light is collimated and there is no change in intensity between the 3 images, A, B and C in Figure 4.1 b). By moving the position of the light source, a phase change is applied to the beam of light causing it to converge or diverge depending if the light source is closer or further from the camera Figure 4.1 b).

This section describes the process for simultaneously measuring three defocused images using a distorted diffraction grating, Blanchard, *et al.* [41], Blanchard, *et al.* [54]. A standard diffraction grating has a series of lines or rules that splits the illuminating light into several beams travelling in different directions. This is shown in Figure 4.1 a). The figure shows plane B being imaged onto a detector. The diffraction grating has split the light into three beams creating three images of plane B. The images are formed in the diffraction orders of the grating where the light has constructively interfered and produced an image.

A distorted diffraction grating, so called due to the distorted lines, applies a different phase change to each diffraction order. The grating can be designed to apply a defocus phase change so that in each diffraction order a different plane is viewed.



a)



b)

Figure 4.1 Schematic of an imaging system using a) a standard straight line grating and b) a distorted diffraction grating. The thicker black line represents a stop between which the phase is to be determined.

Using the grating in combination with a lens simultaneously provides images of different planes in each diffraction order as shown in Figure 4.1 b). By arranging the components so that the images are all contained within the area of the camera the necessary defocused measurements used to calculate the phase can be captured simultaneously, Blanchard, *et al.* [41]. By separating the grating and the lens by the focal length of the lens, the images all have the same magnification, Djidel, *et al.* [55]. If the lens and grating are placed together as close a possible then the two defocused images also have equal magnification but one image is inverted, Blanchard, *et al.* [41]. This is the configuration used in section 4.3.

This is not the only way to make the measurements as, for example, a beam splitter and two cameras can be used, Paxman, *et al.* [56]. For validating the theory, the choice of experimental method here is not important. Therefore, this method using a diffraction

grating is used in both experiments, as it simultaneously provides all the required data and the experiment is simple to conduct.

4.3. Experiment Conducted in the Laboratory

To test the theory, developed in Chapter 3, with experimental data, a summer student, Thomas Bernard, conducted an experiment in the lab to measure sets of defocused images that can be used to calculate phase change applied to a beam of light.

The experimental setup is described in, Blanchard, *et al.* [41], and is shown in Figure 4.2. In this experiment, the source was placed $2\ m$ from a lens, which had focal length $60\ mm$ and a radius of $1\ cm$. The grating had a focal length of $2\ m$, the plane spacing was $2\ m$ and the wavelength was $633\ nm$. The experiment is setup with the light source positioned so that the illumination is a plane wave within the plane of the lens, where the phase will be estimated, and the measured images in planes A and B are the same. By translating the source from this point, an increased defocus phase change is applied to the wavefront. In this experiment, only two images, of planes A and B, are recorded. In this configuration, the image of plane B needs be rotated by 180 degrees as it is inverted in comparison to the image of plane A. The lens edges act as the system stop and the phase is calculated in the plane of the lens.

After making a set of measurements, the first step is to centre the image and calculate the position of the boundary. To calculate the centre and the radius of the disk, in each image, an iterative process was used. This process was only used with the experimental data where the phase was known to be mainly defocus, so in each image the data approximately covers a disk.

The iterative process starts with an estimate for the centre and the radius, the centre of the circle is then moved around the image to maximise the sum of the sample values inside the disk, as the intensity values are greater than or equal to zero. The position with the greatest value is taken as the centre of the image

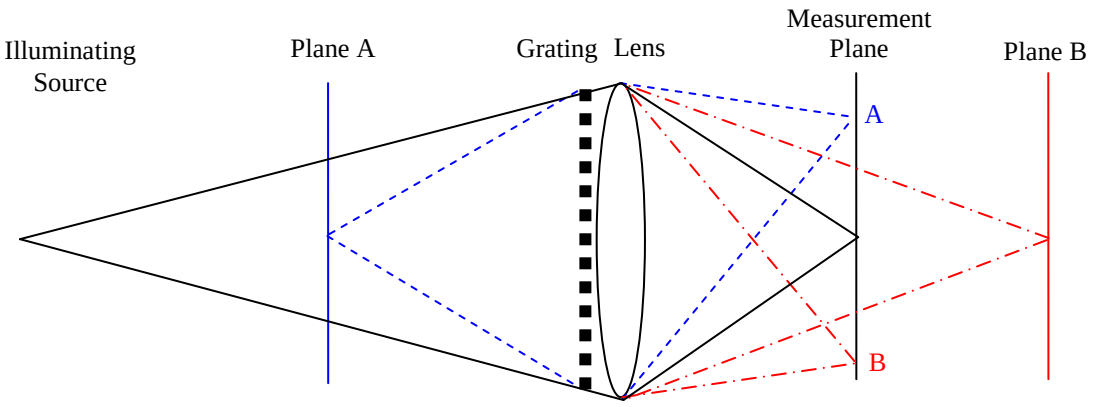


Figure 4.2 Schematic of the experimental setup. The image of plane B is inverted.

The boundary of the disk, in pixels, is positioned by iterating the value of the radius to give the largest change in intensity between the current radius value and current radius value plus one. This is achieved by looking at the difference between the sum of the sample values at the current radius value and the sum of the sample values at the radius value plus one normalised by the number of samples at each radius. The radius is then varied to find the value that produces the greatest difference. The nearest neighbour method, as described in section 3.3.3, is used to define the data at a specific radius.

Having specified the centre and the radius for each image, the images can be centred within a grid and the position of the boundary specified using either the radius from the measurement in the reconstruction plane or, alternatively the average value from the two defocused images depending on the available data. The consequences of incorrectly positioning the boundary and misaligning the data are considered in section 5.3. This pre-processing of the data is required in all cases when using the ITE to calculate the phase.

After pre-processing the data it is analysed using the Fourier transform method described in section 3.2.6.

For comparison with the coefficients calculated using the Fourier transform method, defocus coefficient were calculated by considering a spherical wave, with a radius R, emanating from a point and illuminating an aperture of radius r_0 as shown in Figure 4.3.

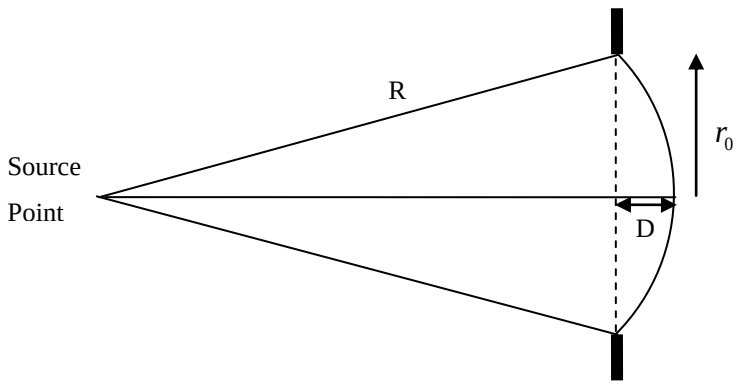


Figure 4.3 Schematic of the model to derive the defocus coefficients

The wavefront sag D approximately relates to the deviation of the wave from a plane wave as the radius of the aperture is much smaller than distance between the aperture and the light source. The distance also relates to the coefficient of a defocus mode and can be calculated from the physical properties of the system using

$$R^2 = r_0^2 + \left(R - D\right)^2 . \tag{4.1}$$

Assuming that D is small so that D^2 is negligible then

$$D \approx \frac{r_0^2}{2R}. \tag{4.2}$$

The change in defocus with respect to the change in the source position is determined by taking the derivative of D with respect to R so that the change of the defocus coefficient, Angarita-Jaimes, *et al.* [57], is approximately

$$\Delta D \approx -\frac{r_0^2 \Delta R}{2R^2} \tag{4.3}$$

where ΔR is the translation of the illuminating source, from the collimated point and ΔD is the change in wavefront sag in metres. The coefficient of the defocus mode, scaled to give the coefficient in radians is approximately

$$\Delta D \approx -\frac{\pi r_0^2 \Delta R}{2\lambda R^2}. \tag{4.4}$$

The defocus coefficients calculated from the experimental data using the Fourier transform method and using equation (4.4) are plotted against the source translation distance as shown in Figure 4.4.

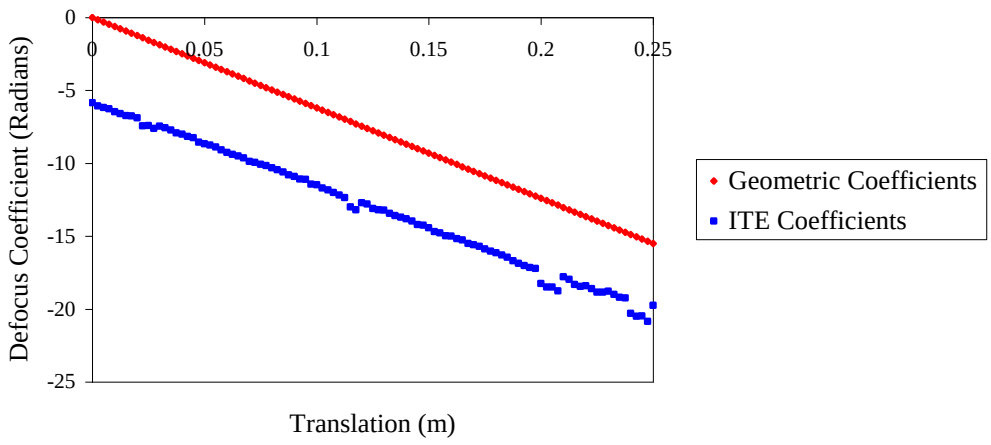


Figure 4.4 Defocus coefficients plotted as a function of the position of the light source. The coefficients calculated using the ITE are marked in blue and the coefficients calculated using the dimensions of the system are marked in red.

The figure shows the gradient of the two lines are similar with a difference of approximately 6%. This shows that the relative change in defocus calculated from the two methods is consistent. The constant offset is likely to be due to errors in the calibration of the experiment, when determining the position that the beam is exactly collimated and positioning the other components. By visually inspecting the data, it is not clear that the translation of the source has reached the collimation point. Therefore, it is difficult to match the absolute values of the defocus coefficients. The combination of the off-set and the difference in gradient is equivalent to an error of 9.8 cm when positioning the collimation point. Additionally, the 6% difference in the gradient is equivalent to an error of approximately 6 cm when measuring R. These are large errors so it is likely the off-set is a cumulative error caused by misplacing a number of components.

4.4. AMO / Wavefront Sciences Inc Experiment

As part of a project with AMO/Wavefront Sciences Inc to develop a phase diversity wavefront sensors for ophthalmic applications, experiments were conducted as part of the process of designing and calibrating the sensor. This provided data could be used to test the accuracy of the calculated phase. This is of interest as the phase encodes the aberration in a patient's eye so this gives a method for calculating the aberrations.

AMO designed and constructed a wavefront sensor using a distorted diffraction grating to simultaneously capture the three measurements, Blanchard, *et al.* [41], Blanchard, *et al.* [54].

Figure 4.5 shows the experimental setup to measure the set of images that were used to determine the applied phase change. In the figure there is a light source placed to the left of the aperture. The light illuminates the aperture and then propagates through the two lenses and the grating to the camera where the measurement is made. In this configuration, the three images are measured on the camera, each is of a different plane along the propagation axis, as indicated by the coloured arrows in the figure.

The system was designed to be telecentric, Djidel, *et al.* [55]. This means that the magnification in each diffraction order is the same, and requires that the distance between the lens on the right and the grating equals the focal length of the lens.

The lens on the left had focal length of 100 mm and the lens on the right had a focal length of 300 mm. The grating had a focal length of 508 mm and a central period of 11 μm. The spacing between the measurement planes was approximately 20 mm, the wavelength of the illuminating source is 845.5 ± 27 nm and the camera pixels are 11.7 μm square. The experiment was designed and conducted by AMO / Wavefront Sciences Inc.

To apply the phase change to the wavefront the light source is moved from the collimated position along the optic axis. When the light is collimated, there is no change in the images between the measurement planes as the phase is constant. As the light source is translated, it creates a converging or diverging wavefront.

Figure 4.6 shows an example of the experimental data, there is some chromatic blurring in the data, as can be seen in the example data. This is due to the band width of the illuminating source and the grating.

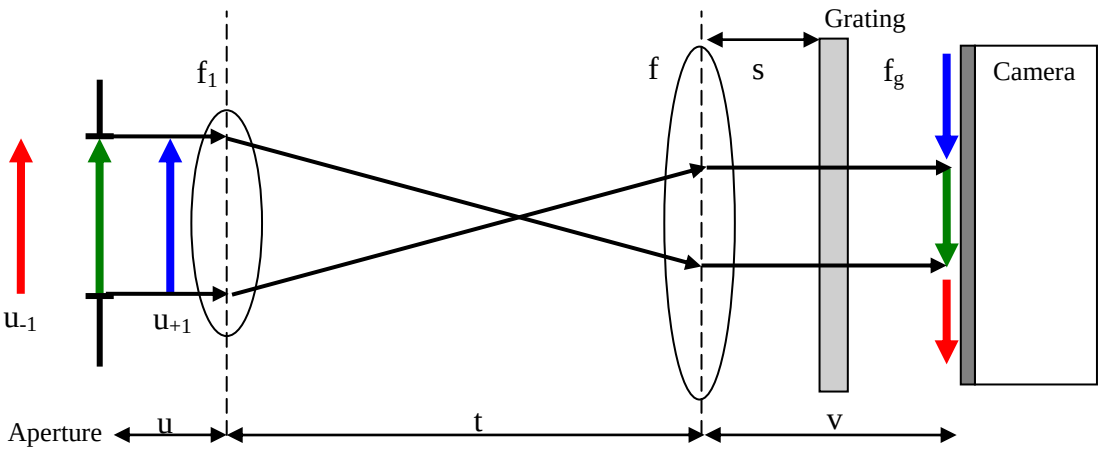


Figure 4.5 Experimental setup (Figure supplied by AMO / Wavefront Sciences Inc)

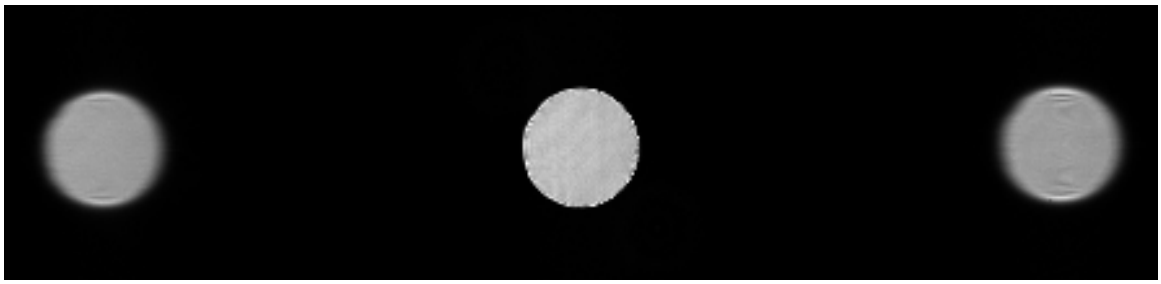


Figure 4.6 An example of the experimental data provided by AMO / Wavefront Sciences Inc

The source was a super luminous diode with a band width of $\pm 27\,nm$. The angle separating each diffraction order from the centre is dependent on the wavelength of the light. So the images are blurred as the light is not a single wavelength. The consequence for the data analysis is minimal. Although the data has been smoothed, the phase change applied to the beam of light is mainly defocus. This is a low order aberration and does not require fine detail of the images.

Sixteen images from AMO were used to calculate the phase. For each set of data, the images on the left and the right are used to calculate the derivative of the intensity and the central part of the image is used for normalisation. The images are analysed using the Fourier transform method, described in Chapter 3, for each position of the light source.

After calculating the phase, the coefficient of the defocus mode is determined, by fitting the defocus mode to the calculated phase. The coefficients are plotted as a function of the source translation distance, as shown in Figure 4.7.

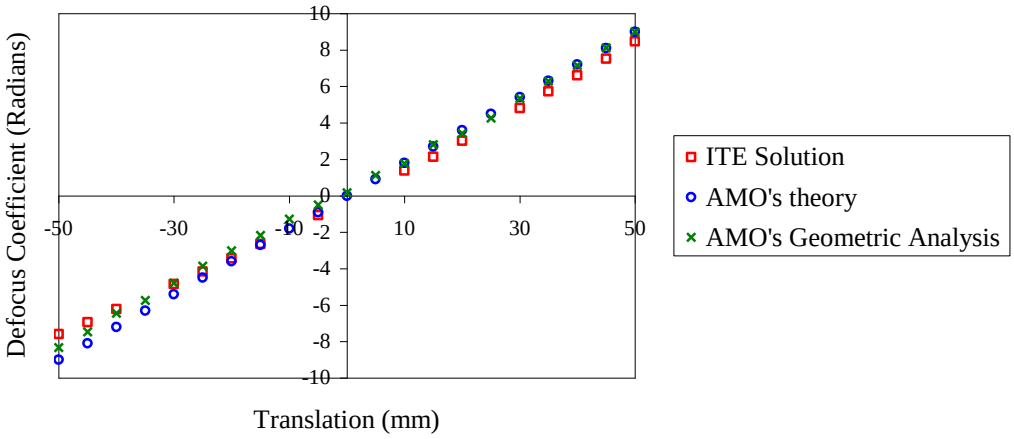


Figure 4.7 Experimental results from a set of data provided by AMO / Wavefront Sciences Inc. The calculated coefficients are plotted as a function of the light source position.

Figure 4.7 shows the coefficients of the defocus mode calculated from the data using the Fourier transform method along with the estimated coefficients supplied by AMO. The theoretical coefficients provided by AMO are estimated from physical properties of the system and the coefficient from the geometric analysis are calculated using the radius of the disk calculated from each of the images. None of the coefficients are therefore known to be exactly correct as all of them have been calculated.

The missing data points in Figure 4.7 are due to receiving data from AMO where two out of the three images were the same. These sets of data were therefore excluded from the data analysed using the ITE.

The calculated phase coefficients are compared to geometric analysis of the data and a theoretical curve both provided by AMO/Wavefront Sciences Inc. All the results in Figure 4.7 are in good agreement, which indicates that the solutions to the ITE can be applied to experimental data as well as simulated data, at least for the defocus mode.

The phase change is not directly set. This makes it difficult to establish which if any of the lines are correct and the deviation from the true phase, but as all the lines agree and were calculated using different methods, the results are at least consistent with each other. This suggests that the Fourier transform method is a practical approach to calculating the phase as well as a theoretical method.

4.5. Conclusions

This chapter has shown that the theory developed in the last chapter can be applied to experimental data and the relative defocus phase can be accurately tracked despite the assumptions and limitations of the model.

The first experiment showed that the relative change in defocus could be tracked but there was an off set in the absolute value of the calculated defocus. This is likely to be due to experimental error, when calibrating the setup. Without detailed knowledge of the imaging system calibration errors can easily be introduced. Therefore, the phase change applied experimentally and the estimated phase change, based on the geometry of the experiment can be different so this phase cannot then be compared to the phase calculated using the ITE.

The second experiment showed the result of the phase calculated using the Fourier transform method closely matched the theoretical line and the geometric analysis both of which were provided by AMO. Although the phase was unknown, the results of each method are consistent and calculated using different and independent methods. This suggests that this method can be applied in practical situations to experimental data as well as providing an exact theoretical model.

Having shown that the analysis from the previous chapters can be applied to experimental data, at least for one simple mode, the following chapter discuses the potential for optimising and improving the experimental setup and the numerical processing to reduce the errors in the calculated phase.

Chapter 5 – Improvements and Limitation

5.1. Introduction

In Chapter 3, exact solutions were derived for the ITE and as shown in Chapter 4 these can be applied to experimental data to give an estimate of the phase that is consistent with expectations. Experimentally and numerically, there are a number of factors common to any experiment of this type that can affect the calculated phase. This chapter considers the effects of these parameters and shows how an experiment can be improved or optimised to increase the accuracy of the calculated phase and reduce unnecessary errors.

This chapter is divided into two parts. The first looks at common experimental parameters that can be altered, which affect the calculated phase, for example the plane separation and the resolution of the images. The second part looks at the numerical data processing, so accuracy of aligning the images is considered along with positioning the boundary.

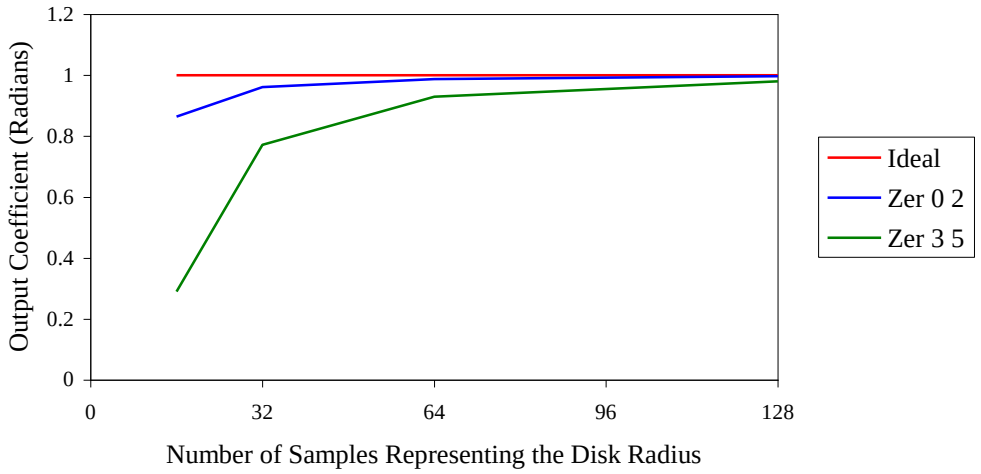
This chapter does not attempt to look at experiment specific factors such as noise or detector properties as this is outside the scope of this thesis.

5.2. Experimental Optimisation

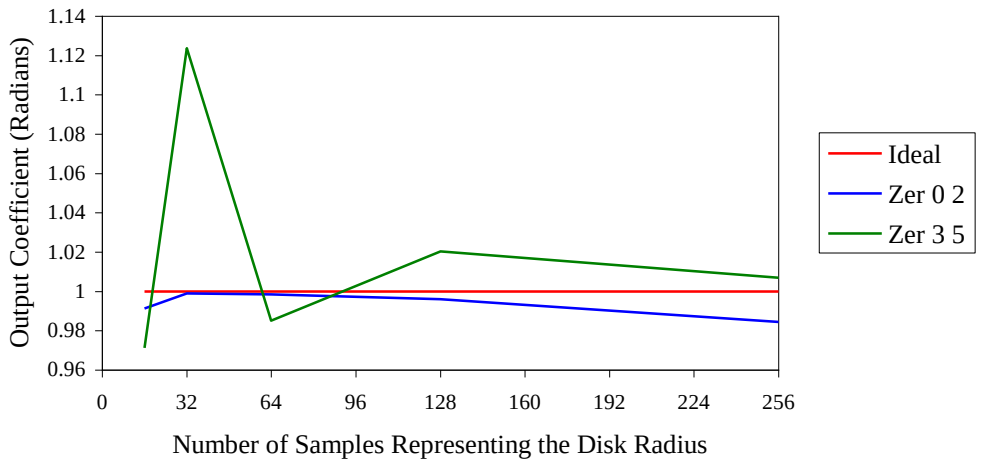
This section looks at generic factors common to any experiment where the phase is to be calculated from two or more defocused measurements. The factors considered are common to all phase diversity experiments of this type. In the process, all other factors are idealised. This rules out application specific factors like noise or situations where there is limited light.

5.2.1. Resolution

One of the factors that will affect the calculated phase is the resolution or number of sample points that represent the data. Experimentally there is a choice for the size of the images, as they can be magnified so they are bigger on the detector so that the sampling is on a finer scale. This will depend on the specific experiment, for example, the physical size of the camera will limit the quantity of data that can be measured and light levels will affect the extent to which the intensity per pixel can be reduced.



a)



b)

Figure 5.1 Output coefficient plotted as a function of the number of sample in the radius of the circle. a) shows the coefficients calculated using the Green's function and b) shows the coefficients calculated using the Fourier transform method.

