Interferometric Metrology Using Reprogrammable Binary Holograms

MATTHEW THOMAS DOLPHIN CASHMORE

Interferometric Metrology Using Reprogrammable Binary Holograms

Matthew T. Cashmore

A Thesis presented for the degree of
Doctor of Philosophy

Centre for Advanced Instrumentation
Department of Physics
Durham University
England

April 2013

Interferometric Metrology Using Reprogrammable Binary Holograms

Matthew T. Cashmore

Submitted for the degree of Doctor of Philosophy

April 2013

Abstract

Interferometric methods for surface metrology have been widely used for many years due to their speed, accuracy and versatility. It is frequently necessary however to produce a known comparison reference surface to minimise the optical path difference and hence enhance the dynamic range. An alternative to this is to use a computer generated hologram to act as the reference wave, or to correct a spherical reference wave to match a highly aspheric optic in order to achieve a null test.

This thesis shall present a novel method of producing such holograms through the use of a binary ferroelectric liquid crystal on silicon spatial light modulator (FLCOS SLM) rather than using the more common lithographically produced plates. One of the primary advantages this could introduce is the ability for arbitrarily reprogrammable holograms to be created upon demand rather than needing to produce a series of holographic plates, saving both time and money in the testing of surfaces. We present results characterising the ability of a FLCOS SLM to produce increasingly large Zernike aberrations as well as quantifying the resulting errors, before using the device to reduce interferometric fringe density allowing us to measure aberrated optics and reveal low amplitude surface variations on the scale of 0.045 waves RMS.

Declaration

The work in this thesis is based on research published by Mark Neil[1] and contains work carried out at the Centre for Advanced Instrumentation, Department of Physics, University of Durham, England, and the National Physical Laboratory(NPL), Teddington. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is the sole work of the author unless referenced to the contrary here or in the main text.

The support of the National Measurement System for this CASE studentship is gratefully acknowledged.

Publications

M.T. Cashmore, S.R.G Hall, G. D. Love. Interferometry using reconfigurable binary holograms for accurate wavefront manipulation. *In preparation.*

Acknowledgements

I would firstly like to thank my supervisor, Gordon Love, for giving me the opportunity to carry out this work and for all the help he has provided over the last four years, in particular his encouragement, advice, and most of all his patience given during the period of writing this thesis up. I would also like to thank Simon Hall at NPL for the invaluable assistance provided during my time working there without which I would not have anywhere near the level of confidence and valuable results as I do now.

Huge thanks go to those in Durham who have provided me with advice, assistance and technical support during the course of my PhD, most notably Tim Morris, Tim Butterly and Chris Saunter. I have also greatly appreciated the members of CfAI who have kept me sane during the time spent in Durham and special thanks to those I have shared an office with; Rob Harris, Penny Lawton, David Nandi, James Osborn, Cyril Bourgenot, Dora Fohring, Scott Silburn, Luke Tyas, Claire Poppett and Andrew Reeves.

I also wish to thank those at NPL and in London who helped greatly, both professionally and socially, during my time there; Kevin O'Holleran, Mike Shaw, Dan Gittings, Pippa Goldenberg, Lloyd England, Shelly Sharma and Frederica Byron.

I greatly appreciate all the support and understanding shown by my friends

and family throughout the course of my PhD, particularly during the last year. I am especially grateful for the tolerance of my housemate, Robbie Glendinning, for putting up with the way I've taken over the living room in our house with thesis material for over half a year, as well as for his assistance in proofreading. I also want to express particular thanks to Caroline Walters for her company during the writing of this thesis, preventing me from becoming a complete hermit and most of all, pans of wine. Further thanks to Neil Robinson, Charlotte Ellis and Rosie Adams for keeping me going during the incredibly brief and rare moments of free time I've been able to see them in.

I wish to make a special acknowledgement to Alex Lloyd-Ribeiro for the many distractions and amusements he has provided over the last 8 years. In particular for evenings down the Bongo, assistance in dentistry and most of all those nights we spent being amused and horrified whilst lost in Tokyo.

Final and most heartfelt thanks go to Jennifer Padgett, whose love and support has quite literally got me through this last year.