To show the improvement to the calculated phase by increasing the resolution two examples are shown with the phase defined as Z_5^3 and Z_2^0. To generate the data the radius of the disk is set at 1 mm, the wavelength is 633 nm and the derivative of the intensity is calculated using the full integral representation, equation (3.13). Using the exact derivative removes the finite difference approximation so the calculated phase will be the best possible approximation of the set phase.

Figure 5.1 shows the output coefficient as a function of the number of samples representing the circle radius. The phase is calculated using the Green's function and the Fourier transform method. The output coefficient is the coefficient of the set mode

calculated from the estimated phase. In each case, when generating the data and calculating the phase using the Green's function, the grid size is four times the number of samples representing the radius of the disk. When calculating the phase using the Fourier transform method the data is zero padded so the grid size is increased to 4096×4096. This increases the sampling in Fourier space, which improves the calculated phase. This will be shown in section 5.3.3.

Figure 5.1 a) shows the output coefficient from the phase calculated using the Green's function. The results show that increasing the sampling improves the estimate of the coefficient for both examples. This is due to the improvement in the Riemann sum that approximates the integral used to calculate the phase, (3.31).

Figure 5.1 b) shows the output coefficient from the phase calculated using the Fourier transform method. The figure shows that when the phase is Z_5^3 and the sampling is increased the estimated coefficient appears to oscillate around the correct value but overall tends towards the set value as the resolution increases. When the phase is Z_2^0 there is an initial improvement in the estimate of the coefficient as the number of samples increases but then there is a deterioration of the estimate. The initial improvement is due to increasing in sampling of the data and the deterioration is due to the relative reduction in sampling in Fourier space. This is shown in section 5.3.3, which shows that increasing the sampling in Fourier space improves the phase estimate for cylindrically symmetric modes, Z_n^0. These results are typical of the effect of increasing the resolution of the images.

Figure 5.2 shows the same setup as Figure 5.1 a). It shows the same effects as before, increasing the resolution improves the phase estimates. Additionally, it shows that more samples are required as the order of the mode that is to be estimated increases. Therefore, the number of samples should be as high as possible to give the most accurate estimation of the phase.

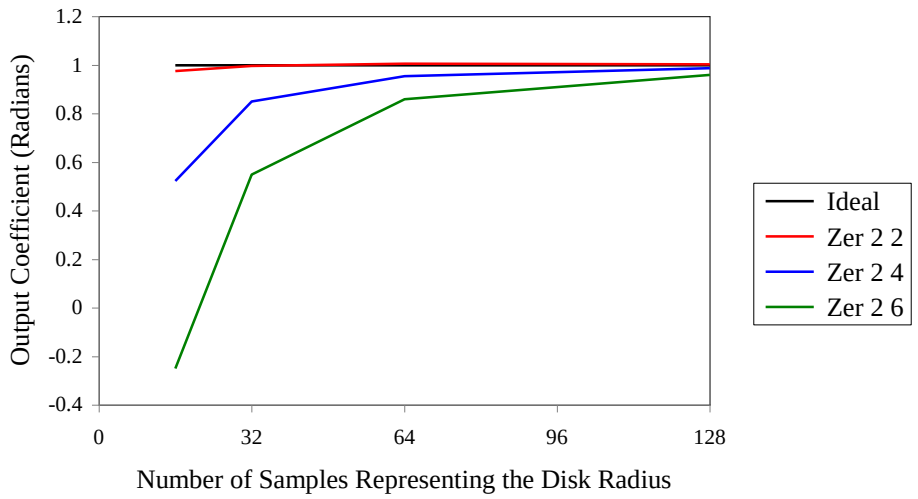


Figure 5.2 Output coefficient plotted as a function of the number of sample in the radius of the circle. The coefficients are calculated using the Green's function and the set phase is defined by Zernike polynomials with increasing order.

5.2.2. Approximating the Intensity, $I(\mathbf{x})$

When making the measurements there is a choice in the number of measurements that can be made. The two defocused measurements are required to approximate the derivative of the intensity and a third measurement can be made in the plane where the phase is to be calculated. This third measurement is used for normalisation, (3.22). In the analysis the intensity, in the plane where the phase is to be calculated, is assumed to be constant. Therefore, it may be unnecessary to make a third measurement as the intensity can be calculated from the other images. This would be beneficial in situation where there is limited light.

The intensity can be approximated using the average intensity value from the other two images so that

$$I_A(\mathbf{x}) \approx \frac{1}{(N+M)}\left(\sum_{n=1}^{N} I(\mathbf{x}_n, \Delta z) + \sum_{m=1}^{M} I(\mathbf{x}_m, -\Delta z)\right) \tag{5.1}$$

where the first sum is of the N sampled intensity values $I(\mathbf{x}_n, \Delta z)$ at position $\mathbf{x}_n$ and similarly the second sum is of the M sampled points, $I(\mathbf{x}_m, -\Delta z)$ at position $\mathbf{x}_m$. The sampled intensity values are the points considered as valid data.

The second approximation of the intensity $I_P(\mathbf{x})$ is a point by point average of the two defocused images,

$$I_P(\mathbf{x}) \approx \frac{I_+(\mathbf{x}) + I_-(\mathbf{x})}{2}. \tag{5.2}$$

The third method is to measure the intensity directly.

The point-by-point average will depend on the phase. The intensity distribution will change depending on the phase and the phase coefficient. This will only be accurate if phase change is small so that the estimated intensity is approximately constant.

In this example, the input phase is defined as Z_2^0 with a coefficient of 5 radians, the radius is set at 1 mm represented by 85 samples in an array of 512×512 elements. The wavelength is 845.5 nm and the plane spacing is 20 mm. These parameters are used as they match the experimental data in section 4.4.

Figure 5.3 shows the RMS error between the calculated and set phase as a function of the input phase coefficient. It shows that as the input coefficient increases the error increases for each method of normalisation. The relative difference in error is due to the normalisation. Using the point by point average for normalisation, (5.2), produces the largest error in the phase calculation, as each point is normalised differently imposing structural variation across the aperture that will be interpreted as an additional variation in the phase.

The average intensity, (5.1), will have a small error as the data is not contained in a well-defined disk so part of the signal will leak outside of the disk and not be counted, resulting in an under estimation of the intensity and an increased error for the average intensity which introduces a systematic error into the calculated phase. There is also the issue of where the boundary of the disk should be placed which will change the values for N and M. Nevertheless, provided the position of the disk is relatively accurate then the estimation of the intensity will be relatively accurate and the normalisation using the average value and the third image should produce similar results.

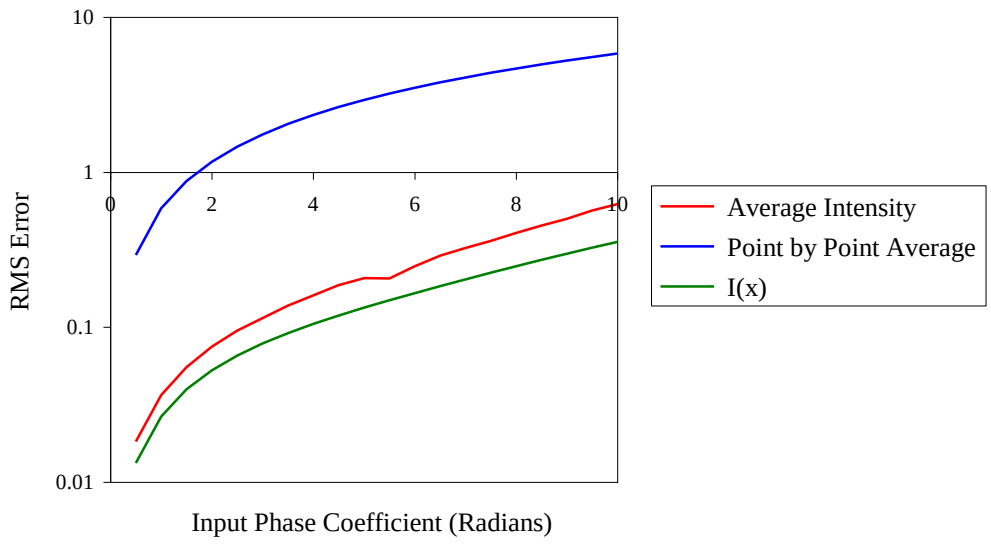


Figure 5.3 RMS error between the set and calculated phase plotted as a function of the input phase coefficients. Three different normalisation methods are used in the process of calculating the phase; the average intensity, I_A, the point by point average, $I_P(\mathbf{x})$ and the third image, $I(\mathbf{x})$.

When deciding on the number of image to use, this figure and other tests suggest that the third image should be included as this gives the smallest error in the calculation. In some cases, there may be a particular reason to reduce the number of images, for example limited light as the addition of the third image will reduce increase the noise in the other two images used to calculate the derivative. When there are only two images Roddier [3], Acosta, *et al.* [58] suggested that the point by point average should be used but clearly the average intensity value should be used as this reduces the error by approximately one order of magnitude.

5.2.3. Improving the Approximation of the Axial Derivative

In an experiment, the exact derivative of the intensity is unavailable, so a finite difference approximation will need to be used. This will introduce a source of error into the calculated phase. Two experimental factors can be controlled to reduce this error; the number of images and the spacing between the measurement planes

So far, the measurements have been made on two planes, which are used to approximate the derivative. This process can be expanded to include more measurements but will be limited in a specific experiment by the available light levels and the signal to noise ratio.

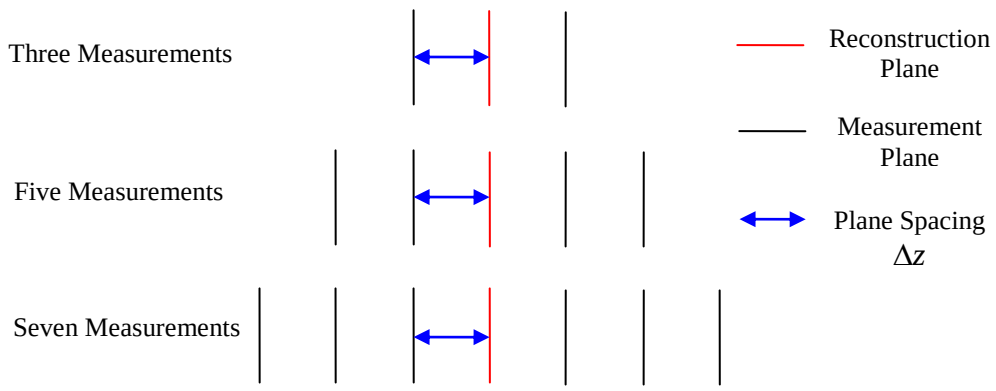


Figure 5.4 Schematic of the simulation setup showing the relative position of the measurement planes to the plane where the phase is estimated

This section considers the improvement to the calculated phase by including more measurements and varying the plane spacing. The three measurement system, the top line in Figure 5.4, is simplest setup considered up to this section but the derivative can be approximated using more measurement to improve the estimation. To approximate the derivative only the measurements in the planes coloured black in Figure 5.4 are required. The measurement in the central plane is for normalisation.

Writing the intensity using a Taylor series expansion, then

$$I\left(\mathbf{x}, z+n\Delta z\right)=\sum_{n=0}\frac{\partial^n I\left(\mathbf{x}, z\right)}{\partial z^n}\frac{\left(n\Delta z\right)^n}{n!}=I\left(\mathbf{x}, z\right)+\Delta z\frac{\partial I\left(\mathbf{x}, z\right)}{\partial z}+\sum_{n=2}\frac{\partial^n I\left(\mathbf{x}, z\right)}{\partial z^n}\frac{\left(n\Delta z\right)^n}{n!}. \tag{5.3}$$

If two measurements are made on the planes $z+\Delta z$ and $z-\Delta z$ then the intensity derivative, on the left side of (3.21), can be approximated using

$$\frac{I\left(\mathbf{x}, z+\Delta z\right)-I\left(\mathbf{x}, z-\Delta z\right)}{2\Delta z}=\frac{\partial I\left(\mathbf{x}, z\right)}{\partial z}+\varepsilon_2. \tag{5.4}$$

with a residual error

$$\varepsilon_2\left(\mathbf{x}, z\right)=\sum_{n=0}^{\infty}\frac{\left(\Delta z\right)^{2n+2}}{\left(2n+3\right)!}\frac{\partial I^{\left(2n+3\right)}\left(\mathbf{x}, z\right)}{\partial z^{\left(2n+3\right)}}. \tag{5.5}$$

In a similar way to using two measurements, the derivative using four measurements can be approximated by,

$$\frac{I\left(\mathbf{x}, z-2\Delta z\right)-8I\left(\mathbf{x}, z-\Delta z\right)+8I\left(\mathbf{x}, z+\Delta z\right)-I\left(\mathbf{x}, z+2\Delta z\right)}{12\Delta z}=\frac{\partial I\left(\mathbf{x}, z\right)}{\partial \mathbf{x}}+\varepsilon_4, \tag{5.6}$$

where the error term is

$$\varepsilon_4(\mathbf{x},z)=\frac{1}{6}\sum_{n=0}^{\infty}\frac{(\Delta z)^{2n+4}}{(2n+5)!}\frac{\partial I^{(2n+5)}(\mathbf{x},z)}{\partial z^{(2n+5)}}\left(2^{2n+5}-8\right). \tag{5.7}$$

This process can be extended to include as many measurements as required. For each higher order approximation of the derivative two extra measurements are required.

There are two ways to increase the number of images. The extra images can be recorded closer to the central plane so that the distance between the outer measurements is unchanged or extra measurements can be made further from the central plane so the spacing between the measurements is the same.

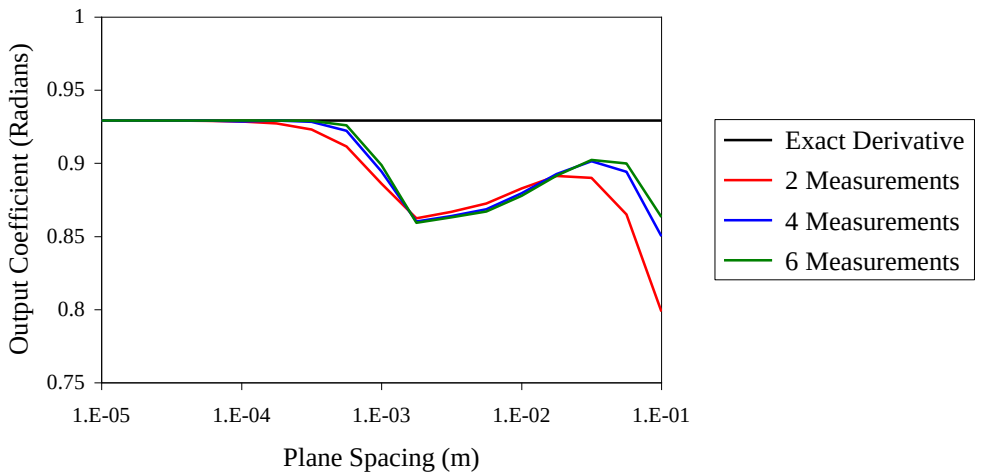
When increasing the number of images the pair closest to the central plane have the largest contribution to the approximation of the derivative. This is shown in equation (5.6) where the inner images have a contribution to derivative eight times greater than the outer images. Therefore, to compare the approximation of the derivative the extra measurements should be added so that the plane spacing remains the same. This will show any improvement from the additional images.

Experimentally, if the measurements are made simultaneously and the total energy is split between each one then adding additional measurements will increase the noise in each measurement. Therefore, as the innermost images have a higher weighting the accuracy of the derivative will be reduced whenever noise is significant.

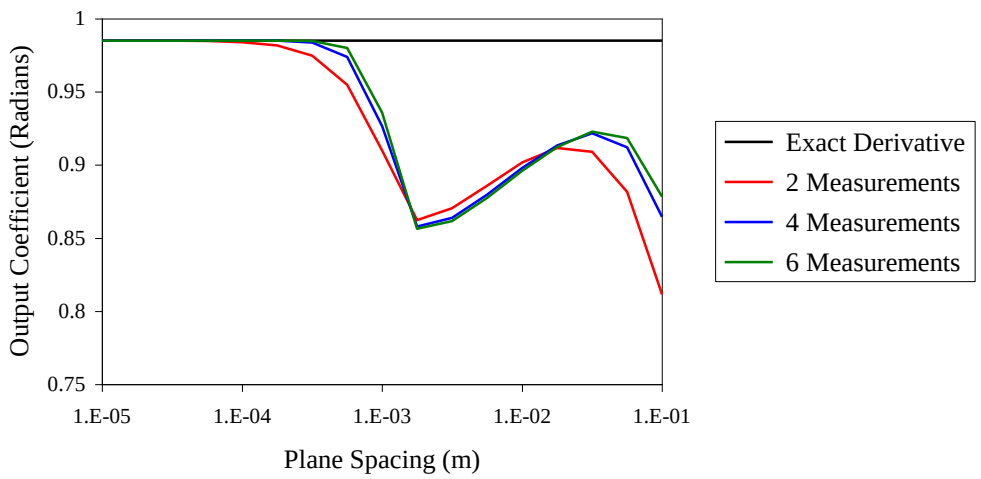
In these simulations the radius of the pupil is 1 mm and represented by 64 samples within a grid of 256×256 samples and the wavelength is 633 nm.

Figure 5.5 shows the output phase coefficient plotted as a function of the plane spacing, for each combination of images as shown in Figure 5.4. The phase is also calculated from the exact derivative using equation (3.13).

Figure 5.5 shows that, using both the Green's function and Fourier transform methods, as the plane spacing increases the estimation of the phase degrades for each approximation of the derivative. The difference between the phases calculated using the two methods is a result of the numerical implementation and depends on the number of sample points.



a)



b)

Figure 5.5 Output phase coefficient as a function of the plane spacing Δz. a) shows the phase calculated using the Green's function and b) shows the phase calculated using the Fourier transform method. The exact axial derivative and the derivative approximated using different numbers of measurements are used as input data and the set phases is Z_5^3

Overall, there is only a small improvement to the calculated phase by increasing the number of images as the inner most images dominate the approximation of the derivative and the contribution from extra images becomes increasingly small. Therefore, the plane spacing between the inner most images is a key factor effecting the calculated phase and the addition of extra images only causes a small improvement so unless the signal to noise ratio is very good only two measurements should be used.

The plane spacing should therefore be chosen to give a good estimation of the derivative. In the ideal noise free case, the measurements would clearly be made close together to get an accurate approximation of the derivative. Although, not considered

here, with the addition of noise, the measurements need to be far enough apart so that the signal dominates the measurements giving a good signal to noise ratio, Paganin, *et al.* [59]. Therefore, there will be a lower limit on the plane separation so that the signal dominates the data. Below this separation, the noise dominates and no useful information can be calculated.

Some of the results in this section have been reached independently as part of the work by Waller, *et al.* [31] which was published in May 2010.

5.2.4. Asymmetric Measurement Planes

After deciding on the number of measurements to use, and there relative position to the plane where the phase is to be calculate there can still be errors in the exact position where the measurement are made. When analysing the data to reconstruct the phase, the model assumes that the measurements, to approximate the derivative, are equidistant from the reconstruction plane. This section considers the sensitivity of the calculated phase due to an error in the placement of the measurement planes, relative to the plane where the phase is to be calculated.

Figure 5.6 shows the position of the reconstruction plane in red and the assumed position of the measurement planes, in blue. In this simulation, there is a positioning error so the data is generated in the planes marked in green but the data is analysed as if it was measured in the correct planes. The measurement planes are moved over a range of -10 to 10 mm. In each case, the output coefficient is plotted as a function of the input phase coefficient.

In this example, the radius of the disk is 1 mm represented by 85 samples in a grid of 512×512 samples. The wavelength is 845.5 nm and the plane spacing is $20\,mm$. The phase is set as defocus, Z_2^0, and the coefficient is set at 5 radians. These parameters are used as they match the experimental data in section 4.4.

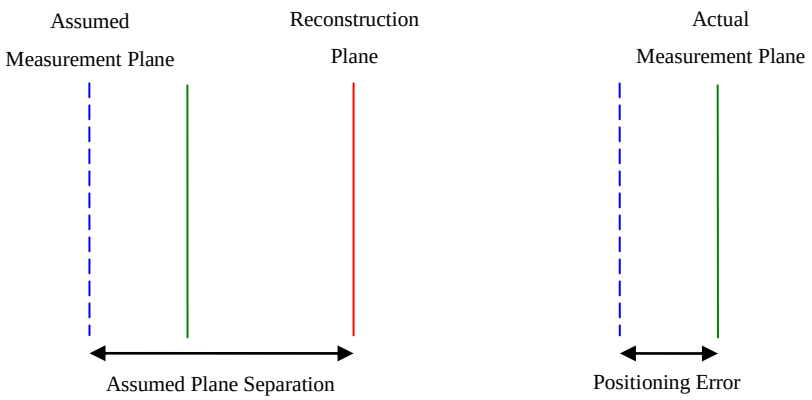


Figure 5.6 Schematic showing the relative position of the reconstruction plane to the assumed measurement plane marked by the blue dashed line and actual measurement planes in green.

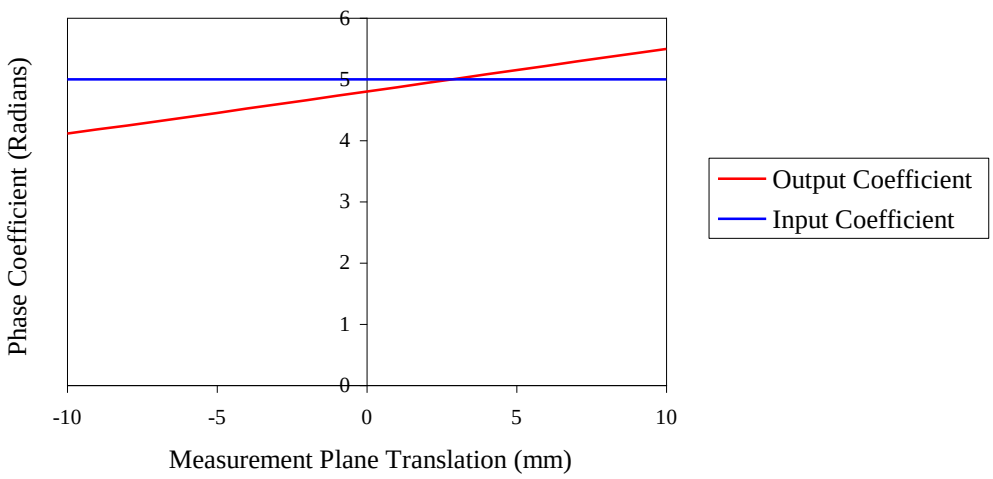


Figure 5.7 Input and output phase coefficients plotted as a function of the measurement plane translation.

Figure 5.7 shows the input and output phase coefficients as a function of the plane translation error. This figure shows that the output coefficient varies linearly as a function of the positioning error this is because moving the measurement planes in this way is equivalent to adding a defocus phase change to the phase estimate. The defocus phase change is approximately linearly related to the translation, Angarita-Jaimes, *et al.* [57], as described in Chapter 4.

When there is no error in the positioning of the measurement planes there is still an error in the calculated phase due to other approximation, for example the finite difference approximation of the intensity derivative and the finite number of sample points.

When the positioning error in 2.5 *mm* the output coefficient matches the input coefficient, this is coincidence. Changing the input phase coefficient will change the crossing point and changing the input phase will change this results.

5.2.5. Summary of Physical Options

This section has discussed the potential for optimising an experiment, to improve the calculated phase, by reducing the approximation from the experimental data. It has shown that increasing the resolution of the images can significantly improve the estimated phase. Increasing the size of the images will increase the processing time therefore any method that can reduce this time is beneficial. Increasing the number of samples will also increase Poisson noise.

There is also an improvement to the phase by measuring a third image in the plane where the phase is to be calculated. Using this image for normalisation in comparison to an average intensity value calculated from the defocused images shows a small improvement and a significant improvement in comparison to using a point by point average. When the images are measured simultaneously, the signal to noise ratio changes depending on the number of measured images. Therefore, the choice between using the third image and the average intensity value depends on the signal to noise ratio for each image.