Contents

List of Figures

List of Abbreviations

- *CCD* Charge coupled device
- *CGH* Computer generated hologram
- *E ELT* European extremely large telescope
- *FLC* Ferroelectric liquid crystal
- *FLCOS* Ferroelectric liquid crystal on silicon
- *HeNe* Helium Neon
- *LC* Liquid crystal
- *LCD* Liquid crystal display
- *LVDT* Linear variable differential transformer
- *NPL* National Physical Laboratory
- *OPD* Optical path difference
- *PSF* Point spread function
- *PSI* Phase shifting interferometry
- *PZS* Piezoelectric stage

- *RMS* Root mean square
- *SLM* Spatial light modulator
- *UV* Ultraviolet

Chapter 1

Introduction

The understanding and characterisation of surfaces has been important in physics for hundreds of years. Long before the need for nanoscale accuracy or the advent of modern optics, the accurate knowledge of the properties of surfaces has been vital to the advancement of understanding and society. Examples of this can be seen dating back well over three thousand years in the remains of dry stone constructions built by ancient civilisations. An area which was for many years handicapped by poor surface measurement and production was the field of ballistics. Before the development of rifling the accuracy of firearms was severely limited [2] as the mismatch between the shape and size of the bore and the size of the shot caused it to ricochet around inside the barrel leading to unpredictable trajectories.

The mechanical efficiency of devices can be crippled by poor quality and inaccurately made parts, and the understanding of friction processes is dependant on the knowledge of the properties of the contact area between two surfaces. With the Industrial Revolution and the increased dependancy of the World on complicated machinery, the ability to produce mechanical parts accurately and

repeatably became vital so as not to hinder the operation of the plethora of new technology being developed and incorporated into daily life. One of the most influential developments during the Industrial Revolution was the concept of interchangeability of parts[3]. The main idea of this is that when a part wears or breaks it can easily be replaced as it has been produced to a specification, and any similar parts can be substituted within the mechanism without the need for modifications. This is only possible with the accurate measurement of master samples so that the parts produced from one manufacturing machine have a functionally identical form to those made from a second. With the development of modern optics there came a new reason to require the knowledge of surface form. Highly rough surfaces in optics cause vast quantities of scatter and prevent the user from obtaining meaningful data, and incorrect surface forms introduce aberrations into a system modifying the properties of the output entirely.

One example of a much more modern application where high quality surface metrology is vital to ensure correct performance is in the analysis of telescope primary mirrors. The most modern generation of planned telescopes shall be significantly larger than any previous ones as a larger primary mirror collects more light and allows astronomers to detect fainter images than ever before. These include the Thirty Metre Telescope[4], the Giant Magellan Telescope with a primary mirror of 25m [5] and the 39.3m primary mirror of the European Extremely Large Telescope(E-ELT) [6]. The huge size of these primary mirrors has led to the need for a segmented design as it is unfeasible to produce a single monolithic reflector of the required scale. This however leads to new complications in the necessary metrology procedure. The E-ELT primary mirror, for example, is comprised of 798 1.4m hexagonal segments to form the f/0.93 39.3m surface, each one of which needs thorough and comprehensive testing to ensure that the maximum wavefront error of a segment not exceed 30nm RMS.

The overall error specifications for the mirror segments is highly complicated, but this value presents a simplified maximum error as given in the E-ELT construction proposal[7]. This necessitates that the metrology involved not only be highly accurate, i.e. that the measurement of the surface must be as close as possible to the actual surface form, but also fast enough to cope with the large number of parts needing to be tested. Furthermore many of the individual mirror segments are different, they belong to "families" with similar surface shapes depending on where in the primary mirror design they are located, so any metrology procedure which incorporates the use of nulling optics must use a uniquely designed nulling lens for each different family of segments.

It is not just in the field of astronomy and optics that highly accurate surface metrology is desired however. As the average life expectancy goes up, so too does the number of body parts being replaced by artificial equivalents. With hip joints for example, it is crucial that an artificial socket has the correct concave surface form to avoid the pain and discomfort caused by a poor fit to the biological joint. Furthermore a minimum surface roughness is critical as not only will this reduce the rate at which the joint is worn, and hence increase the time between replacements, but wear and tear on the artificial joint will release particles into the body causing various detrimental effects depending on material ranging from severe immune reactions to cancers [8].