The improvement using extra images to estimate the derivative was shown to be minimal, as the approximation of the derivative is dominated by the images closest to the center. Overall, the benefit of increasing the number of images is minimal particularly when the measurements are made simultaneously which increases the level of noise. The key factor is the separation between the planes where the measurements are made.

Having considered experimental factors that can be optimized to improve the calculated phase, the following section considers factors in the numerical processing that can affect the phase.

5.3. Numerical Processing

In section 3.3.2, when describing the process for analysing the data, there were a number of steps, centring the images and defining the position of boundary, where there

errors can be introduced. As all the examples, so far, have been simulations, these issues are not important, as all the parameters are known. This section considers the effects these positioning errors introduce into the estimated phase along with factors that can influence the choice of method for calculating the phase, for example, the accuracy of the calculated phase and the computation time.

5.3.1. Incorrectly Choosing the Disk Radius

One of the problems of analysing experimental data is estimating the position of the boundary of the disk, particularly if the central image, has not been measured. When making measurements on different planes, the edge is no longer a well defined shape as diffraction and divergence cause the intensity distribution to change.

To show the effect of incorrectly placing the boundary, simulated intensity data, with a known radius, is created and then analysed varying the position of the boundary. This is shown in Figure 5.8, where the boundary of the aperture is approximated by the purple circle.

In the example, the correct radius is equivalent to 85 pixels and the phase coefficient is 6 radians. Figure 5.9 a) shows the result when the set phase is Z_5^3. At the correct radius of 85 pixels, there is an underestimation of the calculated phase coefficient by approximately 1 radian. This is due to other approximations, for example the finite difference approximation of the intensity derivative and the finite number of sample points. When the estimated radius is too small there is an over estimation of the coefficient, and when the radius is too large the coefficient is underestimated.

Figure 5.9 b) show the result when the phase is Z_{-2}^2. It shows an overall underestimation of the coefficient, which is more significant when the radius is too small as the mode is harmonic all the information is on the boundary, which is missed out. When the radius is overestimated, the boundary data is analysed as if it were curvature data giving a non zero phase estimate.

The figures suggest that a slight overestimation is preferable to an underestimation as the error is smaller but the true radius should be estimated as accurately as possible.

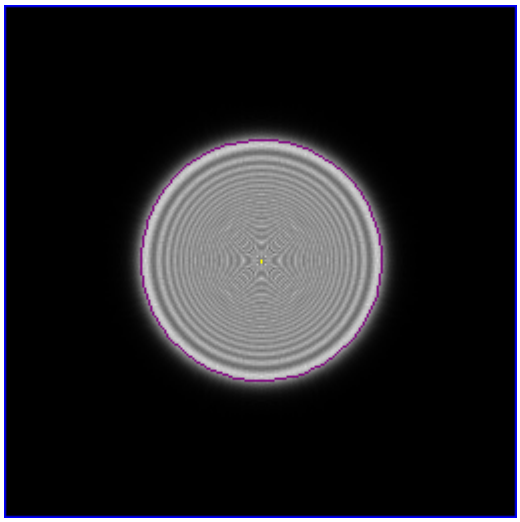
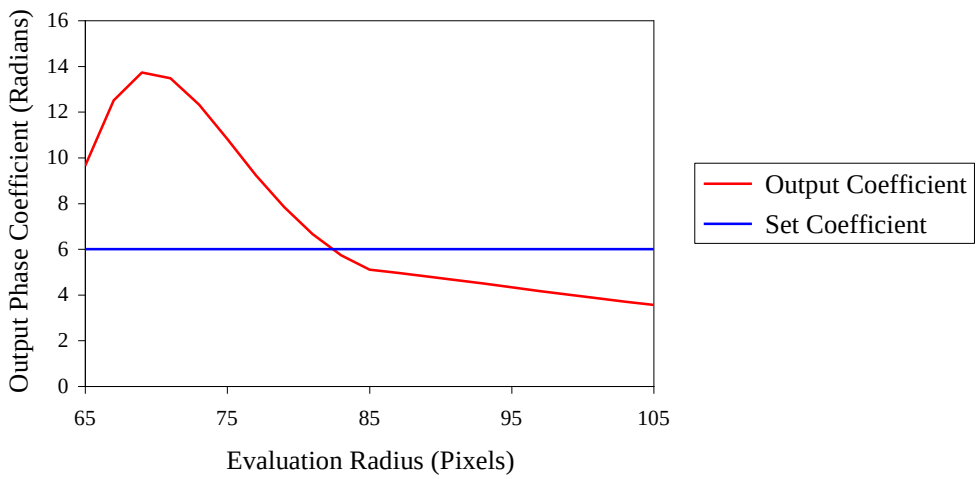
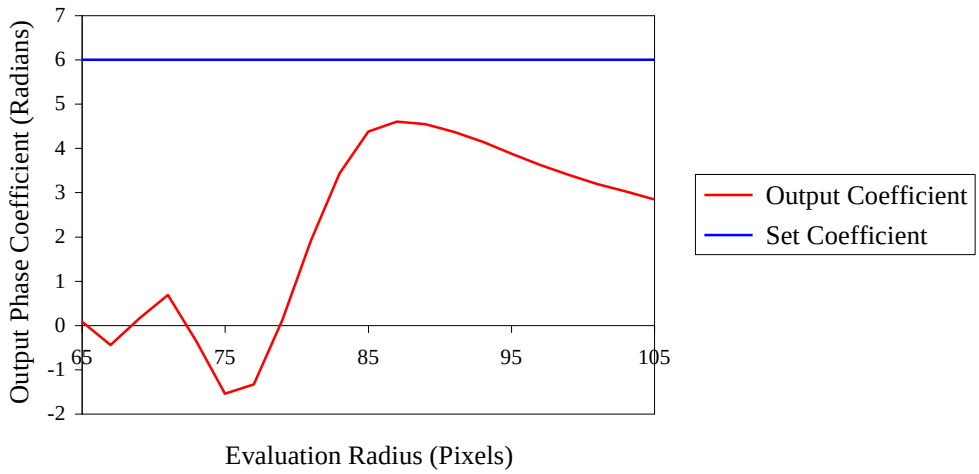


Figure 5.8 An example of intensity data, with an estimation of the boundary marked by the purple circle



a)



b)

Figure 5.9 a) shows the input and output phase coefficient plotted as a function of the evaluation radius when the set phase is Z_5^3. b) shows the input and output phase coefficient plotted as a function of the evaluation radius when the set phase is Z_2^{-2}. The correct radius is 85 pixels.

5.3.2. Misaligned Data

As well as positioning the boundary, as discussed in the previous section, the centre of the disk also needs to be determined. This section considers the effects on the calculated phase caused by misalignment of the images when being used to approximate the axial derivative.

In this example, the input phase is defined as Z_2^0 with a coefficient of 5 radians, the radius is set at 1 mm represented by 85 samples in an array of 512×512 elements. The wavelength is 845.5 nm and the plane spacing is 20 mm. Using these parameters, the intensity data is generated but before taking the difference between the two images, one of the simulated intensities is shifted along the x axis as shown in Figure 5.10. This figure shows a schematic of aligned images on the left. The red and blue circles represent the boundary of two images where the set phase is defocus, so the radius of the disk changes as the wavefront propagates. Therefore, in one image, the disk will get smaller and the other the disk will get larger. The misaligned data, on the right, shows the image with the red boundary translated to the right.

This misaligned data is then analysed and the output phase coefficients are then plotted as a function of the image displacement as shown in Figure 5.11.

Figure 5.11 shows that as the misalignment error is increased the estimated phase becomes poorer. Therefore, the images need to be accurately aligned for the best results but an error of a few pixels makes only a small difference. When there is no misalignment there is still an error, this is due to other approximation, for example, the finite difference approximation of the intensity derivative. As expected the modes with only information on the edge, for example Z_2^{-2}, result in poorer estimates than other modes as the data quickly becomes distorted. When the phase is Z_2^0 or Z_5^3 the larger quantity of curvature data is less affected by the misalignment. There is a small asymmetry when the phase is Z_5^3 due to the asymmetry with respect to the misalignment.

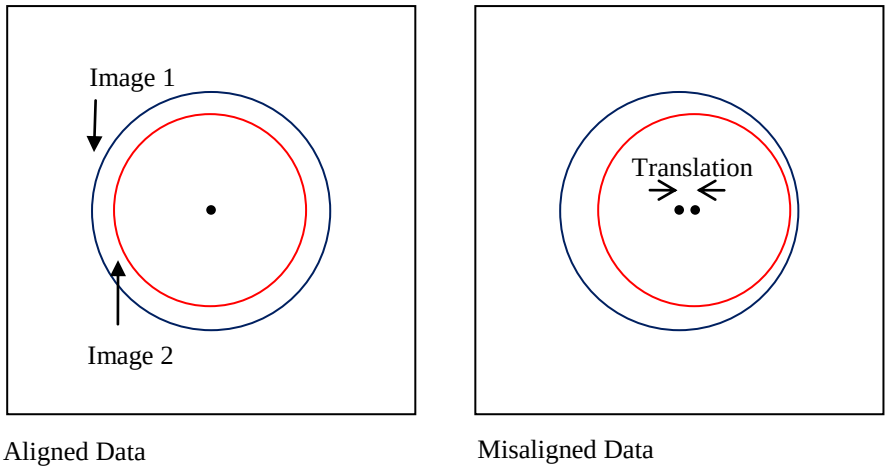


Figure 5.10 Illustration of misalignment of the intensity data

When the phase is defined as a tilt, a combination of Z_1^1 and Z_1^{-1}, then the images are translated across the camera. Therefore, misaligning the images should introduce a tilt into the calculated phase. This will not necessarily show up in the error metric, as the fit of the phase will largely only return the coefficient of a specific mode with no information about the other modes.

As an example of the effects of the misalignment, Figure 5.12 shows cross sections of the calculated phase for increasingly misaligned data. The figure shows that the calculated phase contains a tilt in addition to defocus. In this case removing the tilt produces an almost identical phase to the phase calculated from the aligned images. For larger misalignments, the error is no longer a pure tilt as the curvature and non-curvature data is now mixed up producing additional modes in the reconstructed solution. The previous error metrics showed only a small error from misaligning the images but as shown in Figure 5.12, a small misalignment introduces a substantial error into the calculated phase.

Therefore, the alignment of the data is critical for accurately calculating the phase as a slight misalignment of only pixel will introduce a substantial tilt into the phase estimate. Therefore, a process that accurately positions the data is required which accounts for the translation of the image caused by the presence of a tilt in the phase.

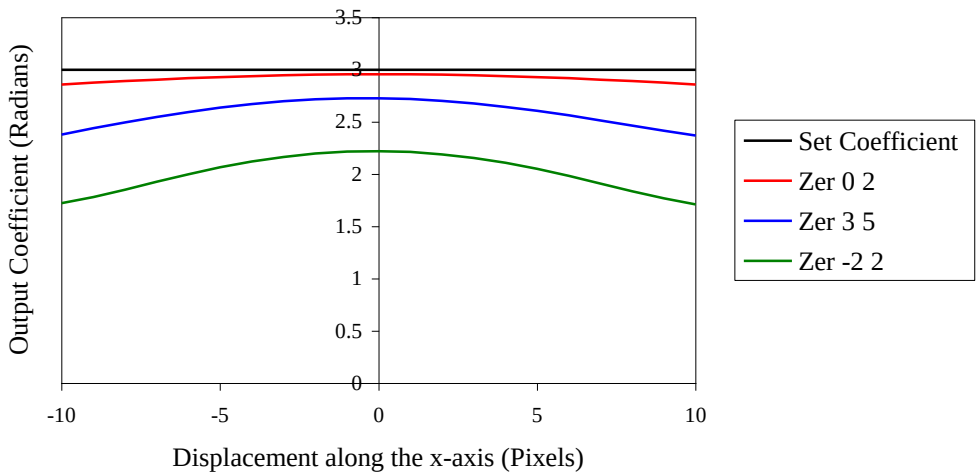


Figure 5.11 Output phase as a function of the displacement of one of the images. For different input phase distributions, Z_2^0, Z_5^3 and Z_2^{-2}.

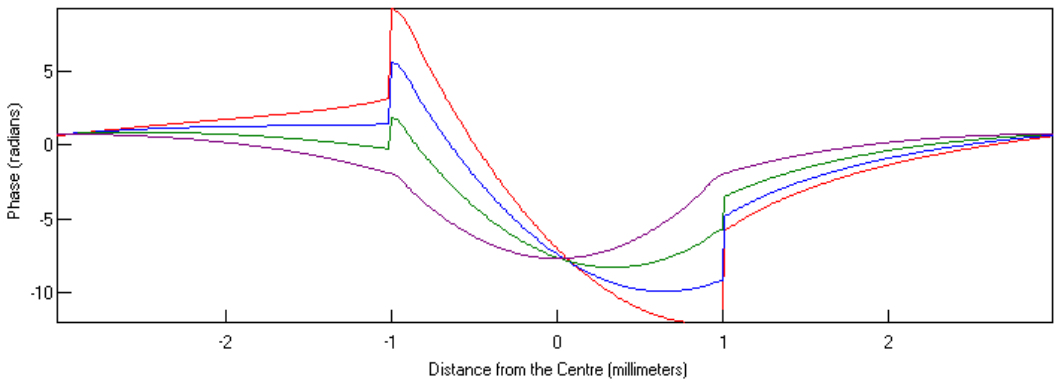


Figure 5.12 Example of the phase calculated from misaligned data. The example corresponds to the defocus case in Figure 5.11. Purple represent the phase calculated when the misalignment is zero pixels, green represent a misalignment of 2 pixels, blue represent a misalignment of 4 pixels and red represent a misalignment of 6 pixels

5.3.3. Sampling in Fourier Space

There are two parts to consider when considering the effect of the number of samples. For the Green's function method, the solution depends on the number of samples considered to be within the disk. The Fourier transform method benefits from a large number of samples within the disk but as the process involves Fourier transforms the data can be padded, by placing it within a larger grid. This has the effect of increasing the sampling in Fourier space. This is a factor that, if important, can be easily changed during post processing.

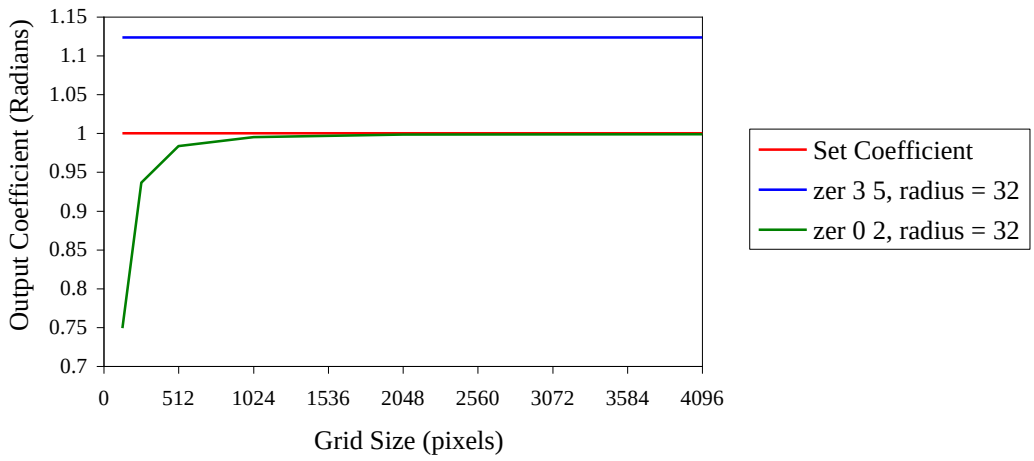


Figure 5.13 Output coefficient plotted as a function of the grid size used when calculating the phase using the Fourier transform method

To test the effect of padding the data, it is defined on a set grid size of 128×128 samples the radius of the disk is 1 mm represented by 32 samples and the exact derivative, equation (3.13), is used define the intensity derivative. When calculating the phase, the data was placed in a grid with size ranging from 128×128 to 4096×4096 samples.

Figure 5.13 shows the output coefficient plotted as a function of the grid size. For the set phase Z_2^0, increasing the number of samples in Fourier space improves the phase estimate. There is a significant improvement from making the grid size bigger but the improvement reduces as the grid size increases. Therefore, there will be a trade off between grid size and accuracy and computational expense. This is a typical example for all the cylindrically symmetric modes, Z_n^0.

The other example, when the phase is defined as Z_5^3 shows no improvement as the grid size is increased, this is the same for all the tested non cylindrically symmetric modes. The consistent error in this calculation, for such modes, is caused by other approximations, for example the finite number of samples in real space, which are not reduced by increasing the sampling in Fourier space.

5.3.4. Computation Time

One of the key factors that influences the choice of method to calculate the phase is the speed of the calculation. This section shows the time taken to calculate the phase from a set of intensity data at different resolutions.

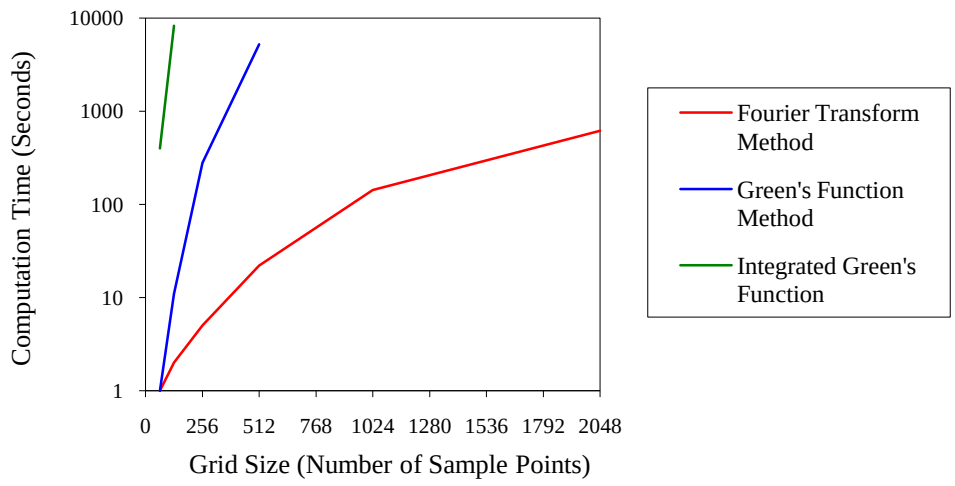


Figure 5.14 Time to complete the calculation of the phase from data generated at different resolutions. The time from the calculation (seconds) is plotted as a function of the grid size. The data is defined over a circle a quarter of the grid size.

The time taken to calculate the phase using the Green's function depends on the number of points in the disk and thus scales as N^4 where N is the number of points representing the radius. Whereas the time taken to use the Fourier transform method depend mainly on the grid size as this determines the time to calculate the Fourier transform of the data which scales as $N \log N$ where N is the number of points in the grid.

In this example, the radius of the disk is 1 mm the wavelength is 633 nm and the plane separation is 20 mm. The grid size is varied and the number of sample representing the radius is a quarter of the grid size.

Figure 5.14 show the time, in seconds, to calculate the phase from the intensity data as a function of the grid size. At each resolution, the same parameters are used to define the data. The only important parameter is the size of the disk that the data is defined over. In each case, the number of samples representing the radius of the disk is a quarter of the grid size.

Three methods are used to calculate the phase, the Fourier transform method and the two Green's function methods. The first sets the singular points to zero the second method uses the Green's function analytically integrated over the area of each pixel.

The figure shows that increasing the number of sample points increases the time for the calculation. The calculation time for even the smallest images, using the integrated Green's function, is at least two orders of magnitude larger than the two other methods. This fact combined with only the small increase in accuracy, as shown in section 3.3.4, makes the benefits of this method minimal.

The difference in time between the other two methods is also significant for all but the smallest images, with the time to complete the calculation increasing much faster using the Green's function than the Fourier transform method. It should be noted that the Green's function can be pre-calculated if flexibility is not required which may speed up the calculation time although the time for loading and saving all the Green's function values has to be considered.

Overall, as the two methods have similar accuracy the Fourier transform method is favourable from the point of view of computation time particularly if flexibility in the data format is required.

5.4. Conclusions

This chapter has considered the possibility of optimising the variable factors in an experiment and the numerical processing of the measured data. By optimising these factors, the error in the calculated phase can potentially be reduced. It should be noted that improving one factor is generally at the expense of another. The factors considered are generic to all experiments where the ITE is being used to calculate the phase.

This chapter has shown that the resolution of the image should be as high as possible even when low order modes are considered. The phase calculated using the Green's function and the Fourier transform method both improve with increased resolution. The Fourier transform method also requires high resolution Fourier space to improve the accuracy. The cost for increasing the resolution, to get better accuracy, is increased computation time; this is particularly severe when using the Green's function, which can be orders of magnitude slower that the Fourier transform method.

The chapter also showed that the improvement when using extra images on more planes to approximate the derivative is minimal and the key factor is the spacing between the

measurement planes. The spacing should be as close as possible but it will depend on the signal to noise ratio.

It has been shown that for the purpose of normalisation using a third image in the plane where the phase is to be calculated introduces a smaller error than approximating it using the defocused images. If only two images are measured, then the average intensity value should be used in the calculation rather that adding the two images together pixel by pixel and dividing by two. This avoids normalising each sample point differently that causes a variation across the data that is interpreted as a phase change.

The addition of extra measurements beyond the minimum requirement of two depends on the level of noise. When making the measurements simultaneously the signal to noise ratio, for each image, will be reduced depending on the number of images.

In term of the numerical processing, the chapter showed that positioning the radius of the disk is required to be accurate, but can tolerate an error of a couple of sample points as shown in Figure 5.9. Whereas, the alignment of the data is critical, as large error can be introduced, in the form of a tilt, from only a single pixel misalignment, this error gets worst for larger misalignments

The only factors that affect the choice of method to calculate the phase are accuracy and the calculation time. Overall, the Fourier transform method is preferable compared to the Green's function method. Both exactly solve the ITE and both have similar accuracy, although this depends on the resolution. However, a key factor is calculation time where the Fourier transform method is orders of magnitude faster than the Green's function.

Chapter 6 – Determining a 3D Object

6.1. Introduction

The calculation of the phase from sets of phase diverse data has been discussed in the previous three chapters, using sets of phase contrast images and defocused images. All these processes result in an estimate of the phase in a 2D plane.

There are applications where the phase change of the wave is a consequence of the light passing through a three dimensional object. For example, in biology, the phase of the illuminating light is modified when passing through an almost transparent cell. The phase change therefore relates to the structure of the cell, Popescu, *et al.* [34], Rylander, *et al.* [60], Sung, *et al.* [61].

This chapter considers, through a series of simulations, whether it is possible to reconstruct properties of the object by propagating the calculated wave backwards. The calculated object cannot be unique, as in general this requires a priori knowledge and more than a single view of the object, Devaney [62].

6.1.1. Literature Review – 3D Imaging

This problem of determining an object from its scattered field is found in many areas for various applications. In diffraction tomography many views of the object, over 360°, are used to reconstruct the object, Devaney [63]. This problem also appears in geophysics for determining underground structures, where clearly taking measurement all the way round the object is impossible, Devaney [64], but even with a limited number of views, information about the structure can be determined, Devaney [65]. The problem is also found in crystallography where prior knowledge, of the general periodic structure of the object, can be used to determine details, Millane [66].

Two commonly used methods for 3D imaging in biology use a confocal microscope, Figure 6.1, or a wide field microscope, Figure 6.2, Stephens, *et al.* [67]. These methods require preparation of the sample by introducing fluorescent dye. When illuminated the dye fluoresces and emits light at a different wavelength, some of which will travel back along the path of the incident light to the detector.