In measuring a manufactured surface there are four terms generally used to describe the features seen. Lay refers to the predominant orientation of major tooling marks left by the machining process, for example the concentric rings seen in some diamond machined products. The remaining three; roughness, waviness and form can be visually depicted as in figure 1.1 and described as follows. Surface texture, also known as roughness[9], is the high frequency variations in surface height of the form under measurement and plays an important

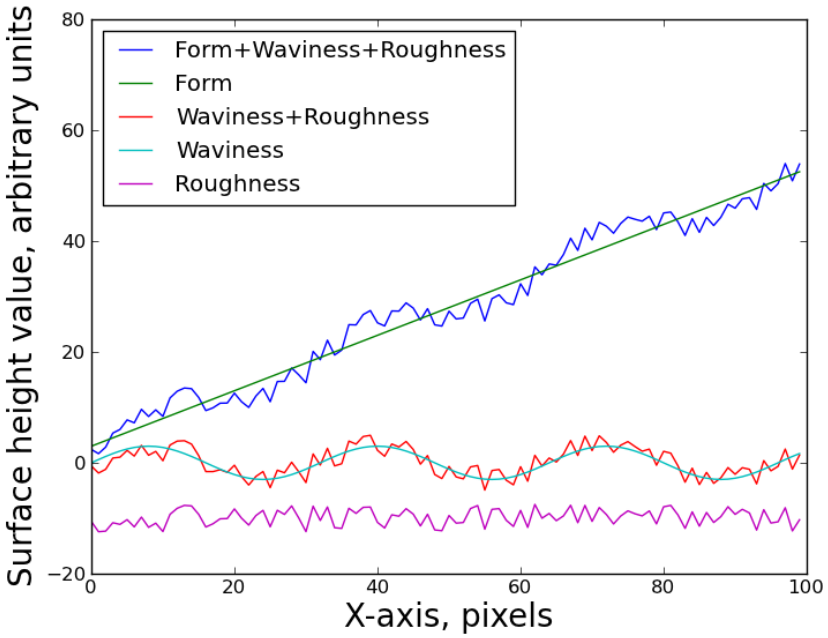


Figure 1.1: A visual representation of three of the basic surface form parameters used in manufacturing. Form refers to the overall shape of the surface, here a sloped surface, roughness corresponds to high frequency variations in surface height and waviness describes the low frequency deviations from the form the surface displays. Lay is not included here as its form and orientation of marks are characteristic of the manufacturing process used. Here the absolute value on the Y scale is unimportant, and included for reference

part in the surface interactions with the rest of the system. A rough object will have a much higher coefficient of friction than a smooth one, and will wear quickly and produce heat in situations where it is contacted with another component. In optics a rough object scatters light undesirably and renders the surface useless when interacting with small wavelength light such as UV. Roughness generally occurs due to manufacturing techniques although it may also arise from unique properties of the material being shaped. Waviness is the description of the behaviour of the surface shape at longer wavelengths than roughness and can arise due to variations in the environment during production, heat

treatments or by the size of the tool used in manufacturing.The overall surface shape, ignoring roughness and waviness, is called the form. Whilst roughness and waviness may reduce the efficiency of the surface or affect its performance adversely, errors in form cause serious problems. Parts with form deviations may not fit properly in the environment they are designed to work, with effects such as mechanical jamming or, when present in optics, form errors completely change the characteristics of the resultant wavefront, adding in prominent aberrations. Furthermore, over time parts degrade which will also alter its form. It is important to understand the behaviour of this wear over time as it may result in serious errors. An example of this can be seen in the diamond machining of optics. As the tool wears it deviates from initial shape, and the path in space it traces out over the production optic is no longer that which is desired.

Often in metrology the measurement of a surface is limited by the dynamic range of the system being employed. For example when measuring an aspheric surface using an interferometer we cannot use a flat reference surface, as the aspheric deviation would be huge and prevent us from observing a fringe pattern. In this case a common approach is to use a spherical reference surface and measure the deviations from sphericity that we find in the test piece. This is not always sufficient however and further elements may be required to modify the wavefront in order to aid metrology.