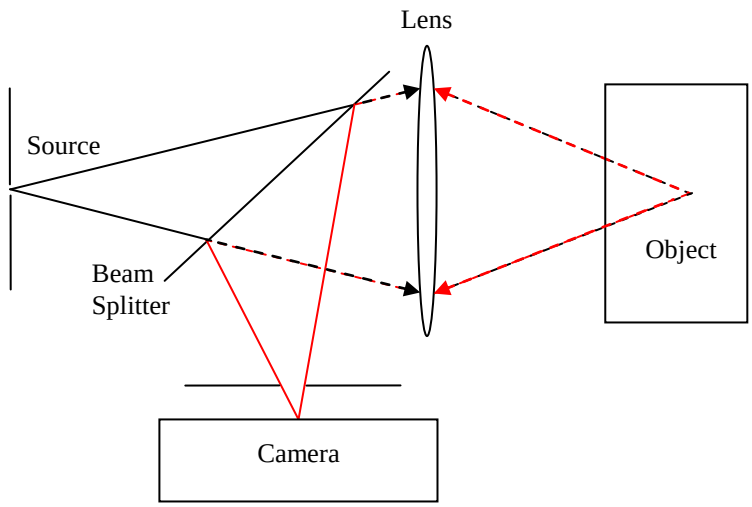


Figure 6.1 Schematic of a confocal microscope

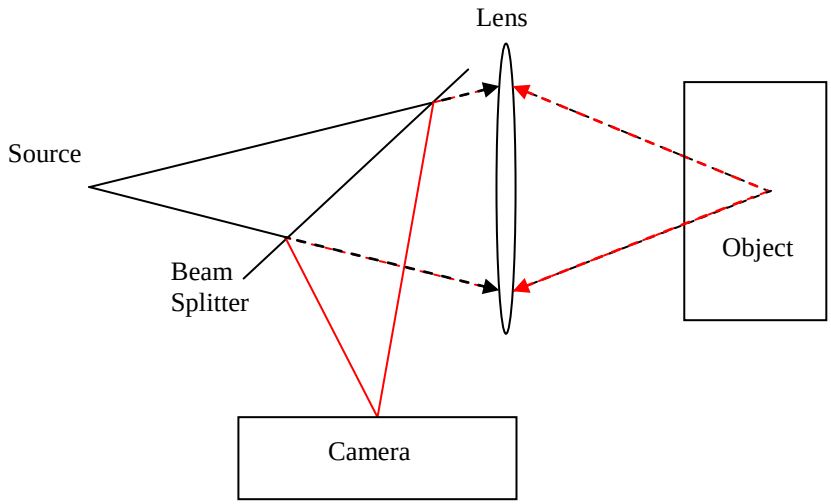


Figure 6.2 Schematic of a wide field epi-fluorescence microscope

The difference between these methods is that a confocal microscope has a pinhole after the source and before the camera; this excludes the out-of-focus light from the image. As both the source and detector pinholes need to be conjugate to get a high signal, this arrangement provides good depth discrimination. The wide field microscope illuminates the object and collects all the emitted light. Therefore parts of the images will be out-of-focus so further data processing is required to determine the object.

The advantage of a confocal microscope is that it produces high resolution 3D images, but because of the pinhole the object needs to be scanned to build up a 3D model of the object. Wide field microscopes do not exclude any light but the measurements will

require processing to obtain the 3D image. The disadvantage of these methods is the addition of the florescent dyes to the specimen. Specimens therefore need to be prepared before they can be viewed. Additionally, illumination of the dye can cause damage to the cell as free radicals are produced during fluorescence.

An alternative approach is to view an object along two orthogonal axis, Kikuchi, *et al.* [68]. In this configuration the poorer depth resolution from one view is compensated by the superior lateral resolution from the other view. The specimen can also be rotated around one axis while the detector remains stationary so that many views of the object can be obtained, Sharpe, *et al.* [69].

When viewing single cells or clusters of cells, an optical trap can be used to hold them. An optical trap uses a focused laser beam to hold the object within the beam. By trapping the cells they can then be rotated to produce multiple views of the object, Kreysing, *et al.* [70]. This process of trapping a specimen has been demonstrated Charrière, *et al.* [71], Charrière, *et al.* [72] where a hologram is generated for each view before analysing the data. The use of a hologram is only to determine the scattered field. Multiple views of the object are still required to calculate the object.

Although it may not be possible to view the object from all angles but it has been demonstrated that, even using a limited number of views, the structure of live cells can be reconstructed, Sung, *et al.* [61].

When viewing transparent objects with a microscope, the slice of the volume that is in focus i.e. giving the strongest signal can be changed by adjusting the focus of the microscope. To adjust the focus the position of the microscope optics is changed, this changes the slice viewed by the camera, as shown in Figure 6.3. Under some circumstances, the structure of an object can be determined by looking at the in-focus areas of the image, and adjusting the focus of the microscope is to scan through the volume containing the object, Streibl [32], Sheppard, *et al.* [73].

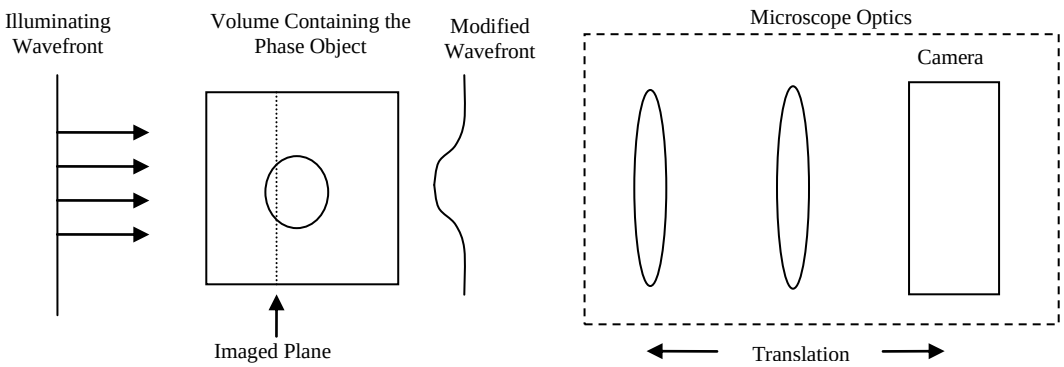


Figure 6.3 Schematic of a microscope, which is defocused to adjust the imaged plane

The work of Streibl [32], Sheppard, *et al.* [73] only considered the limitations of 3D imaging for various microscope configurations but a method for determining the object was not considered.

6.1.2. Aims and Objectives

In the previous chapters the phase and therefore complex amplitude have been calculated, in a single 2D plane, this is equivalent to determining the scattered field from a single view with a microscope. This chapter will look at the possibility of determining the structure of the object using this information.

There will be limitations to this process, as in general 3D imaging requires more information in the form of the scattered field from multiple views but even if the class of objects that can be determined is small, there may still be applications for this process.

This chapter considers to what extent the phase, calculated in a single plane, can be used to determine information about the 3D structure of an object by numerically propagating the wavefront backwards through the volume that contains the object, this is equivalent to adjusting the focus of a microscope and recording an image. Numerically propagating the wavefront back through the volume gives the intensity, equivalent to the information from a microscope, but also the phase information throughout the volume.

For those cases where the object can be seen by eye, an iterative scheme that works on the same principle using the in-focus points, is developed to calculate the structure of the object. The information to determine the object by this method is clearly limited,

therefore a method for increasing the number of views of the object is considered to obtain more information, to improve the accuracy of the structure of the object.

6.2. Model of an Object Viewed with a Microscope

This chapter looks at a method for determining the structure of a three-dimensional object, knowing the phase and intensity in a two dimensional plane. This section introduces the basic model of the process to view an object. The next stage then considers how to determine properties of the object from the data.

Figure 6.4 shows the specific modification of the wavefront mention in Chapter 2 and Chapter 3, where an object is placed in a volume and illuminated the object is transparent so only changes the phase of the wavefront. In the previous chapters, the modified wavefront is propagated through further optics to a camera, where measurements are made and the phase calculated. This chapter assumes that the phase can be calculated using a method described in a previous chapter or some other method and considers to what extent this information can be used to specify the structure of the object.

In the previous chapters, the phase is calculated in a plane after the object. After calculating the phase from the measured intensity, these can be combined to give the complex amplitude of the wave. This wave can be propagated backwards, through the specimen volume. From the back propagated wave the intensity can be calculated, equivalent to the information from refocusing a microscope. In addition, this method will provide the extra phase information.

When looking at a slice of the volume containing the back propagated data parts of the image will be in-focus and others out-of-focus, the out-of-focus parts of the image are due to the object having depth, so the parts of the object in different planes will be defocused relative to the plane being viewed.

This chapter starts with the assumption that under some conditions the in-focus parts of the image correspond to parts of the specimen and that the in-focus points can be separated from the out-of-focus points. The points considered to be in-focus have the largest absolute phase value

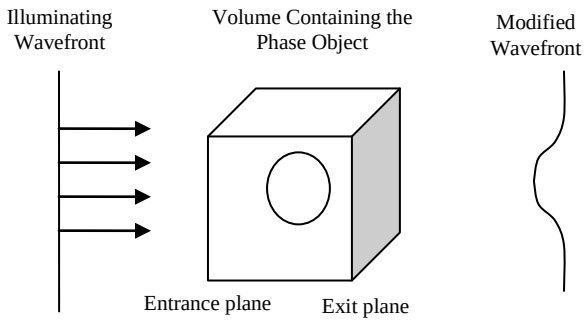


Figure 6.4 Schematic of the imaging system

Therefore, it should be possible to develop an algorithm to calculate a 3D model of the object, in the cases where the object can be determined by eye. There will be a problem with determining a unique object from a single view as, in general, an object calculated from a single image is known to be non unique, Devaney [62].

Subsequent sections describe the imaging process and look at the correlation between the in-focus points and the position of the object, and look at using the point with the maximum absolute phase to define the in-focus points to determine the object.

6.2.1. Simulating the Imaging Process

This section describes the numerical process for propagating the complex amplitude through an object and the back propagation process, equivalent to viewing an object through a microscope. The section then shows examples where the in-focus points correspond to the position of the object.

In a similar way to the previous chapters the complex amplitude of the illuminating wave at the entrance plane of the volume, Figure 6.4, is defined as

$$\psi_0(\mathbf{x}) = \sqrt{I(\mathbf{x})}\exp\left(-i\alpha(\mathbf{x})\right). \tag{6.1}$$

As before, in Chapter 2, known phase aberrations in the illumination are defined as $\alpha(\mathbf{x})$ and the intensity is defined as $I(\mathbf{x})$. If $\alpha(\mathbf{x}) = 0$ and $I(\mathbf{x})$ is constant then the object is illuminated by a plane wave. The subscript 0 defines the first slice of the volume. In the simulations that follow the volume is modelled a discrete 3D array and the propagation through the object is simulated recursively as

$$\psi_{n+1}(\mathbf{x}) = \wp\left\{\psi_n(\mathbf{x})\right\}\exp\left(-i\phi_{n+1}(\mathbf{x})\right). \tag{6.2}$$

The complex amplitude in the nth plane is denoted by $\psi_n(\mathbf{x})$, the specimen-induced phase change by $\phi_n(\mathbf{x})$ and $\wp\{\ \}$ denotes the propagation between the planes. The propagation between the planes is simulated using the angular spectrum of plane waves, Goodman [37].

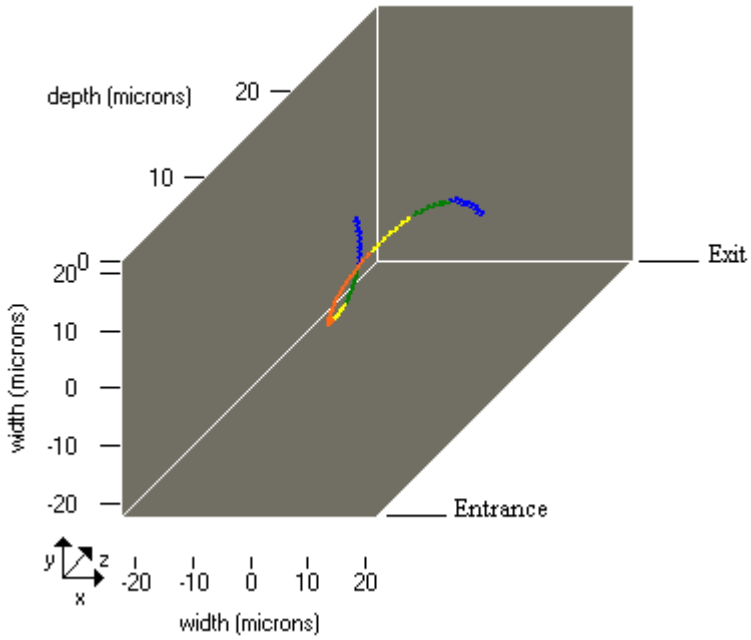
Depending on the configuration of the experiment, further optics could be used to reimage the exit plane wave to produce, for example, phase contrast images that can be analysed as discussed in Chapter 2 but for the purpose of this chapter, the complex amplitude of the wave in the exit plane is assumed to be known.

The wave in the exit plane is then propagated backwards through the same volume as a free-space propagation so no object is present in the volume for the back propagation. The back propagated data will be used to calculate the phase object.

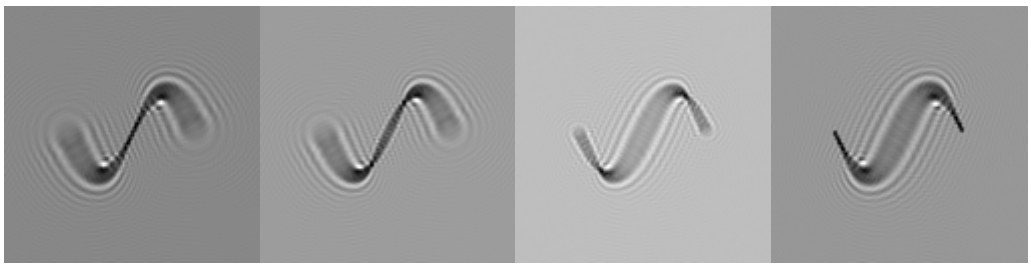
Figure 6.5 shows an example of the correspondence between the position of the object and the in-focus points from the back propagation. Using a helix as the phase object, slices of the original object are compared with the same slices from the back propagated data.

Figure 6.5 a) shows a 3D volume contain the phase object drawn with "cavalier perspective" with no reduction of z scale to imply depth, Giesecke, *et al.* [74], because of this 3D objects drawn in this chapter can look stretched along the z-axis. The colours indicate depth with orange at the front then yellow, green and at the back blue. Throughout the chapter, the entrance plane is taken to be at the front and the exit plane is at the back with the propagation direction along the z axis.

The object is illuminated by a plane wave and the complex amplitude in the exit plane recorded. This is then propagated backwards through the same volume assuming a homogeneous medium. Figure 6.5 b) shows slices in the xy plane from the back propagated data and Figure 6.5 c) shows the corresponding slices of the original object.



a)



b)

c)

Figure 6.5 a) 3D view of the phase object b) Images showing the phase of the back propagated complex amplitude in different planes. c) Corresponding slices of the phase object. The depth of the volume is 30 μm and the slices in b) and c) are, from left to right, 9.6 μm, 11.25 μm, 17.34 μm and 19.68 μm from the entrance plane.

The set of pictures in Figure 6.5 b) clearly show the contribution from all the surrounding planes with different levels of defocus. Scanning throught the volume containing the back propagated data, the points in a given plane that are in-focus map to the points of the original phase object as shown in Figure 6.5 c). It should therefore be possible in some cases to build up a model of the object by scanning through the volume to “pick out” the in-focus points.

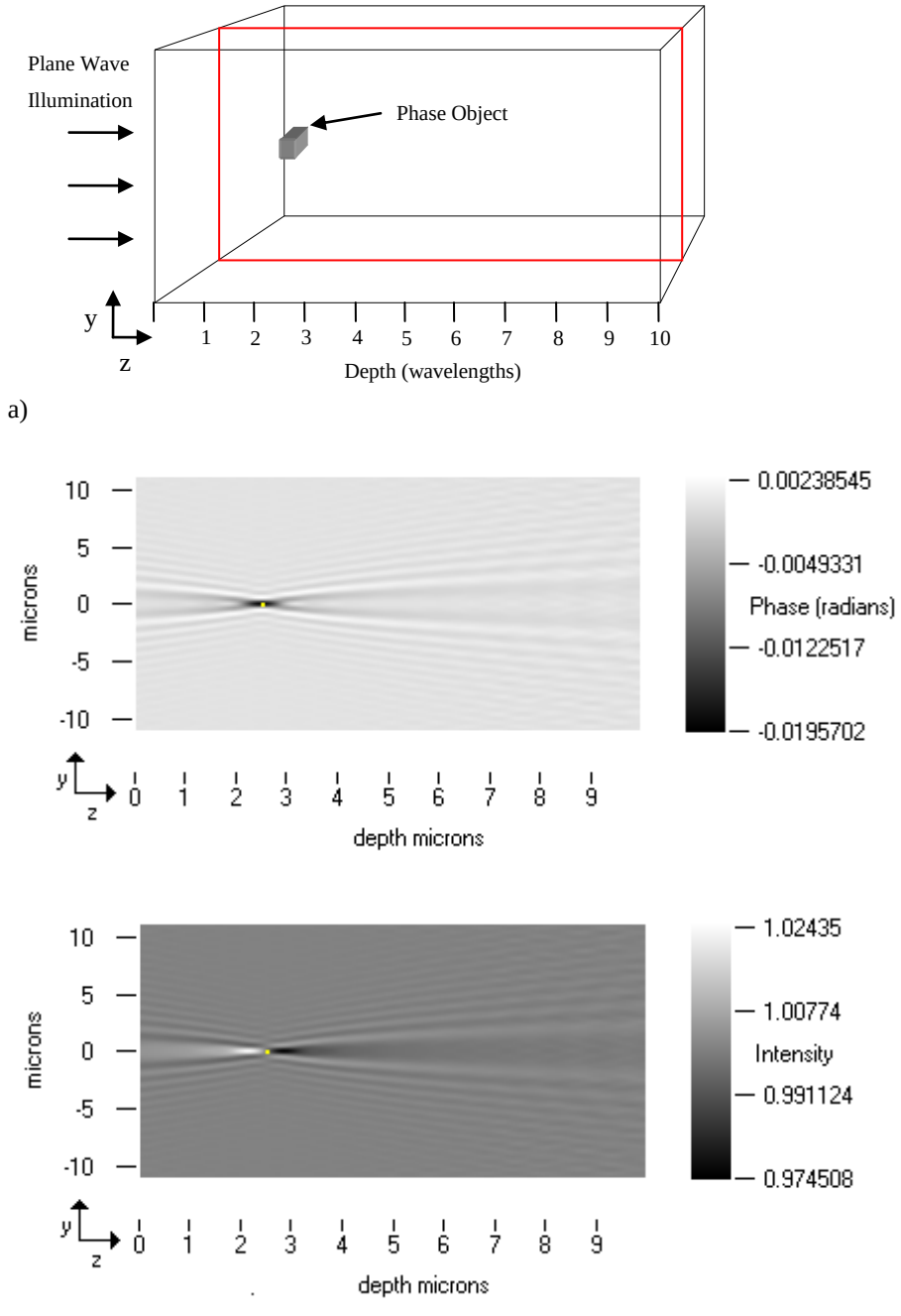


Figure 6.6 a) Schematic of the imaging system showing a phase object one volume element in size is placed 2.5 microns into the volume from the entrance plane. b) Across section of the back propagated phase (radians). c) a cross section of the back propagated intensity. The focal points are marked in yellow, 2.5 microns from the entrance plane. The entrance plane is at the left and the exit plane at the right.

To "pick out" an in-focus point, a criterion for distinguishing these points is required. A suitable criterion is determined by considering an object defined as one volume element with size $(\lambda/2, \lambda/2, \lambda/3)$, placed inside the volume as shown in Figure 6.6 a). It is then illuminated by a plane wave and the resulting complex amplitude in the exit plane back propagated through the volume. Figure 6.6 b) and c) shows cross sections of the back

propagated intensity and the phase though the yz plane containing the object, the plane is outlined in blue in a).

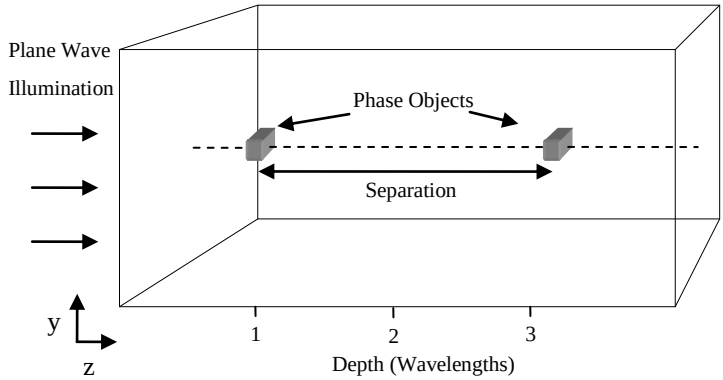
The figure shows that, for an object the size of one volume element, the position of maximum absolute phase, marked in yellow, 2.5 microns into the volume, corresponds to the in-focus point which is in the same position as phase object. The value of the intensity at the position of the object matches the background intensity level making the intensity difficult to use to select in-focus points unless the intensity data is viewed, for example, in phase contrast.

In Figure 6.6 and in the previous example, the points with the maximum absolute phase value match the position of the object. This suggests that this can be used as a criterion for positioning the object. This represents a single-scattering situation and the interpretation is straightforward as expected.

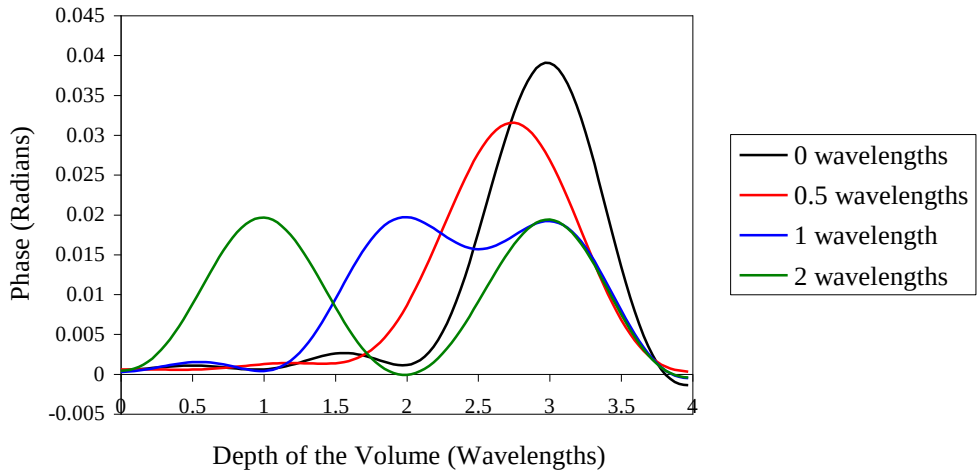
To be effective as a practical tool an algorithm designed to determine a three-dimensional object will need to be able to distinguish between objects that are placed one behind the other. As a test of this power of discriminating two phase objects were considered to be placed in a volume, as shown in Figure 6.7 a), each one was one volume element in size and caused a phase change of 0.1 radians

The first object is placed three wavelengths into the volume from the entrance plane and the second object is placed directly behind the first. The separation of the object is varied to determine if the position of the object can be determined and if the two objects can be separated from the back propagated data.

Figure 6.7 b) shows the phase along the z axis through the volume as shown in a) by the dashed line. For a single object, zero separation, the position of the peak matches the position of the object as expected from the previous example. As the second object is moved away from the first and while the separation is less than one wavelength, the peak is at the mean depth of the two objects. When the separation exceeds approximately one wavelength, there are two distinct peaks. Although the positions of the peaks are not in exactly the correct position as the two objects interact until the objects are sufficiently far apart.



a)

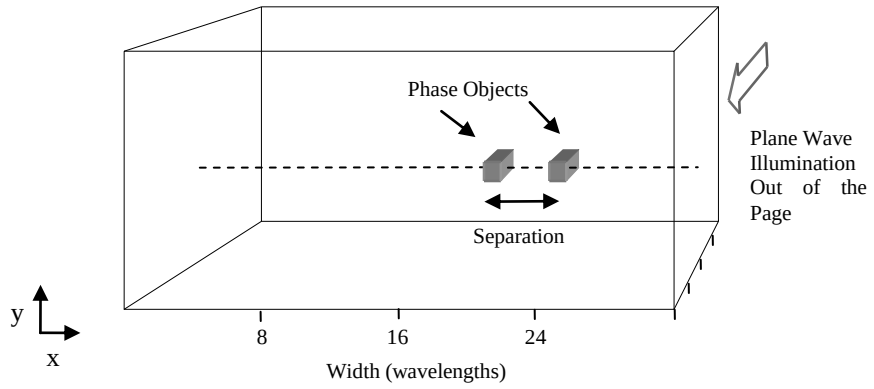


b)

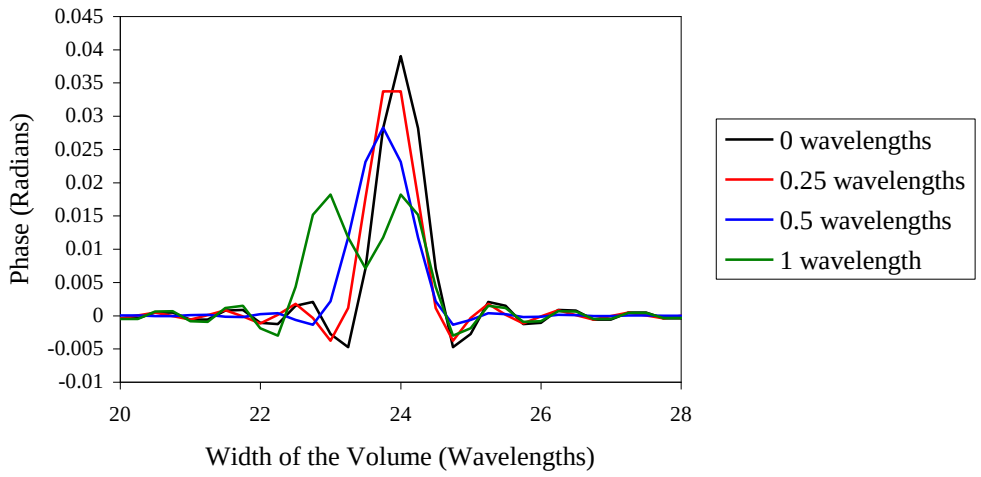
Figure 6.7. a) shows a schematic of the imaging system where two objects are placed in the volume directly behind each other. Each phase object is one volume element and both apply the same phase change of 0.1 radians. b) shows a plot of the phase through the volume along the dashed line shown in a)

Therefore, provided the objects are small enough and far apart then it should be possible to determine the position. When small objects are close together, then it will become more difficult to separate them, as there is an ambiguity between two small objects, which are close together, and a single object causing a larger phase change.

Figure 6.8 shows a similar simulation to Figure 6.7, two objects are placed in the volume are moved apart laterally, the first object is placed 24 wavelengths from one side and the second is place a known distance along the x axis as shown in Figure 6.8 a). In this figure the volume is illuminated from the back so the wave is propagating out of the page.



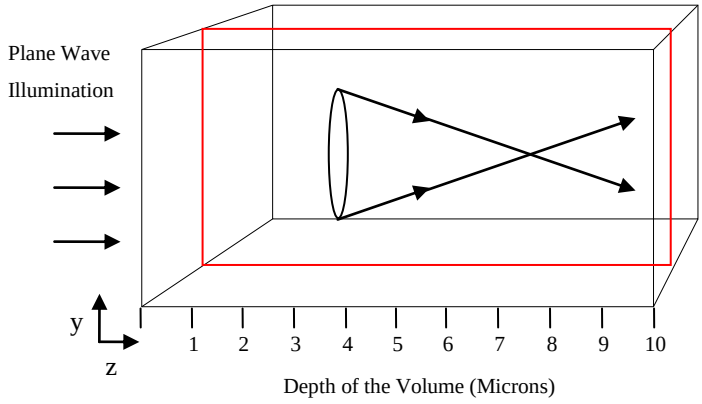
a)



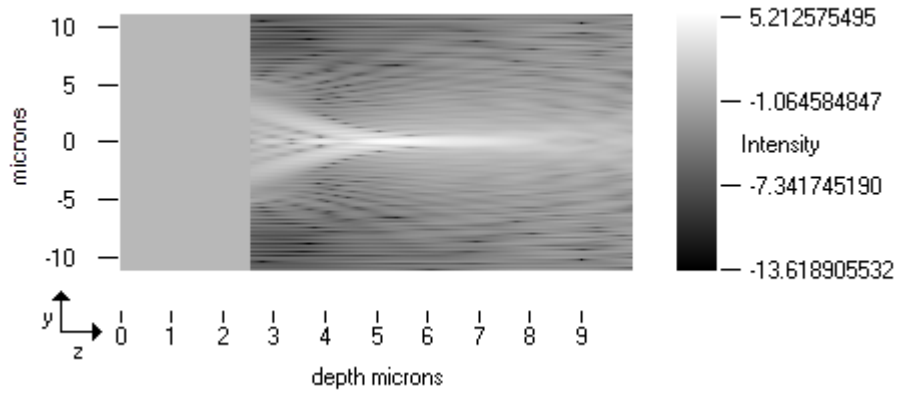
b)

Figure 6.8 a) Schematic of the imaging system where two objects placed in the volume and separated laterally. Each object is one volume element in one size and causes a phase change of 0.1 radians the illuminating wave is propagating out of the page. b) shows a plot of the phase through the volume along the dashed line in a).

Figure 6.8 b) shows the phase along the x axis through the objects marked by the dashed line in a). The figure shows cross sections of the phase for different separations of the objects. It shows that provided the objects are further than one wavelength apart there are two distinct peaks and the position of each object can be determined. If the separation is less than half a wavelength, then positioning the objects is difficult as this could be interpreted as a single object.



a)



b)

Figure 6.9 a) Schematic of a lens placed in a volume at 2.5 μm from the entrance plane and illuminated by a plane wave. b) shows a cross sections of the intensity, on a log scale, of the plane marked in red in a)

So far, the examples have used phase objects that are small and confined to a single volume element so the focal points match the position of the objects. As an example of a larger object, Figure 6.9 a) show an idealised lens placed in the volume. It is then illuminated from the left and the light focuses to the right of the lens. Figure 6.9 b) shows a cross section of the intensity as the light propagates through the object from left to right. The figure shows the intensity instead of the phase because the phase is wrapped so it is not clear where the focal point is. The figure shows that the focal point does not match the position of the object as the lens has focused the light.

In general, as the object becomes larger it will cause a focusing effect, so the in-focus points will not match the position of the object. Therefore, there will be a limitation on the process to determine the 3D object due to its size. This can be seen in the examples in section 6.3.1.

This section has discussed numerically back propagating the complex amplitude through a volume to determine the structure of an object, a process equivalent to adjusting the focus of a microscope. The section has shown that in some cases, when the phase objects are sufficiently small and far apart then the in-focus points correspond to the position of the object, so from this information the structure of the object can be determined. Therefore, the following section develops a process for determining 3D objects, which are more complicated than a single volume element, from the in-focus points within the volume, for this limited class of objects.

6.3. Iterative Calculation of the Object from the Data

The previous section described a process of numerically propagating a wavefront through a volume that is equivalence to adjusting the focus of a microscope. The in-focus points were indicated by the maximum absolute phase value and were shown in some cases to correspond to the position of the object. Therefore, this section develops an iterative process to determine the structure of an object from the in-focus points and examines the correlation between the focal points and the position of the object. The iterative process is required to build up an approximation of the object, from many in-focus points, and to separate the interaction between parts of object.

In the cases where the in-focus points match the object there are still parts of the image that are out-of-focus. It has been shown, Devaney [63], that the back propagated data is related to the phase object by an integral equation. To calculate the three-dimensional phase object this equation needs to be inverted,

$$BP(\mathbf{x}) = \int_V O(\mathbf{x}')h(\mathbf{x},\mathbf{x}')d\mathbf{x}'. \tag{6.3}$$

This shows that the back propagated phase, $BP(\mathbf{x})$ is related to the object $O(\mathbf{x}')$ and $h(\mathbf{x},\mathbf{x}')$, where $h(\mathbf{x},\mathbf{x}')$ describes the imaging system. In astronomy, there is a problem with a similar formulation, where measured intensity data can be described as a convolution of a sharp image with a function that describes the imaging process. To calculate the sharp image, the measured intensity data needs to be deconvolved with the function describing the imaging process.

As deconvolution is ill posed, iterative algorithms such as “clean”, Starck, *et al.* [75], have been developed that avoid direct deconvolution. This algorithm works on the measured intensity. It finds the brightest point and on a separate image, a fraction of that

value is placed in the same position. This model of the data is then convolved with the function representing the imaging process and the resulting intensity is subtracted from the original measurement. The process is then repeated for the next brightest point until a model of the sharp image is produced and the effects of the imaging process are removed. In this way, direct deconvolution and regularisation can be mitigated but the process is limited as it assumes that the image is made up of single points.

The method considered here for reconstructing the phase object is similar to the “clean” algorithm. The position with the largest absolute phase value is found and some fraction of this value is place in a new “clean” volume. This assumes a one to one correspondence between the in-focus points, and the position of the object. The current object model consisting of all the points in the volume is then illuminated in the same way as the original object to estimate the exit plane data. The estimated wave in the exit plane is then back propagated and the complex conjugate of the wave is multiplied by the original back propagated data to remove the effect of the estimated phase object from the back propagated data. This process is then repeated for the next largest absolute phase value.

This process continues until a set number of iterations is reached or the error between set and estimated phase in the exit plane reaches a preset threshold. The phase value that is placed in the “clean” volume can be negative allowing incorrect values to be removed.

In an experiment, the set phase is calculated from the measured data using, for example, one of the methods described in Chapter 2 or Chapter 3 and the estimated phase is calculated by illuminating the estimated phase object.

The phase of the back propagated data needs to be unwrapped as the propagation distances are longer than a wavelength. The data is unwrapped by back propagating the complex amplitude at the exit plane when there is no object in the volume. The complex conjugate of the back propagated wave, with no object, is multiplied by the back propagated data to leave only the phase changes due to the object.

In the following examples, the phase objects are weak so the total phase change due to the object, over the depth of the volume, is not wrapped. Therefore, it is still valid to

look for the largest absolute phase values. This method also assumes the object only changes the phase and there is no absorption so that the change in the wave from the entrance plane to the exit plane is purely due to a phase change. This reduces the ambiguity of the object. These simple examples allow the fundamental problem, of back propagation through a volume containing sources, to be separated from the objects strength and the problem of phase unwrapping.

6.3.1. Simulations

As a test of the 3D phase reconstruction algorithm described in the previous section, two phase objects are used to assess the potential for calculating 3D objects. Figure 6.12 shows the two test objects. The scales show the sample number but in the following examples, physical units are used with the same objects. Therefore, the objects are physically different sizes but the reconstructed objects can be compared point by point with the set object.

The first object, shown in Figure 6.12 a), is a single period of helix, the centre points, (x, y, z), of the tube are defined as

$$\begin{aligned} x &= x_0 + A\cos\left(\frac{2\pi\left(y_0 - L/2 + t\right) f}{L}\right) \\ y &= y_0 - L/2 + t \\ z &= z_0 + B\sin\left(\frac{2\pi\left(y_0 - L/2 + t\right) f}{L}\right) \end{aligned} \tag{6.4}$$

the length L is represented by 64 samples, $A = 15$ samples and $B = 8$ samples. The frequency of the helix, f is 1 cycle per length, L. The centre of the grid is the point (x_0, y_0, z_0) and t indexes the number of samples in the length, L. The radius of the tube around these points is one sample. The phase per volume element is set at 0.1 radians. It is anticipated that this will act essentially as a single scattering object.

The second object a hollow sphere, shown in Figure 6.12 b), the relative phase change is only caused by the surface of the sphere as the inside has the same properties as the volume outside of the sphere. The object is defined as the surface of a sphere 20 samples from the centre of the array where each volume element has a phase value of 0.01 radians. This provides an example of a multiple-scattering object.

Figure	Sample Spacing	Scattering Angle (radians)
Figure 6.13	$\lambda/2$	$\pi/2$
Figure 6.15	4λ	$\sim 1/8$
Figure 6.17	$\lambda/2$	$\pi/2$
Figure 6.18	4λ	$\sim 1/8$

Figure 6.10 Parameter used in the simulations

Figure	NA	Depth of field	Magnification
Figure 6.13	1	$\pm\lambda/2$	50 – 100
Figure 6.15	$1/8$	$\pm 32\lambda$	5 – 10
Figure 6.17	1	$\pm\lambda/2$	50 – 100
Figure 6.18	$1/8$	$\pm 32\lambda$	5 - 10

Figure 6.11 Additional quantities describing the simulated systems

In each case, the wavelength is 633 nm. The sample spacing in each direction and the maximum scattering angles are defined in Figure 6.10. The larger the scattering angle the quicker the wave can diverge, making the object easier to position. As the medium within the volume is homogenous up to the exit plane these examples simulate a system using a dipping lens, Pawley, *et al.* [76].

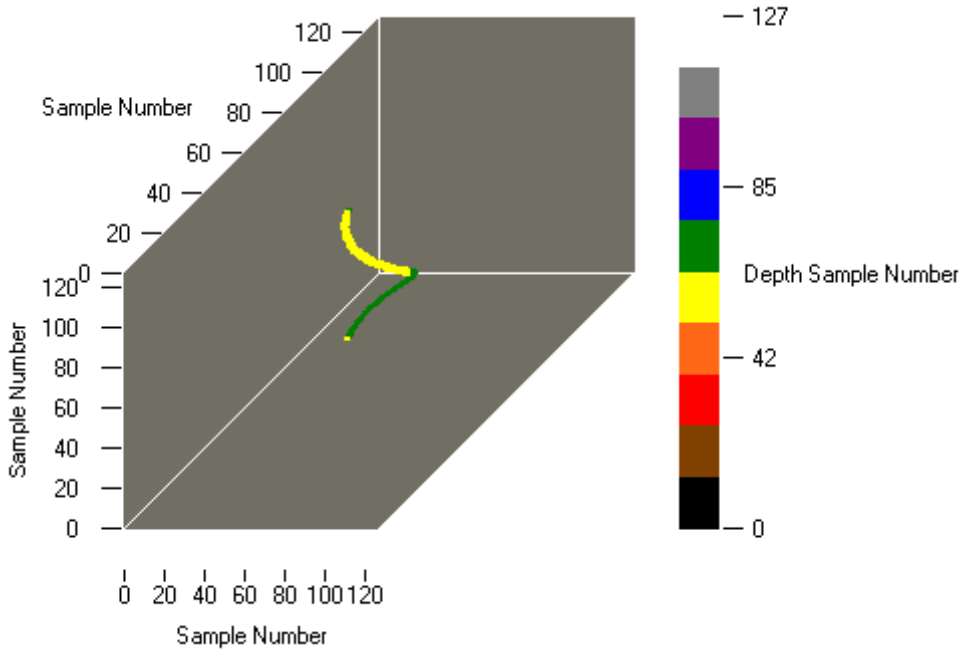
The numerical aperture, NA, for each of these systems can be calculated, Hecht [4] , using

$$NA = n\sin(\sigma) \tag{6.5}$$

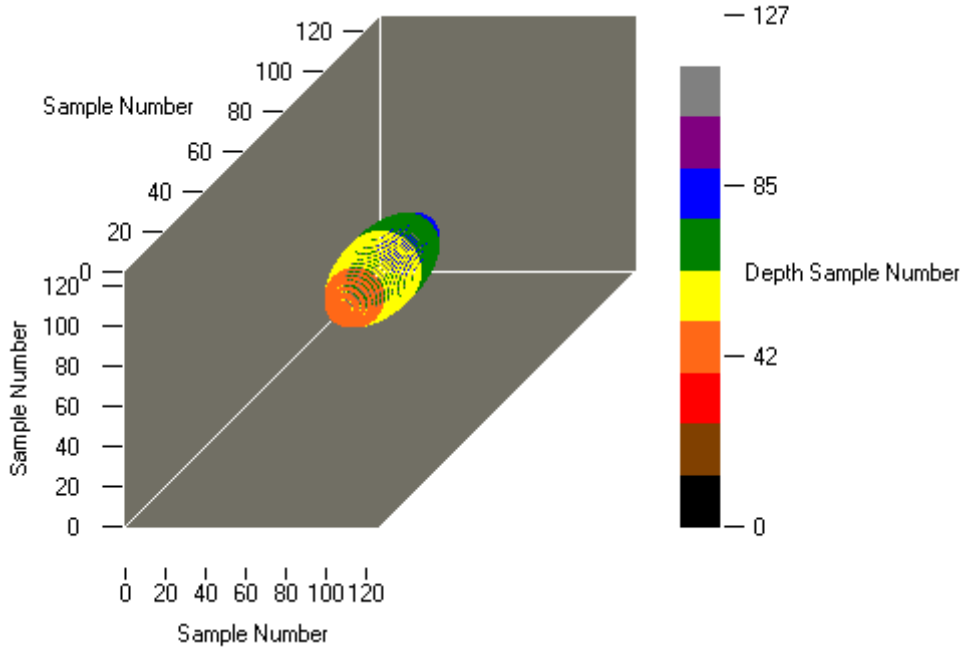
where n is the refractive index, this is set to one and σ is the maximum scattering angle. The depth of focus can also be calculated, Marshall [77] , using

$$D = \pm\frac{\lambda}{2(NA)^2}. \tag{6.6}$$

For the systems modelled here the objective lens would have a magnification between 50 and 100 when the NA is 1 and between 5 and 10 when the NA is $1/8$, Olympus [78]. For the low NA system, experimentally this would have to be assembled in bulk optics as the magnification is too low to view the object using a standard microscope.

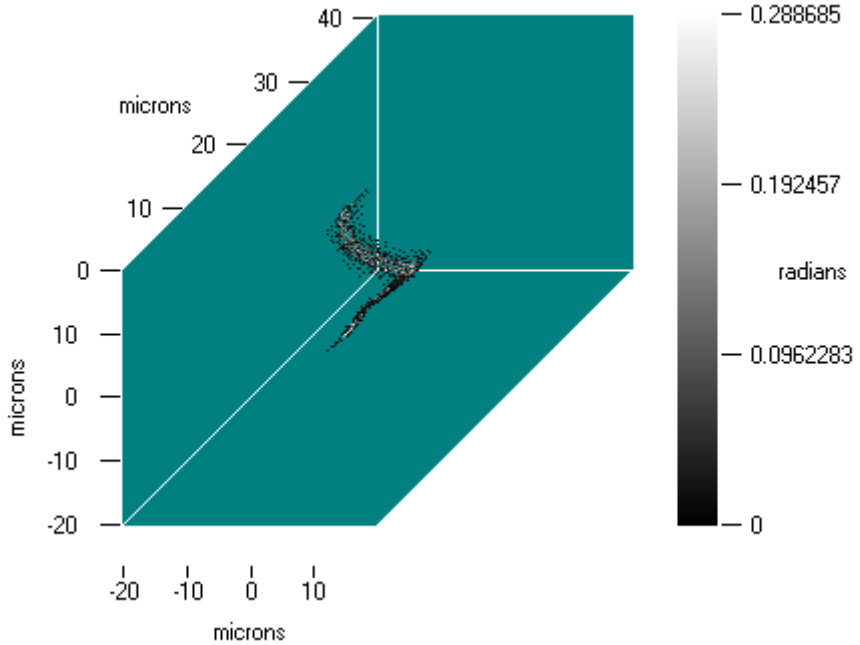


a)

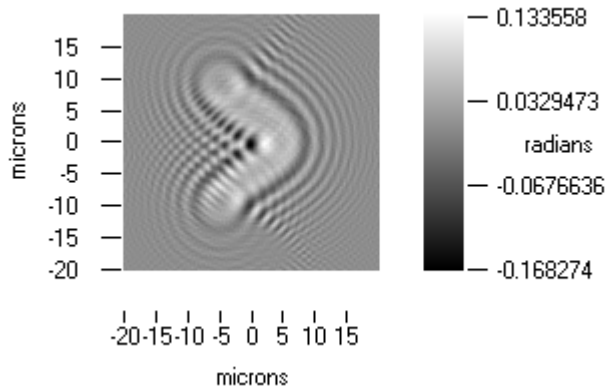


b)

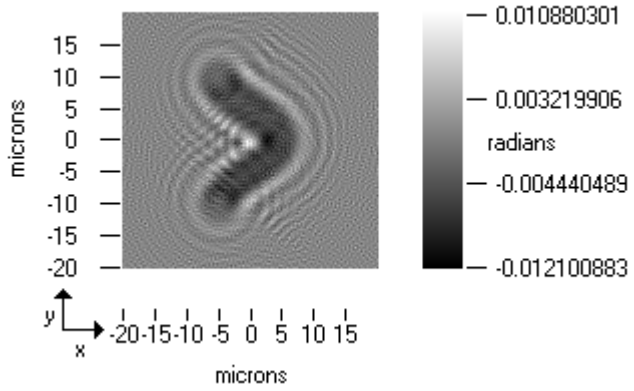
Figure 6.12 Phase objects: a) a helix and b) a hollow sphere. The colour scale indicates the position of the object in depth, along the z axis.



a)



b)



c)

Figure 6.13 Results of the iterative process to estimate the object shown in Figure 6.12 a). a) shows the estimated object. b) shows the phase in the exit plane after illuminating the estimated object. c) shows the residual error between the phases in the exit plane corresponding to the original and estimated objects. NA = 1.

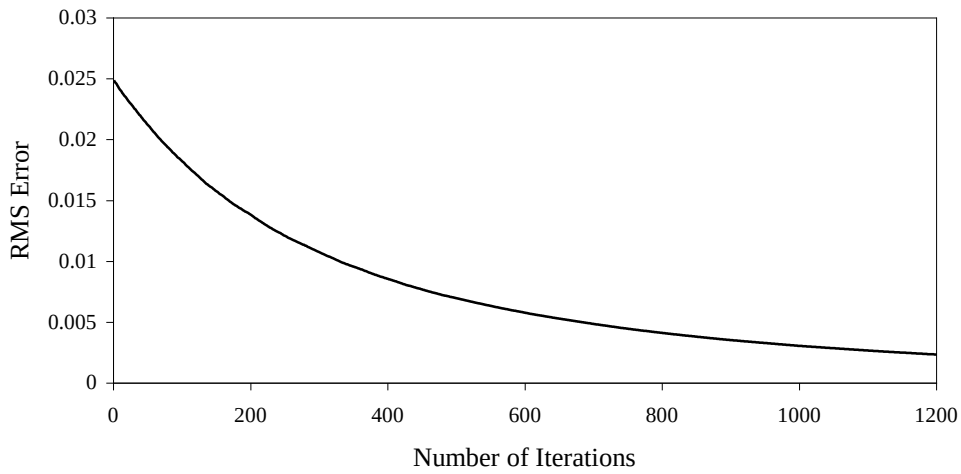
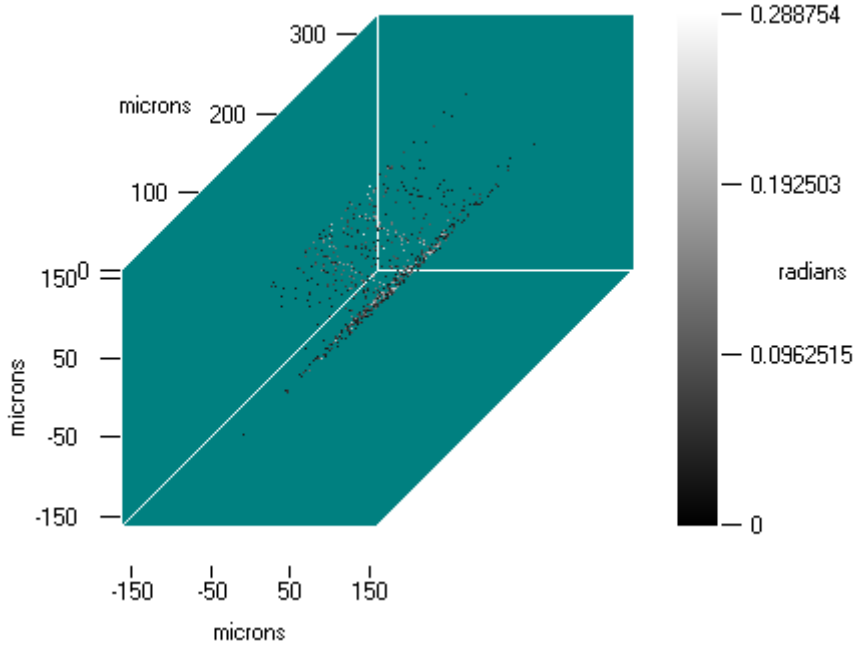


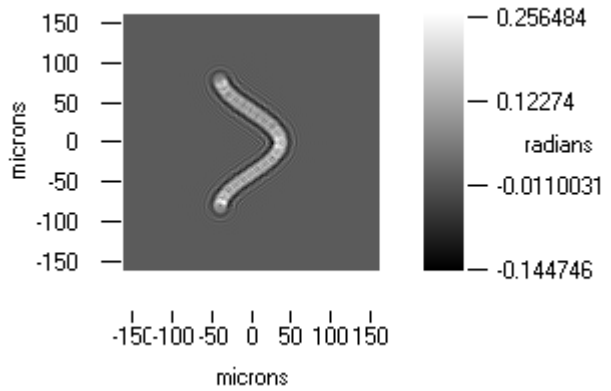
Figure 6.14 Convergence of the iterative scheme for example in Figure 6.13

The first test of the 3D reconstruction algorithm, Figure 6.13 a), shows a reasonable qualitative approximation of the phase object structure, although there are many outlying points and the values of the individual pixels are not uniform. The error, shown in Figure 6.13 c), between the estimated exit plane phase and the set phase is small. Therefore, the phase in the exit plane Figure 6.13 b) is a good approximation of the set phase. This example also shows the potential ambiguity of the reconstruction process as the objects are different but produce similar exit plane data.