This thesis presents an investigation into a novel means of wavefront manipulation, using a binary ferroelectric liquid crystal on silicon (FLCOS) spatial light modulator (SLM) to create arbitrary and reprogrammable holograms which can be used in interferometry to null aberrations. This is a pixellated device which uses a ferroelectric liquid crystal as a quarter wave plate in reflection mode to modify the phase of an incoming wave, with the overall function of a pixellated half wave plate[10]. The liquid crystal layers, which preferentially show a helical

form, has this behaviour suppressed [11] and is constrained into being switched between two stable orientations through the use of an applied voltage allowing for binary phase control of an incident beam. By focussing the binary wavefront and isolating the first diffracted order in the Fourier plane we create an analogue wavefront with a desired phase profile[1]. The primary purpose of this thesis is to evaluate the accuracy of these wavefronts and examine the errors produced alongside the desired wavefront term. We perform both simulated and experimental investigations using a Mach-Zehnder interferometer and consider various methods of optimising results. We finish off the thesis by using the holograms produced in such a manner to interferometrically measure aberrations on two separate refractive transmissive elements and evaluate the potential of the device to be used in a more practical application.

A previous investigation into the capability of a FLCOS SLM device to produce arbitrary wavefronts has shown that the significant source of error into the resultant wavefront is due to the inclusion of higher order terms being passed through the spatial filter, and can limit the accuracy to 0.1 waves RMS in the production of aberrations as small as 0.32 waves (2 radians RMS) spherical aberration[12]. In order to reduce these further we need to consider using up to 24 frames of data for each interferogram as well as alternate binarization techniques. This thesis outlines results indicating that by using a larger hologram and simple three step interferometry we can obtain wavefronts of up to 9 waves RMS Zernike astigmatism with a residual error from aliasing artifacts of less than 0.1 waves RMS, and the use of random binarization techniques will result in a larger residual error rather than a diminished one. For smaller amplitude aberrations such as 1 wave of defocus we show that it is possible to measure variations in optical components down to 0.045 waves RMS, and find that the main limitation in experimental accuracy is from camera noise rather

than holographic artifacts. Simulated results show that in the absence of camera noise the limiting effect of the higher order wavefront contamination is less than 0.02 waves RMS using three step interferometry, but that this value is strongly dependant on the size of the spatial filter used.

There are numerous techniques by which the surface of a part can be measured, far too many to cover in this thesis, however chapter 2 shall give an overview of a small number of them. Some classical tests from before the availability of coherent light sources such as lasers as well as computers for the recording and processing of data are discussed where the detection of aberrant wavefronts is made primarily through detecting the wavefront slopes. Photographs of aberrated wavefronts were measured and then mathematically analysed to determine the original shape that would cause the beam to propagate in such a manner as to match the observed aberration. Three of the most basic, versatile and popular modern techniques are also covered, namely profilometry, Shack-Hartmann wavefront sensing and interferometry, which allow for high resolution, high accuracy maps of surface forms to be produced down to nanometre level precision.

Chapter 3 shall cover interferometry in more detail, in particular some techniques to expand the versatility of it as a measurement tool. The introduction of phase shifting interferometry allows direct and straightforward conversion between a fringe pattern and a phase map of the wavefront, by adjusting the path length of one arm by a known fraction of a wavelength. When interferometrically testing surfaces it is often necessary to use additional elements to pre-modify one of the wavefronts so that the dynamic range of the technique is improved. We describe how holograms can be introduced into the technique of interferometry to reach this goal and modern ways of manufacturing them.

Chapter 4 presents the FLCOS SLM details and how it can be used to

display holographic patterns for use in the interferometer, as well as giving the outline of the Mach-Zehnder interferometer system used to characterise the device. By binarizing a wavefront and then passing it through a lens we can use a spatial filter to isolate the first order diffracted term in the Fourier plane and then re-image to create an analogue wavefront with the desired phase profile. We also show the simulated ability of the SLM to produce up to 9 waves RMS Zernike astigmatism and trefoil measured interferometrically and the growth of the residual aberrations they produce in a simulated null test. We also investigate a method of introducing random background noise into the binarization of the hologram onto the SLM in an attempt to reduce the contaminating effects of passing structured higher order harmonics through the spatial filter.