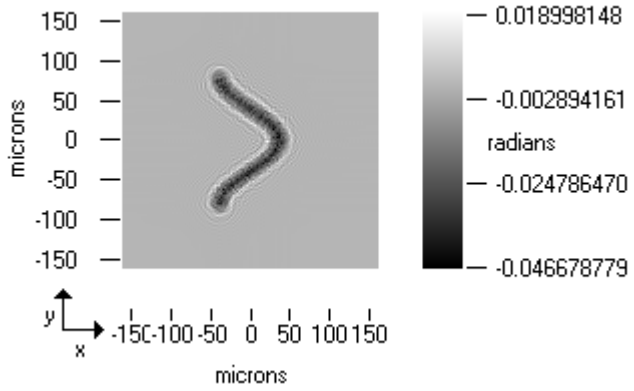
Figure 6.14, shows the RMS error between the set and estimated phase in the exit plane as a function of the number of iterations for the example in Figure 6.13. The figure shows the error asymptotically tending towards zero. As only a half of the phase value is added per iteration, this suggests that the estimated phase in the exit plane will tend towards the set phase given enough time. The example could be left to iterate further, but without additional information, it is unlikely to improve the estimation of the object, it will only generate an object that produces exit plane data consistent with the set exit plane data and reduce the residual error.



a)



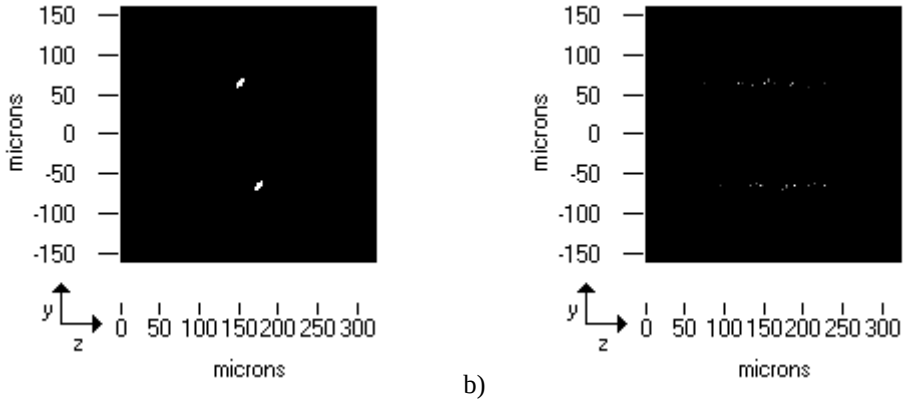
b)



c)

Figure 6.15 Results of the iterative process to estimate the object shown in Figure 6.12 a). a) shows the estimated object. b) shows the phase in the exit plane after illuminating the estimated object. c) shows the residual error between the phases in the exit plane corresponding to the original and estimated objects.

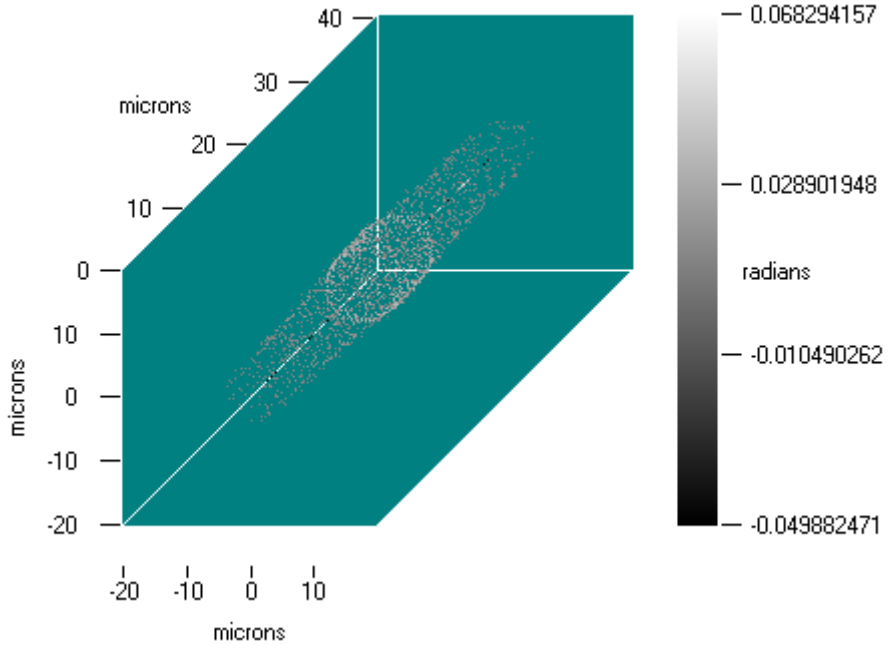
NA = $1/8$



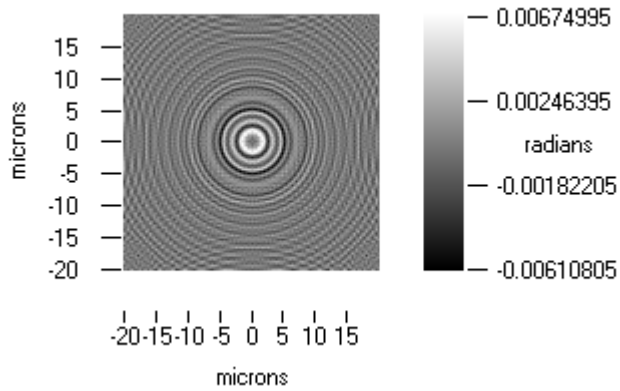
a) b)

Figure 6.16 a) shows a cross section of the set phase object for the example in Figure 6.15. b) shows the same cross section through the estimated phase object for the example in Figure 6.15. The cross sections are through the plane $x = -32.9\mu m$

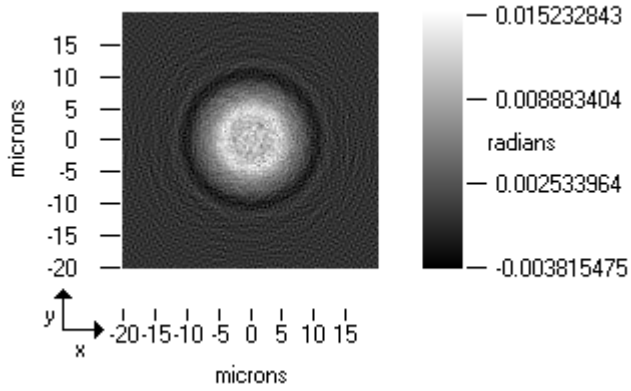
Figure 6.15 uses similar shaped example object, the sample spacing is greater therefore, the dimensions of the volume are increased and the maximum scattering angle of the system is smaller so the object should be harder to position. Figure 6.15 a) shows the reconstructed object. The xy position of the object have been reconstructed better than the depth of the object, as this is harder to determine and therefore the object has been stretched out. This stretching can be clearly seen in Figure 6.16 b) where the reconstructed clean map has sources distributed along the z axis compared to Figure 6.16 a) where set phase object is defined over a small volume. Even with these differences, the error in the exit plane data, shown in Figure 6.15 c), is small. Again, this shows the potential for specifying an object that can produce the same exit data but is different from the real object.



a)

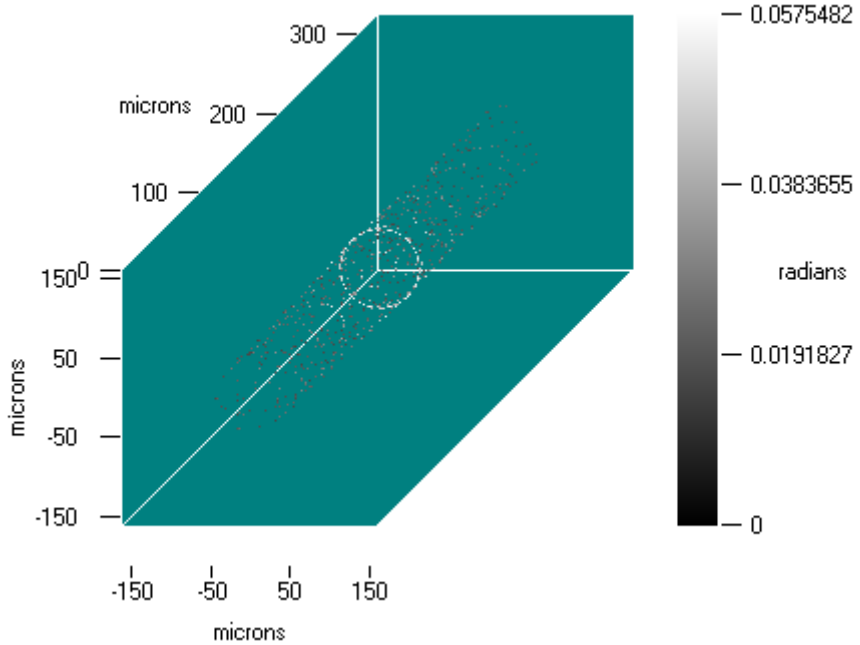


b)

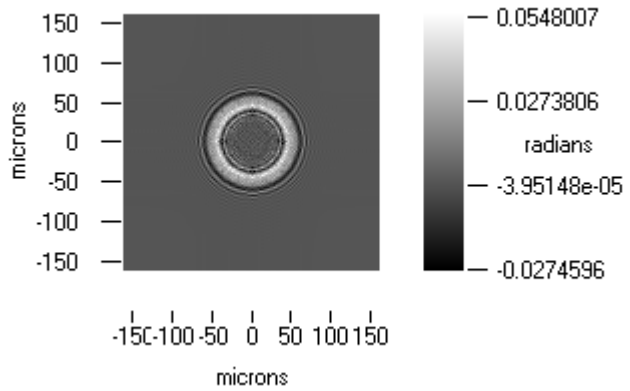


c)

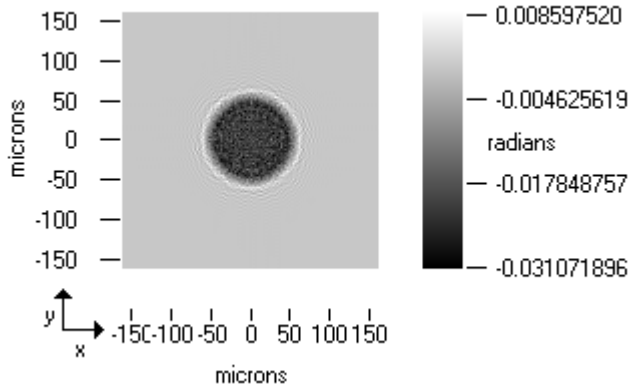
Figure 6.17 Results of the iterative process to estimate the object shown in Figure 6.12 b). a) shows the estimated object. b) shows the phase in the exit plane after illuminating the estimated object. c) shows the residual error between the phases in the exit plane corresponding to the original and estimated objects. NA = 1



a)



b)



c)

Figure 6.18 Results of the iterative process to estimate the object shown in Figure 6.12b). a) shows the estimated object. b) shows the phase in the exit plane after illuminating the estimated object. c) shows the residual error between the phases in the exit plane corresponding to the original and estimated objects. NA = $1/8$.

The hollow sphere example is not expected to be accurately reconstructed, as it should act something like a lens, so the in-focus points will not map directly to the position of the object.

Figure 6.17 a) and Figure 6.18 a) show that the estimated objects have been stretched along the z axis as in the previous example. The estimated exit plane data in Figure 6.17 b) and Figure 6.18 b) are similar to the set data. The error shown in c) is not small but there is little structure. However, the reconstructed object is clearly different to the set phase object, due to the expected non-uniqueness of calculation and the selection criterion for the positioning of the sources. Even with a large maximum scattering angle, the reconstructed object in Figure 6.17 a) is still stretched along the z axis.

This algorithm can give reasonable qualitative estimation of the original object, but due to the potential ambiguity the calculated objects are clearly different even though the exit plane data can be similar. The positioning of the objects along the z axis was improved when modelling a system with a higher maximum scattering angle. Thus, the results in Figure 6.13 are qualitatively better than Figure 6.15 and Figure 6.17 is qualitatively better than Figure 6.18

This section has shown that, under certain conditions, the structure of a three dimensional transparent object can be qualitatively estimated from knowledge of the phase and intensity in a single plane. For this process to work the focusing effect of the object will need to be small so that the focal points match the position of the object.

For larger objects, this process will not work as the position of the focal points do not match the position of the object. Nevertheless, exit plane data can be generated that is consistent with the measured data.

To determine objects that are larger and more complicated more information is required and a more sophisticated method for processing the data. Methods for obtaining more views of the object are considered in the following section.

6.4. Calculating a 3D Object from Phase Contrast and Defocused Images

In the previous section, an iterative method was described that estimates the structure of a three dimensional phase object from the intensity and phase known in a single plane.

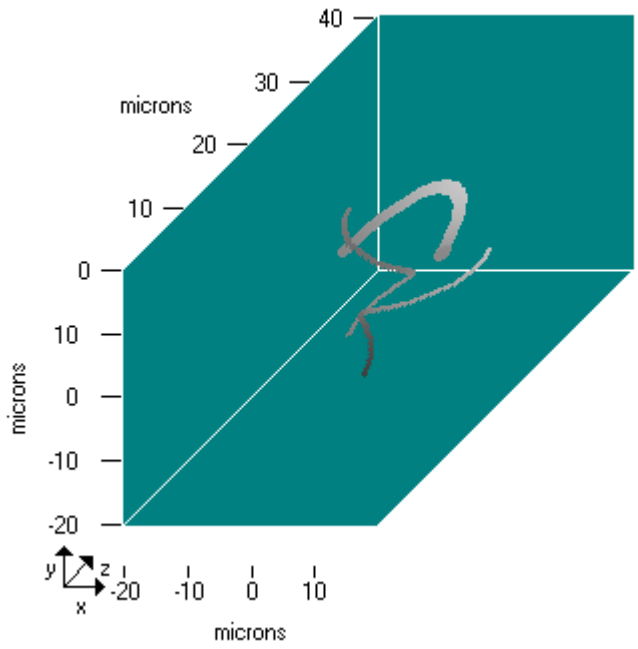
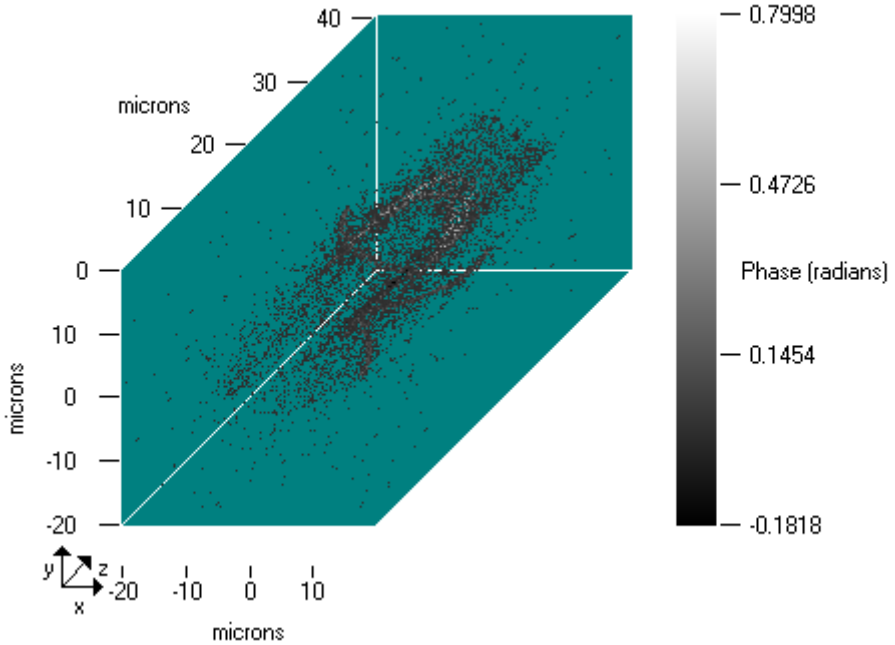


Figure 6.19 Set 3D phase object. The colour represents depth

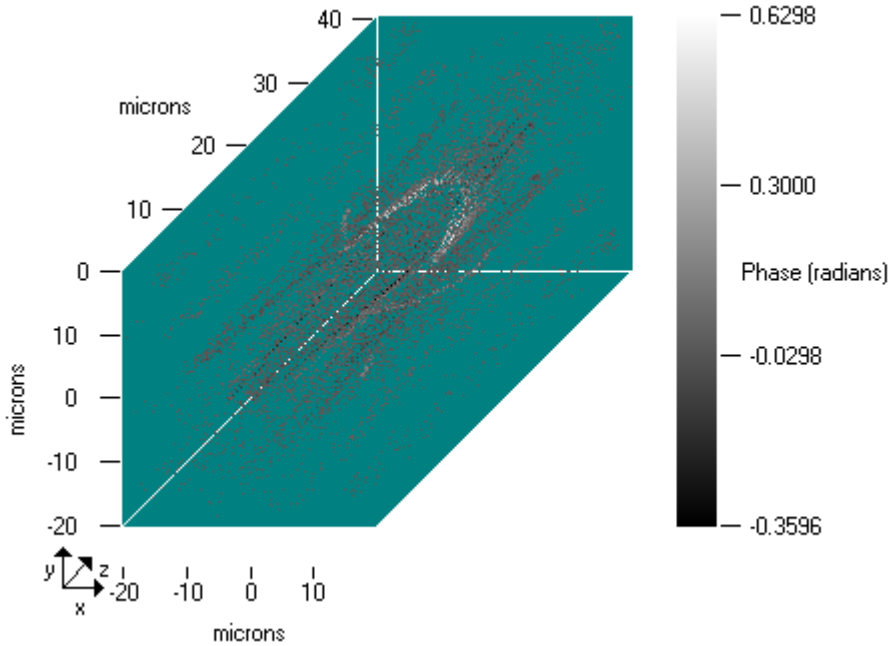
In the previous section, the known phase in the exit plane was used to calculate the object. In this section, a phase object will be illuminated to generate data in the exit plane of the volume. This data will then be used to produce phase contrast images and defocused images that can be analysed using the methods described in Chapter 2 and Chapter 3 to calculate the phase in the exit plane. Using this calculated phase and the known intensity the iterative process described in the previous section is used to estimate the object.

In this section the phase object, shown in Figure 6.19 is used as an example. This is a more complicated object than the single helix used in the previous section as parts of the object are directly behind other parts.

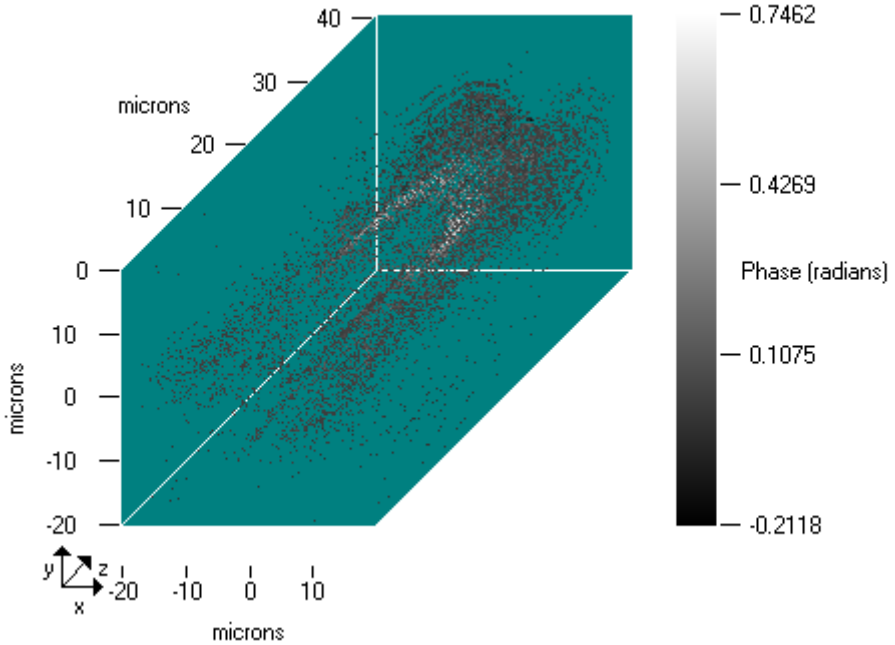
In this example, three helixes are used as the phase object. These are placed in a volume that is represented by 128 samples in each direction. The sample spacing is $\lambda/2$ and the wavelength is 633 nm. The phase change per volume element is 0.1 radians. For the phase contrast images the phase shift is $\pi/4$ radians and the radius of the reference wave area is $3\lambda/8r_0$, as discussed in section 2.5.1. The defocused images are made on planes 10 μm apart.


40
30
microns
20
10
0
10
0
-10
-20
microns
y
z
x
-20
-10
0
10
microns
0.7998
0.4726
Phase (radians)
0.1454
-0.1818

a)


40
30
microns
20
10
0
10
0
-10
-20
microns
y
z
x
-20
-10
0
10
microns
0.6298
0.3000
Phase (radians)
-0.0298
-0.3596

b)



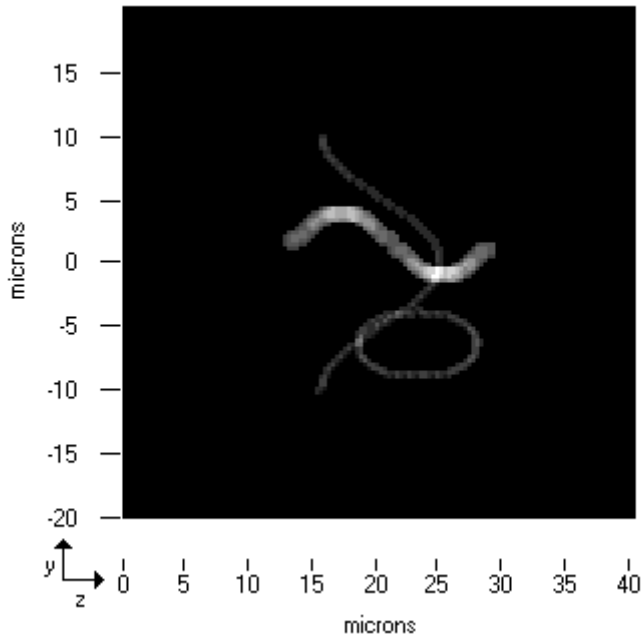
c)

Figure 6.20 a) shows the object calculated from the known phase and intensity in the exit plane. b) shows the object calculated from the phase calculated using phase contrast images and the known intensity in the exit plane. c) shows the object calculated the phase calculated using defocused images and the known intensity in the exit plane

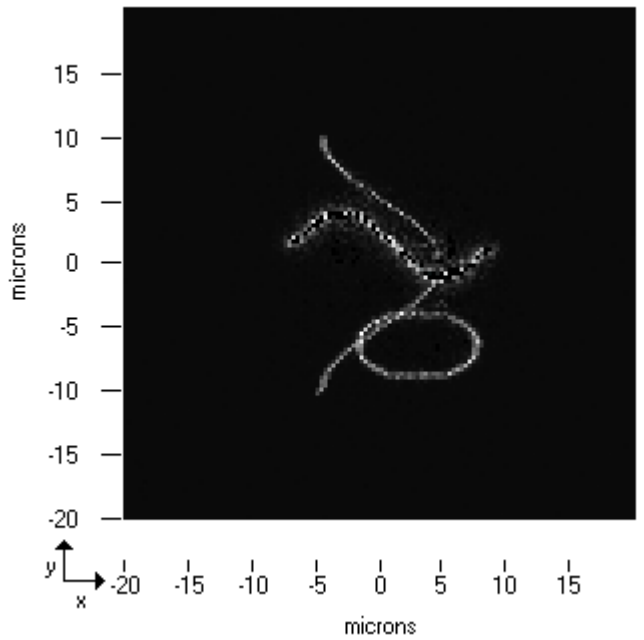
To simulate the process of passing through a circular aperture the exit plane data is placed in a grid 1024×1024 samples. Within the disk where the radius is represented by 256 samples, each point is given the value of the complex amplitude in the exit plane when there is no object. The exit plane data when there is an object is then placed in the centre of the grid. In this way, there is no jump in intensity or phase at the edge of the original grid.

Figure 6.20 a) shows the estimated object calculated from the known phase and intensity in the exit plane. Even in these ideal conditions, it is still difficult to make out the object. Figure 6.20 b) uses the known intensity and the phase calculated from phase contrast images using the three images method described in Chapter 2. Fig c) uses the known intensity and the phase calculated from defocused images using the Fourier transform method described in Chapter 3.

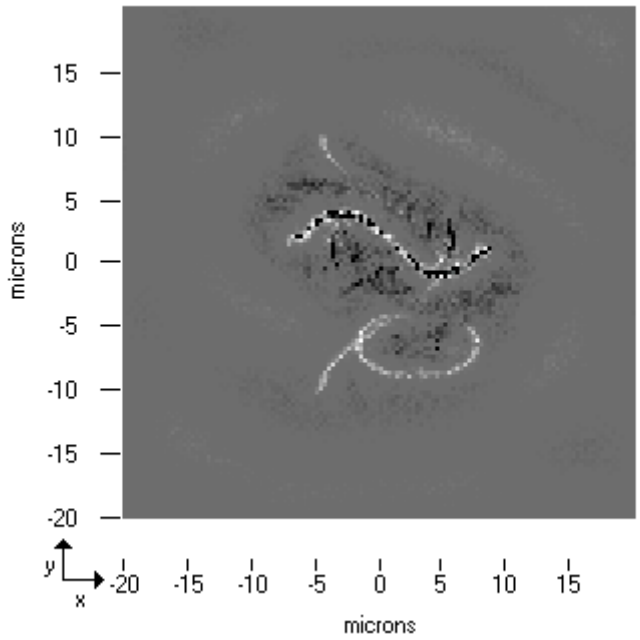
From these pictures, it is hard to distinguish how well the objects have been reconstructed and which method may be better. The estimated objects look no worse than the object calculated from the correct data in the exit plane.


15
10
5
0
-5
-10
-15
-20
microns
y
z
0
5
10
15
20
25
30
35
40
microns

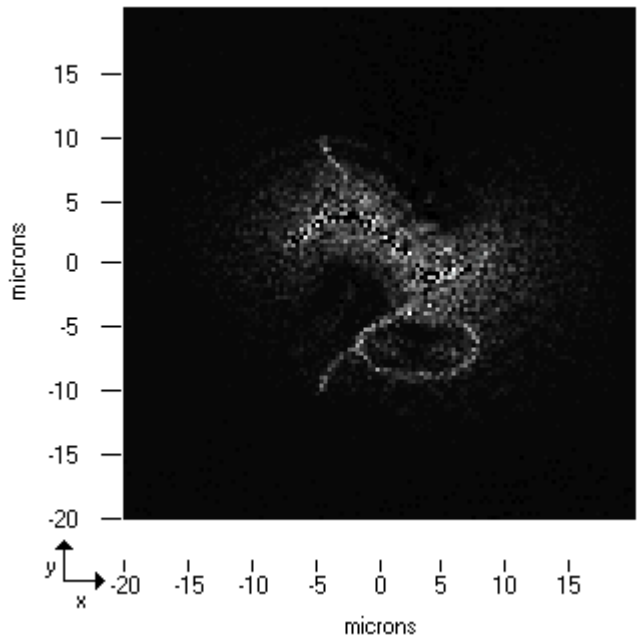
a)


15
10
5
0
-5
-10
-15
-20
microns
y
x
-20
-15
-10
-5
0
5
10
15
microns

b)



c)



d)

Figure 6.21 a) shows the projection of the set object along the z axis. b) shows the projection along the z axis of the object calculated from the known phase in the exit plane. c) shows the projection along the z axis of the object calculated from phase contrast images. d) shows the projection along the z axis of the object calculated from defocused images. The colour scale is arbitrary to allow the fainter parts of the object to be visualised.

To determine the structure of a three dimensional object there are three things that need to be determined the first and easiest is the position in the xy plane the second is the position in depth and the third is calculating the correct phase value at that point. To

visualise the positioning of the object the estimated phase object is projected onto a 2D plane by summing all the sample values along the axis.

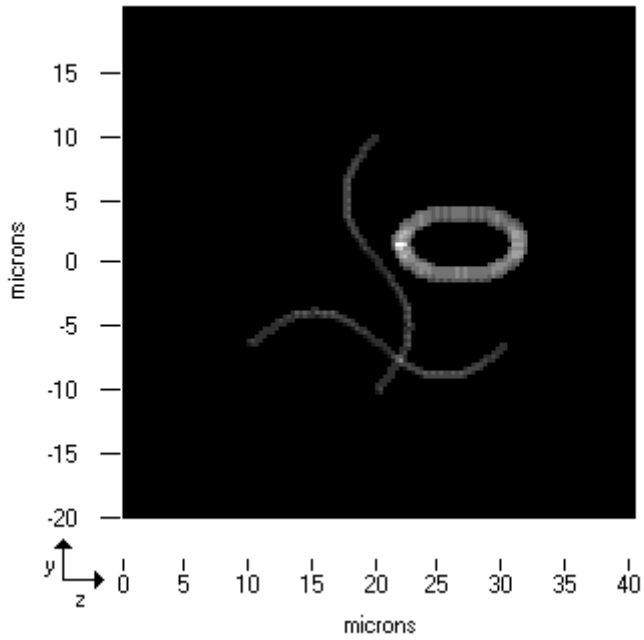
Figure 6.21 shows the projection along the z axis of the set phase object, the phase object calculated from the known phase in the exit plane, and the phase object calculated from the defocused and phase contrast images. There is a large error in the calculated values of the phase object in all the examples. Additionally, the difference between the largest and the smallest values is too large to allow the object to be seen while give a meaningful scale for the phase values therefore all the examples have been normalised differently so that the structure of the object can be seen.

Using the exact phase in the exit plane the projected phase object, Figure 6.21 b), shows that the thicker helix has been estimated as a thin helix. This is due to the focusing effect of the object. Overall the xy position is well determined and there is only a small error in the centre of the image mainly where the objects are directly behind each other.

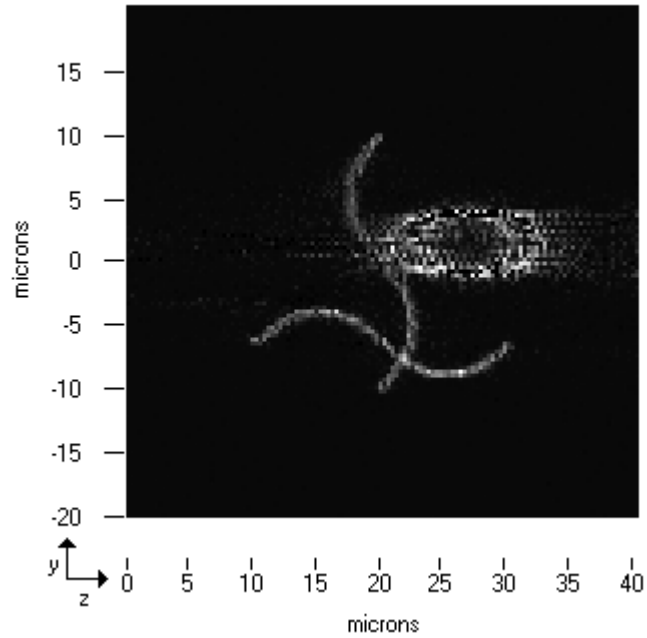
Overall, the shape of the object has been well estimated but there is a significant error in the phase values at each point, the estimated value can be overestimated or under estimated.

The projection of the phase object calculated from the phase contrast and defocused images, Figure 6.21 c) and d), both show a distinct wave from the thicker helix. The ellipse below has also been reasonably well positioned although there are parts that have been missed. The vertical helix in both cases has parts missing this is more significant for the projection of the object calculated from the defocused images, d), which also shows more error in the centre of the image.

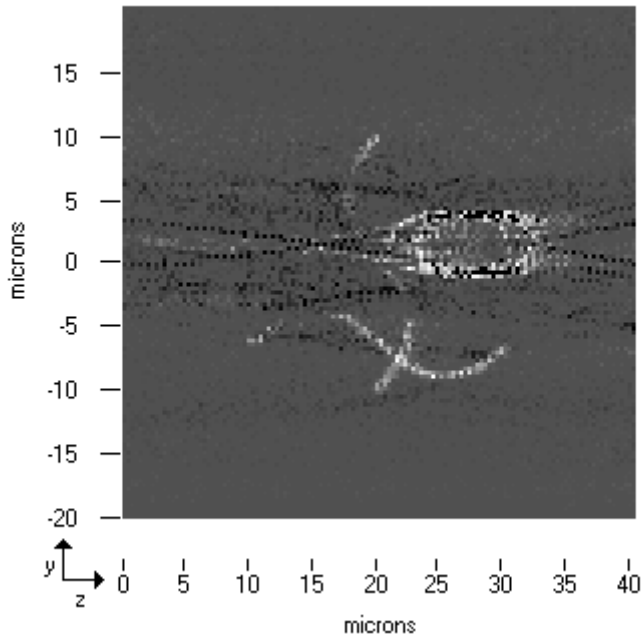
Overall, the projection of the object calculated from the phase contrast images appears to give slightly better xy positioning although neither method has produced a quantitative estimation of the phase.


15
10
5
0
-5
-10
-15
-20
microns
y
z
0
5
10
15
20
25
30
35
40
microns

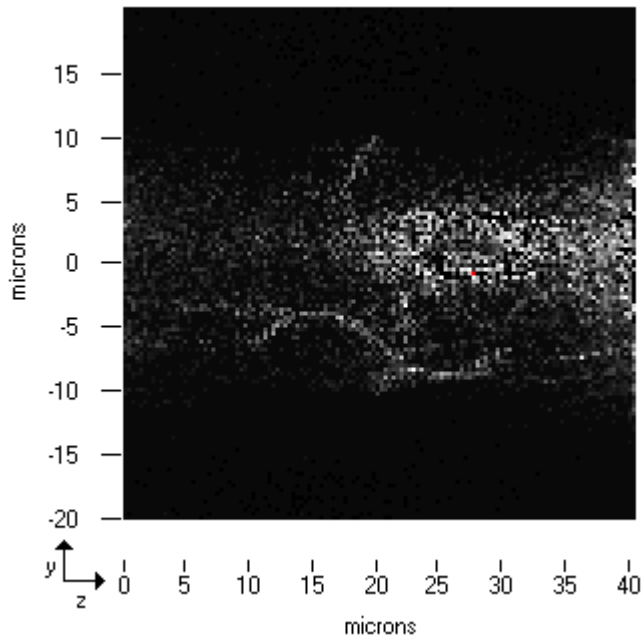
a)


15
10
5
0
-5
-10
-15
-20
microns
y
z
0
5
10
15
20
25
30
35
40
microns

b)



c)



d)

Figure 6.22 a) shows the projection of the set object along the x axis. b) shows the projection along the x axis of the object calculated from the known phase in the exit plane. c) shows the projection along the x axis of the object calculated from phase contrast images. d) shows the projection along the x axis of the object calculated from defocused images. The colour scale is arbitrary to allow the fainter parts of the object to be visualised.

Figure 6.22 shows the same set of figures as Figure 6.21 except the projection is along the x direction. This set of pictures can be used to visualise the accuracy that the iterative process has determined the depth of the object. Figure 6.22 b) shows the projection of the phase object calculated from the known phase in the exit plane. In

comparison to the projection of the set phase object, the y position is well estimated but there is some stretching along the z axis.

The projection of the phase object calculated using from the phase contrast images, Figure 6.22 c), shows a reasonably clear visualisation, although there is some positioning error of the thicker helix in the z axis. The separation between the thicker helix and the object behind it is also not clear. This projection in comparison to the projection of the object calculated from the defocused images, shown in Figure 6.22 d) is much clearer. The projection of the object calculated from the defocused images cannot easily be distinguished from the background error and there is slightly larger error in determining the depth.

Overall, the object calculated from the phase contrast images appears to be a better estimation of object with slightly better positioning in the xy plane and much better positioning in depth. Both methods give a poor estimation of the phase values.

The error between the phase in the exit plane calculated by illuminating the current estimate of the object and the known phase or the phase estimated from the phase contrast or defocused images is plotted as a function of the number of iterations in Figure 6.23. The figure shows that as the number of iterations increases the error reduces therefore the estimated phase in the exit plane that is tending towards the starting phase. This does not mean that the estimated phase object will resemble the set phase object.

The iterative scheme is attempting to generate an object that produces data in the exit plane that is consistent with the set data. The known phase in the exit plane, calculated by illuminating the set object, relates to the set object where the in-focus points approximately match the position of the object. When the phase in the exit plane is estimated using phase contrast or defocused images, and this data is used to determine the 3D object, there is a higher error between the phase in the exit plane calculated from the current estimate of the object and the phase calculated from the defocused of phase contrast images.

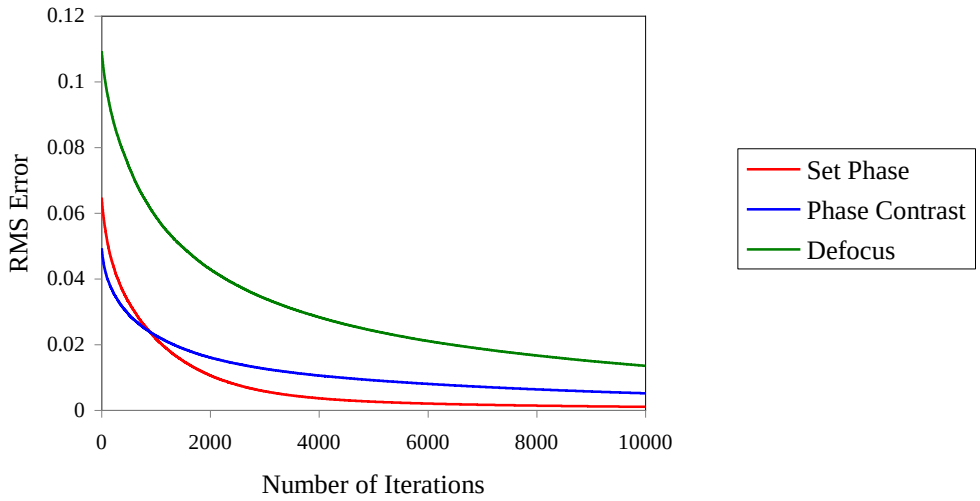


Figure 6.23 RMS error between the phase in the exit plane calculated by illuminating the estimated object and the known or estimated phase calculated from the defocus or phase contrast images

It is likely that the phase calculated from the phase contrast or defocused images does not relate to a phase object that is small and relatively compact and therefore require a larger number of points to represent the object that will generate consistent exit plane data. This figure suggests that the object calculated from the phase contrast images is closer to a small and compact object and is therefore a better approximation of the set object as the phase contrast error curve is closer to the error from the set phase.

This is only a single example but shows the complete process of determining the phase in a single plane from phase contrast or defocused images and the potential for estimating the 3D structure of a phase object.

The example is arbitrarily constructed, as such the integrated phase change along the z axis is large and in some places larger than the phase contrast method should be able to accurately estimate. The system is also modelled to have a large maximum scattering. angle. The ITE is derived under the condition that the scattering angle is small. Additionally the solutions to the ITE assume the intensity is constant in the plane where the phase is to be calculated, in the example this is not the case. Despite being at the limit where these methods can be accurately used this example shows the potential for extracting 3D information.

This process has not been optimised and the choice to use phase contrast and defocus images in the process to calculate the 3D object cannot be made based on a single example.

If this is used with experimental data then there will be a limit on the accuracy of the estimated object and the estimated data in the exit plane due to the presence of noise. Due to noise there will be a limit on the accuracy that can be achieved. This provides a threshold below which the iterative process should be stopped.

6.5. Conclusions

The first half of this chapter considered the possibility of determining a 3D phase object by scanning through the back propagated data to pick out in-focus point in a way analogous to defocusing a microscope. Using the clean algorithm, developed for astronomy to remove the unwanted effect in the imaging process, a similar method was developed that constructs a model of the phase object that is consistent with the data and removes the influence of the different layers on each other.

To reconstruct the object this algorithm relies on a one to one correspondence between the object position and the points with a maximum absolute phase in the back propagated data. When this is not valid, an object will be reconstructed that is consistent with the exit plane data but different from the true object.

The final section of this chapter combined the methods for calculating the phase in a 2D plane described in the previous chapters with the iterative scheme described in this chapter. For a single example it was shown that some three dimensional information can be determined. The phase object calculated from the phase contrast images showed clearer visualisation of the object and slightly better positioning in depth.

It is possible, for simple objects to be reconstructed giving a good structural representation of the object but a quantitative estimation of the phase values is in general to be beyond the capabilities of the iterative scheme, due to the lack of information. In general, there is not enough information from a single view of an object to determine the original 3D structure and so this method will not work in every case but for applications like sperm imaging it may be possible to determine the structure of the tail.

Although this process is limited, it could be improved by including multiple image and prior knowledge of the object. If this process is to be applied to experimental data, further testing through simulation to test the process and optimise it will be required. These areas are beyond the scope of this thesis and remains as an area for future work.

Chapter 7 – Conclusions and Future Work

The aim of this thesis was to improve methods to determine the phase from defocused or phase contrast images. The thesis aimed to understand the approximations and limitations in current methods for calculating phase and consequently improve the calculation process by removing or reducing the approximation. Many factors can influence the calculated phase from both the experiment and the numerical processing of the data. This thesis also aimed to show which of these factors required careful control and it showed which factors could be optimised to improve the estimated phase. The final aim of this thesis was to use the phase calculated in a single plane to determine information about the structure of a three dimensional phase object.

Chapter 2 developed the theory for calculating phase that used a combination of two phase contrast images and a third image with no phase shift for normalisation. It was shown that multiple images required fewer approximations than single image phase contrast and therefore produced a better approximation of the phase. There is a significant reduction in the error, using three images in comparison to one image, by a factor of between 2 and 20 depending on the details of the specific case.

It was also shown that the phase calculated from three images is independent of the phase shift applied to the reference wave. This is particularly useful experimentally as it enables the phase shift to be chosen so that the images can have high contrast without affecting the accuracy of the calculated phase.

A fundamental limitation of this method is that the calculation of the phase is limited to weak phase objects. This is due to requirement that there must be a known reference wave before the determination of the scattered wave is possible. As the reference wave is itself phase dependent, only when the phase change is small can the reference wave be approximately known.

To overcome this limitation, an iterative scheme was developed to refine the estimated phase based on successive linear estimates of the phase error. Repeated use of the derived equations is still limited but the phase coefficient can be increases to approximately one radian and the iterative scheme will still provide an improvement.

Within this limitation, there was an improvement in the phase error by approximately an order of magnitude within several iterations. The iterative scheme has not been proven to converge so there is no guarantee that this process will work in every case.

As a development of this process, it was shown that phase contrast imaging using three images could be expanded to incorporate an arbitrary number of images. Although not considered any further this is a potential area for future work. There is a significant reduction in the phase error by changing from one image to three images and so there may be a further reduction in the error by including more images or else there may be applications where the additional images give other benefits.

Chapter 3 considered calculating the phase from defocused images using the intensity transform equation (ITE). It has been shown, in general, that this equation can be solved using a Green's function, Woods, *et al.* [30]. For the specific case of a circular aperture illuminated with constant intensity, there are three different Green's functions stated in the literature, without proof, that claim to solve this problem.

The chapter showed that two of these Green's functions were correct, Stakgold [46], Trim [47], and that the third was wrong as it fails to satisfy all the necessary conditions for a valid solution, Zhu, *et al.* [48]. The two correct Green's functions were shown to calculate equivalent phases that differ by a constant.

A second approach to solving the ITE was also considered using Fourier transforms. This method is considered as there is no condition requiring the phase to be continuous. The Fourier transform method has been partly applied by Dorrer, *et al.* [27] and Allen, *et al.* [28] where the boundary conditions were not included by and by Diaz-Douton, *et al.* [24] where the boundary data is included iteratively. By excluding the boundary conditions an entire set of functions are missed that can only be calculated by considering the boundary conditions. Therefore, a full analytic solution was derived that calculates the contributions of these functions to the phase, giving a full phase solution.

The application of these methods to experimental data was described. The implementation of the Green's function required the evaluation of singular point. The contributions from these points have previously been set to zero, Teague [49]. It was shown that by integrating the Green's function over the area of each pixel the

singularities can be removed giving well defined values that can be used in the numerical process. Although this process removes the singularities, the increase in accuracy is minimal in return for the computational time to perform the calculation. As a compromise, the simple numerical integration should be used for non singular points and the analytically integrated Green's function should be used for the singular points. In this way, the accuracy is improved without severely increasing the computational expense.

In Chapter 4, the data from two experiments was analysed using the Fourier transform method. The first set of experimental data showed a consistent relative change in defocus as the source was translated but the absolute value was different due to a constant offset. This is not likely to be due to the analysis and is probably due to miss calibration of the experiment when defining the position where the defocus phase change is zero. As well as defocus, there were other modes detected in the analysis. The accuracy of these modes is unknown but they appear consistent with the data. The reconstructed estimate of the phase from the second experiment was in good agreement with the expected results; although it is unknown which, if any, are correct. This shows that the analysis process can be applied to experimental data as well simulated data.

Having developed the theory and shown that it can be applied to experimental data Chapter 5 considered optimising the experimental process and showed the sensitivity of the estimated phase to factors common to all experiments using defocused images and this type of analysis.

Each parameter that can be varied, to improve the estimated phase, will usually be at the expense of another. The chapter showed that increasing the resolution of the data improved the accuracy of the phase but the time for the calculation will also increase. It is overall not beneficial to include more images and the accuracy of the approximation of the derivative mainly depends on the plane spacing.

The factors that influence the choice between analysing the data using a Green’s function or the Fourier transform method are accuracy and computation time. In general, the accuracy is similar for both methods as this depends in the resolution and the phase, but there is a considerable difference in the computation time. The Fourier transform method, depending on the size of the grid, can be orders of magnitude, faster

than the Greens function therefore this method is preferable. The Fourier transform method also has the potential to be applied in case where the aperture is not a disk, and it does not require the phase to be continuous.

Chapter 6 considered the possibility of using the phase information calculated using the methods from the previous chapters, to determine the structure of a 3D phase object. The first half of the chapter developed an iterative algorithm that used the back propagated complex amplitude and "picked out" the in-focus points to construct the object. There is a known issue of uniqueness with this approach so an object can be constructed with little resemblance to the original set object but still generate a good estimation of the data in the exit plane.

The algorithm depends on mapping the in-focus points to the position of the phase object. When the object is locally compact so that the focusing effect is minimal, then a reasonably good estimation of the structure of the object was calculated. As the object gets larger, causing the focal points to move away from the position of the object, the mapping between the in-focus points and the position of the object breaks down. When this happens, the iterative scheme cannot be used to determine even the approximate structure of the object.

It was shown for a single example that the phase in the exit plane can be estimated from a set of phase contrast images or a set of defocused images and from the calculated phase, that three dimensional objects can be estimated. This shows the potential for extracting some structural information about the object from a set of images taken simultaneously and that are experimentally simple to make.

This is likely to be applicable to simple structures for example single cells and sperm tails. The head of the sperm is a strong phase object and may be difficult to reconstruct but determining the structure and position of the tail may be possible.

7.1. Future Work

The work in Chapter 6 only produced a method for determining the structure of a limited set of 3D object but this is still an area of interest particularly, when the specimen is moving. Determining the 3D object as it moves in real time or making the measurements in real time, which can be processed after, is still an area for further

research. Further work in this area could develop methods for simultaneously viewing the specimen from an increased number of angles either completely encircling the object or over a limited range of angles and the development of a suitable process for analysing the data, that incorporates prior knowledge.

Appendix A – Second Order Approximation of the ITE

In the derivation of the intensity transport equation (ITE) there is an first order approximation applied to the equation (3.13). If all the additional terms are retained then

$$k\frac{\partial I(\mathbf{x},z)}{\partial z}=-\frac{i}{2}\left(\psi^*(\mathbf{x},z)\nabla^2\psi(\mathbf{x},z)-\psi(\mathbf{x},z)\nabla^2\psi^*(\mathbf{x},z)\right)$$
$$+ik^2\sum_{n=2}^{\infty}\frac{(-1)^n(2n)!\lambda^{2n}}{(2n-1)(n!)^2\,4^n(2\pi)^{2n}}\sum_{j=1}^{2}\left\{\psi^*(\mathbf{x},z)\frac{\partial^{2n}\psi(\mathbf{x},z)}{\partial x_j^{2n}}-\psi(\mathbf{x},z)\frac{\partial^{2n}\psi^*(\mathbf{x},z)}{\partial x_j^{2n}}\right\}. \tag{A.1}$$

The higher order derivatives in the final summation are taken with respect to variable x_j, where $\mathbf{x}=(x_1,x_2)$ in Cartesian coordinates.

The ITE is only a first order approximation of the axial derivative, as the phase coefficients becomes larger and the scattering angles increase the higher order terms will become important. This appendix shows the effect of retaining extra terms, from the series expansion of the square root (3.14). The sum that accounts for all the higher order structure which was previously neglected is

$$\nu(\mathbf{x},z)=ik^2\sum_{n=2}^{\infty}\frac{(-1)^n(2n)!\lambda^{2n}}{(2n-1)(n!)^2\,4^n(2\pi)^{2n}}\sum_{j=1}^{2}\left\{\psi^*(\mathbf{x},z)\frac{\partial^{2n}\psi(\mathbf{x},z)}{\partial x_j^{2n}}-\psi(\mathbf{x},z)\frac{\partial^{2n}\psi^*(\mathbf{x},z)}{\partial x_j^{2n}}\right\}. \tag{A.2}$$

Consider, a second order approximation where $n=2$ then the first correcting term is

$$\nu_2(\mathbf{x},z)=\frac{i}{8k^2}\sum_{j=1}^{2}\left\{\psi^*(\mathbf{x},z)\frac{\partial^4\psi(\mathbf{x},z)}{\partial x_j^4}-\psi(\mathbf{x},z)\frac{\partial^4\psi^*(\mathbf{x},z)}{\partial x_j^4}\right\}. \tag{A.3}$$

Substituting in the expression for the complex amplitude, $\psi(\mathbf{x},z)=\sqrt{I(\mathbf{x},z)}\exp\left(-i\phi(\mathbf{x},z)\right)$, without loss of generality the plane where for the phase is reconstructed can be defined as $z=0$ therefore

$$\nu_2(\mathbf{x})=\frac{i}{8k^2}\sqrt{I(\mathbf{x})}\sum_{j=1}^{2}\left\{\begin{array}{c}\exp\left(i\phi(\mathbf{x})\right)\dfrac{\partial^4\sqrt{I(\mathbf{x})}\exp\left(-i\phi(\mathbf{x})\right)}{\partial x_j^4}\\ -\exp\left(-i\phi(\mathbf{x})\right)\dfrac{\partial^4\sqrt{I(\mathbf{x})}\exp\left(i\phi(\mathbf{x})\right)}{\partial x_j^4}\end{array}\right\}. \tag{A.4}$$

This can be evaluated, dropping the arguments for $\phi(\mathbf{x})$ and $I(\mathbf{x})$, to give

$$\nu_2(\mathbf{x},z)=\frac{1}{8k^2I^2}\sum_{j=1}^{2}\left\{\begin{array}{l}2I^3\frac{\partial^4\phi}{\partial x_j^4}-3I\left(\frac{\partial I}{\partial x_j}\right)^2\frac{\partial^2\phi}{\partial x_j^2}-6I\frac{\partial I}{\partial x_j}\frac{\partial^2 I}{\partial x_j^2}\frac{\partial\phi}{\partial x_j}+4\frac{\partial^3\phi}{\partial x_j^3}\frac{\partial I}{\partial x_j}I^2+6\frac{\partial^2\phi}{\partial x_j^2}\frac{\partial^2 I}{\partial x_j^2}I^2\\-12\frac{\partial^2\phi}{\partial x_j^2}\left(\frac{\partial\phi}{\partial x_j}\right)^2I^3+4\frac{\partial\phi}{\partial x_j}\frac{\partial^3 I}{\partial x_j^3}I^2-4I^2\left(\frac{\partial\phi}{\partial x_j}\right)^3\frac{\partial I}{\partial x_j}+3\frac{\partial\phi}{\partial x_j}\left(\frac{\partial I}{\partial x_j}\right)^3\end{array}\right\}. \tag{A.5}$$

The modified intensity transport equation is therefore

$$\begin{aligned}&I\left\{\frac{\partial^2\phi}{\partial x_j^2}+\frac{1}{4k^2}\frac{\partial^4\phi}{\partial x_j^4}-\frac{3}{2k^2}\frac{\partial^2\phi}{\partial x_j^2}\left(\frac{\partial\phi}{\partial x_j}\right)^2\right\}+\frac{\partial\phi}{\partial x_j}\frac{\partial I}{\partial x_j}\\&-k\frac{\partial I(\mathbf{x},z)}{\partial z}\approx\sum_{j=1}^{2}+\frac{1}{8k^2I^2}\left\{\begin{array}{l}+4\frac{\partial\phi}{\partial x_j}\frac{\partial^3 I}{\partial x_j^3}I^2-4I^2\left(\frac{\partial\phi}{\partial x_j}\right)^3\frac{\partial I}{\partial x_j}+3\frac{\partial\phi}{\partial x_j}\left(\frac{\partial I}{\partial x_j}\right)^3-3I\left(\frac{\partial I}{\partial x_j}\right)^2\frac{\partial^2\phi}{\partial x_j^2}\\-6I\frac{\partial I}{\partial x_j}\frac{\partial^2 I}{\partial x_j^2}\frac{\partial\phi}{\partial x_j}+4\frac{\partial^3\phi}{\partial x_j^3}\frac{\partial I}{\partial x_j}I^2+6\frac{\partial^2\phi}{\partial x_j^2}\frac{\partial^2 I}{\partial x_j^2}I^2\end{array}\right\}.\end{aligned} \tag{A.6}$$

This has a similar form as (3.21), so for a uniformly illuminated aperture the first term on the top line defines the “curvature” part of the phase and the remaining terms define the “non-curvature” part of the phase. This is now a fourth order, nonlinear partial differential equation. The corrective terms are very small and scale as $1/k^{2n}$, as shown in (A.2). In comparison to the ITE terms the benefits of extending the approximation to second order is negligible. Finally, the additional complexity from this second term is unlikely to be tractable to produce a closed form analytic solution. Therefore, when the first order approximation is no longer valid an alternative method is required.

Appendix B – Comparison of Terms in the Green's function and Fourier Transform Solution

The solution to the differential equation using Fourier transforms can be shown to be equivalent to part of the Green's function solution. The free space Green's function, equation (3.76), can be written in vector form as

$$U\left(\mathbf{x},\mathbf{x}'\right)=\frac{1}{2\pi}\log\left|\mathbf{x}-\mathbf{x}'\right|. \tag{B.1}$$

By definition $\nabla^2 U\left(\mathbf{x},\mathbf{x}'\right)=\delta\left(\mathbf{x}-\mathbf{x}'\right)$ and from equation (3.23) it follows directly that

$$\begin{aligned}\phi\left(\mathbf{x}\right)=&\int_A U\left(\mathbf{x},\mathbf{x}'\right)\nabla^2\phi\left(\mathbf{x}'\right)d^2\mathbf{x}'+\int_A G_C\left(\mathbf{x},\mathbf{x}'\right)\nabla^2\phi\left(\mathbf{x}'\right)d^2\mathbf{x}'\\&-\oint_C G\left(\mathbf{x},\mathbf{x}'\right)\nabla\phi\left(\mathbf{x}'\right)\cdot d\hat{\mathbf{n}}'.\end{aligned} \tag{B.2}$$

where $G_C\left(\mathbf{x},\mathbf{x}'\right)$ represents the remaining terms in the Green's function not including the free space Green's function. The first area integral is also a convolution and it can be shown Chan, *et al.* [79] that

$$\mathfrak{F}^{-1}\left\{-\frac{1}{4\pi^2\left|\boldsymbol{\xi}\right|^2}\right\}=\frac{1}{2\pi}\log\left|\mathbf{x}\right|, \tag{B.3}$$

so the solutions to the differential equation part of the ITE using the free space Green's function and Fourier transform solution are equivalent.

References

1. G. Popescu, T. Ikeda, C. A. Best, K. Badizadegan, R. R. Dasari, and M. S. Feld, "Erythrocyte structure and dynamics quantified by Hilbert phase microscopy," Journal of Biomedical Optics **10**, 060503 (2005).
2. C. Campbell, "Wave-front sensing by use of a Green's function solution to the intensity transport equation: comment," J. Opt. Soc. Am. A **24**, 2480-2481 (2007).
3. F. Roddier, "Curvature sensing and compensation: a new concept in adaptive optics," Applied Optics **27**, 1223-1225 (1988).
4. E. Hecht, "Optics," (Addison-Wesley, 1998).
5. E. L. Lago and R. de la Fuente, "Wavefront sensing by diffracted beam interferometry," Journal of Optics A: Pure and Applied Optics **4**, 299-302 (2002).
6. F. Zernike, "Das Phasenkontrastverfahren bei der mikroskopischen Beobachtung," Z. Tech. Phys **16**, 454 (1935).
7. M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation Interference and Diffraction of Light*, 6th ed. (Cambridge University Press, 1997).
8. C. R. Mercer and K. Creath, "Liquid-crystal point-diffraction interferometer for wave-front measurements," Appl. Opt **35**, 1633–1642 (1996).
9. R. M. Neal and J. C. Wyant, "Polarization phase-shifting point-diffraction interferometer," Appl. Opt. **45**, 3463-3476 (2006).
10. N. Lue, W. Choi, G. Popescu, T. Ikeda, R. R. Dasari, K. Badizadegan, and M. S. Feld, "Quantitative phase imaging of live cells using fast Fourier phase microscopy," Applied Optics **46**, 1836-1842 (2007).
11. G. Popescu, L. P. Deflores, J. C. Vaughan, K. Badizadegan, H. Iwai, R. R. Dasari, and M. S. Feld, "Fourier phase microscopy for investigation of biological structures and dynamics," Optics Letters **29**, 2503-2505 (2004).
12. P. J. Rodrigo, D. Palima, and J. Glückstad, "Accurate quantitative phase imaging using generalized phase contrast," Optics Express **16**, 2740-2751 (2008).
13. B. C. Platt, "History and principles of Shack-Hartmann wavefront sensing," Journal of Refractive Surgery **17**, 573-577 (2001).
14. P. Hariharan, "Lateral and radial shearing interferometers: a comparison," Applied Optics **27**, 3594-3596 (1988).
15. W. Zou and Z. Zhang, "Generalized wave-front reconstruction algorithm applied in a Shack-Hartmann test," Applied Optics **39**, 250-268 (2000).
16. R. K. Tyson, *Principles of Adaptive Optics*, 2nd ed. (Academic Press, 1998).
17. G. Rousset, "Wavefront sensing," Adaptive optics for astronomy **423**, 115–137 (1994).
18. R. A. Gonsalves, "Phase retrieval and diversity in adaptive optics," Optical Engineering **21**, 829-832 (1982).
19. S. M. Jefferies, M. Lloyd-Hart, E. K. Hege, and J. Georges, "Sensing wave-front amplitude and phase with phase diversity," Applied optics **41**, 2095-2102 (2002).
20. L. J. Allen, M. P. Oxley, and D. Paganin, "Computational Aberration Correction for an Arbitrary Linear Imaging System," Physical Review Letters **87**, 123902 (2001).
21. T. C. Petersen and V. J. Keast, "Astigmatic intensity equation for electron microscopy based phase retrieval," Ultramicroscopy **107**, 635-643 (2007).
22. D. L. Misell, "An examination of an iterative method for the solution of the phase problem in optics and electron optics: I. Test calculations," Journal of Physics D: Applied Physics **6**, 2200-2216 (1973).

23. H. I. Campbell, S. Zhang, A. H. Greenaway, and S. Restaino, "Generalized phase diversity for wave-front sensing," Optics Letters **29**, 2707-2709 (2004).
24. F. Diaz-Douton, J. Pujol, M. Arjona, and S. O. Luque, "Curvature sensor for ocular wavefront measurement," Optics letters **31**, 2245-2247 (2006).
25. T. E. Gureyev and K. A. Nugent, "Phase retrieval with the transport-of-intensity equation. II. Orthogonal series solution for nonuniform illumination," J. Opt. Soc. Am. A **13**, 1670-1682 (1996).
26. T. E. Gureyev, A. Roberts, and K. A. Nugent, "Phase retrieval with the transport-of-intensity equation: matrix solution with use of Zernike polynomials," JOSA A **12**, 1932 (1995).
27. C. Dorrer and J. D. Zuegel, "Optical testing using the transport-of-intensity equation," Opt. Express **15**, 7165-7175 (2007).
28. L. J. Allen and M. P. Oxley, "Phase retrieval from series of images obtained by defocus variation," Optics Communications **199**, 65-75 (2001).
29. D. J. Lee, M. C. Roggemann, B. M. Welsh, and E. R. Crosby, "Evaluation of least-squares phase-diversity technique for space telescope wave-front sensing," Applied optics **36**, 9186-9197 (1997).
30. S. C. Woods and A. H. Greenaway, "Wave-front sensing by use of a Green's function solution to the intensity transport equation," Journal of the Optical Society of America A **20**, 508-512 (2003).
31. L. Waller, L. Tian, and G. Barbastathis, "Transport of Intensity phase-amplitude imaging with higher order intensity derivatives," Opt. Express **18**, 12552-12561.
32. N. Streibl, "Three-dimensional imaging by a microscope," J. Opt. Soc. Am. A **2**, 121-127 (1985).
33. R. Danev and K. Nagayama, "Single particle analysis based on Zernike phase contrast transmission electron microscopy," Journal of Structural Biology **161**, 211-218 (2008).
34. G. Popescu, T. Ikeda, R. R. Dasari, and M. S. Feld, "Diffraction phase microscopy for quantifying cell structure and dynamics," Optics Letters **31**, 775-777 (2006).
35. A. Y. M. Ng, C. W. See, and M. G. Somekh, "Quantitative optical microscope with enhanced resolution using a pixelated liquid crystal spatial light modulator," Journal of microscopy **214**, 334-340 (2004).
36. C. Maurer, A. Jesacher, S. Bernet, and M. Ritsch-Marte, "Phase contrast microscopy with full numerical aperture illumination," Opt. Express **16**, 19821-19829 (2008).
37. J. W. Goodman, *Introduction to Fourier optics* (Roberts & Co, 2005).
38. E. O. Brigham, *The fast Fourier transform and its applications* (Prentice Hall Englewood Cliffs, NJ, 1988).
39. J. Glückstad and P. C. Mogensen, "Optimal phase contrast in common-path interferometry," Appl. Opt **40**, 268–282 (2001).
40. M. Beleggia, "A formula for the image intensity of phase objects in Zernike mode," Ultramicroscopy **108**, 953-958 (2008).
41. P. M. Blanchard, D. J. Fisher, S. C. Woods, and A. H. Greenaway, "Phase-diversity wave-front sensing with a distorted diffraction grating," Applied optics **39**, 6649 (2000).
42. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt **21**, 2758-2769 (1982).
43. J. C. He, S. Marcos, R. H. Webb, and S. A. Burns, "Measurement of the wave-front aberration of the eye by a fast psychophysical procedure," J. Opt. Soc. Am. A **15**, 2449-2456 (1998).

44. J. Schwiegerling, J. E. Greivenkamp, and J. M. Miller, "Representation of videokeratoscopic height data with Zernike polynomials," Journal of the Optical Society of America A **12**, 2105-2113 (1995).
45. C. S. Anderson, "Fringe visibility, irradiance, and accuracy in common path interferometers for visualization of phase disturbances," Appl. Opt **34**, 7474–7485 (1995).
46. I. Stakgold, *Green's functions and boundary value problems* (John Wiley & Sons, Inc, 1998).
47. D. W. Trim, *Applied Partial Differential Equations* (PWS, 1990).
48. B. Zhu, A. Shirakawa, M. Musha, K. Ueda, and K. D. Cole, "Wavefront Curvature Sensing Based on Diffraction Grating and Fractional Fourier Transforms," Optical Review **11**, 344-347 (2004).
49. M. R. Teague, "Deterministic phase retrieval: a Green's function solution," J. Opt. Soc. Am **73**, 1434-1441 (1983).
50. R. N. Bracewell, "*The Fourier Transform and its Applications*," McGraw-Hill New York (1978).
51. R. S. Strichartz, *A Guide to Distribution Theory and Fourier Transforms* (CRC-Press, 1994).
52. G. F. Roach, *Green's functions* (Cambridge Univ Pr, 1982).
53. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes: the art of scientific computing* (Cambridge Univ Pr, 2007).
54. P. M. Blanchard and A. H. Greenaway, "Simultaneous Multiplane Imaging with a Distorted Diffraction Grating," Appl. Opt. **38**, 6692-6699 (1999).
55. S. Djidel, J. K. Gansel, H. I. Campbell, and A. H. Greenaway, "High-speed, 3-dimensional, telecentric imaging," Opt. Express **14**, 8269-8277 (2006).
56. R. G. Paxman, T. J. Schulz, and J. R. Fienup, "Joint estimation of object and aberrations by using phase diversity," Journal of the Optical Society of America A **9**, 1072-1085 (1992).
57. N. Angarita-Jaimes, E. McGhee, M. Chennaoui, H. I. Campbell, S. Zhang, C. E. Towers, A. H. Greenaway, and D. P. Towers, "Wavefront sensing for single view three-component three-dimensional flow velocimetry," Experiments in Fluids **41**, 881-891 (2006).
58. E. Acosta, S. Ri´os, M. Soto, and V. V. Voitsekhovich, "Role of boundary measurements in curvature sensing," Optics Communications **169**, 59-62 (1999).
59. D. Paganin, A. Barty, P. J. McMahon, and K. A. Nugent, "Quantitative phase-amplitude microscopy. III. The effects of noise," Journal of Microscopy **214**, 51-61 (2004).
60. C. G. Rylander, D. P. Davé, T. Akkin, T. E. Milner, K. R. Diller, and A. J. Welch, "Quantitative phase-contrast imaging of cells with phase-sensitive optical coherence microscopy," Optics Letters **29**, 1509-1511 (2004).
61. Y. Sung, W. Choi, C. Fang-Yen, K. Badizadegan, R. R. Dasari, and M. S. Feld, "Optical diffraction tomography for high resolution live cell imaging," Optics Express **17**, 266-277 (2009).
62. A. J. Devaney, "Nonuniqueness in the inverse scattering problem," Journal of Mathematical Physics **19**, 1526 (1978).
63. A. J. Devaney, "Reconstructive tomography with diffracting wavefields," Inverse Problems **2**, 161-183 (1986).
64. A. J. Devaney, "Geophysical diffraction tomography," IEEE Transactions on Geoscience and Remote Sensing, 3-13 (1984).
65. A. J. Devaney, "The limited-view problem in diffraction tomography," Inverse problems **5**, 501-521 (1989).

66. R. P. Millane, "Phase retrieval in crystallography and optics," J. Opt. Soc. Am. A **7**, 394-411 (1990).
67. D. J. Stephens and V. J. Allan, "Light Microscopy Techniques for Live Cell Imaging," Science **300**, 82-86 (2003).
68. S. Kikuchi, K. Sonobe, S. Mashiko, Y. Hiraoka, and N. Ohyama, "Three-dimensional image reconstruction for biological micro-specimens using a double-axis fluorescence microscope," Optics Communications **138**, 21-26 (1997).
69. J. Sharpe, U. Ahlgren, P. Perry, B. Hill, A. Ross, J. Hecksher-Sorensen, R. Baldock, and D. Davidson, "Optical projection tomography as a tool for 3D microscopy and gene expression studies," Science **296**, 541 (2002).
70. M. K. Kreysing, T. Kiessling, A. Fritsch, C. Dietrich, J. R. Guck, and J. A. Käs, "The optical cell rotator," Opt. Express **16**, 16984-16992 (2008).
71. F. Charrière, A. Marian, F. Montfort, J. Kuehn, T. Colomb, E. Cuche, P. Marquet, and C. Depeursinge, "Cell refractive index tomography by digital holographic microscopy," Opt. Lett. **31**, 178-180 (2006).
72. F. Charrière, N. Pavillon, T. Colomb, C. Depeursinge, T. J. Heger, E. A. D. Mitchell, P. Marquet, and B. Rappaz, "Living specimen tomography by digital holographic microscopy: morphometry of testate amoeba," Opt. Express **14**, 7005-7013 (2006).
73. C. J. R. Sheppard and X. Q. Mao, "Three-dimensional imaging in a microscope," J. Opt. Soc. Am. A **6**, 1260-1269 (1989).
74. F. E. Giesecke, A. Mitchell, H. C. Spencer, and I. L. Hill, *Technical Drawing*, 5th ed. (The Macmillan Company, 1967).
75. J. Â L. Starck, E. Pantin, and F. Murtagh, "Deconvolution in Astronomy: A Review," Publications of the Astronomical Society of the Pacific **114**, 1051-1069 (2002).
76. J. B. Pawley, R. J. Palmer, J. A. J. Haagensen, T. R. Neu, and C. Sternberg, "Confocal Microscopy of Biofilms — Spatiotemporal Approaches," in *Handbook of Biological Confocal Microscopy* (Springer US, 2006), pp. 870-888.
77. G. Marshall, *Optical Scanning* (Dekker, 1991).
78. Olympus, "http://industrial-microscope.olympus-global.com/en/ga/product/uis2.cfm," Accessed 30/9/10.
79. Y. S. Chan, L. J. Gray, T. Kaplan, and G. H. Paulino, "Green's function for a two-dimensional exponentially graded elastic medium," Proceedings of the Royal Society of London-A **460**, 1689-1706 (2004).