In chapter 5 we outline the results of experimentally testing the ability of the SLM to produce Zernike modes, as well as attempting to quantify the errors produced alongside them. We see that the dominant source of error in the production of a specific aberration is the production of erroneous Zernike terms in the wavefront, however these are trivial to remove as we can directly subtract them from the residual phase map. A description of the local approach used to unwrap the wrapped phase profiles obtained through the phase shifting algorithm is given, along with a comparison of the residual errors given from a global Fourier based method.

In chapter 6 we see how the SLM can aid interferometric measurements of optics as well as how the inclusion of it into the system affects the residual errors. We describe how we combine numerous interferometrically measured subframes into an overall phase map of a larger optic through a stitching technique with the aid of motorised translation stages that can position the test piece to an accuracy of less than a micron. The first piece used is a refractive transmissive produced

with trefoil aberration inherent in it, and we create individual holograms for each subframe to completely null the aberration present before stitching them to form a composite phase map. We then use a vacuum window piece which could not be measured with the departmental interferometer due to a combination of poor fringe modulation and high fringe density. Rather than creating individual holograms for each subframe we apply one single hologram to every subframe to reduce the fringe density in the interferograms.

Finally chapter 7 provides a conclusion and summary, as well as suggestions for further work we feel would be necessary to accurately characterise the use of an FLCOS SLM in interferometric surface analysis.

Chapter 2

General optical testing techniques

2.1 Introduction

This chapter shall outline some of the basic ideas and methods used in the measurement of surfaces and wavefronts. To begin with, an overview of the Zernike polynomial system shall be given as it provides a basis for wavefront modelling and is used extensively in this thesis. Next some of the theory behind classical methods of wavefront measurement through slope detection are covered, namely the Foucault knife edge test, the Ronchi test and the Platzeck-Gaviola caustic test, before moving on to describing the basics of three of the most versatile and widely used techniques in optical metrology; profilometry, Shack-Hartmann testing and interferometry.

2.2 Zernike polynomials in metrology

In order to ease the understanding of results obtained in optical metrology, wavefronts are often modelled as the sum of a series of polynomial expressions. One of the most commonly used of these are the Zernike polynomials, largely due to their similarity to commonly seen optical aberrations [13, 14]. One example of their importance can be found in the field of astronomical adaptive optics, as they can accurately describe aberrations induced through atmospheric turbulence[15]. The names and associated Zernike term for the first 15 terms are shown in table 2.1 and a visualisation of them can be seen in figure 2.1. A highly useful property of Zernike polynomials is that they form an infinite series of orthogonal expressions over a continuous unit circle which allows for very easy deconstruction of a wavefront into a series of different Zernike components. Any individual Zernike polynomial is expressed in terms of two real functions in polar co-ordinates, $R_n^m(\rho)$ and $\Theta_m(\theta)$, where ρ is the normalised radius, θ the angular co-ordinate and m and n are integers specifying the degree of the polynomial. This two index notation, using m and n to express the polynomial degree, gives an expression for the radial term of the polynomial to be

$$R_n^m(\rho) = \sum_{k=0}^{(n-m)/2} \frac{(-1)^k (n-k)!}{k!((n+m)/2-k)!((n-m)/2-k)!} \rho^{n-2k}, \tag{2.1}$$

with

$$z_n^m(\rho,\theta) = R_n^m(\rho)\cos m\theta, \tag{2.2}$$

and

$$z_n^{-m}(\rho,\theta) = R_n^m(\rho)\sin m\theta. \tag{2.3}$$

Zernike Term	Aberration
Z_0	Piston
Z_1	X-tilt
Z_2	Y-tilt
Z_3	Astigmatism
Z_4	Defocus
Z_5	Astigmatism at 45°
Z_6	Trefoil
Z_7	Coma
Z_8	Coma at 45°
Z_9	Trefoil at 45°
Z_{10}	Quadrafoil
Z_{11}	Secondary Astigmatism
Z_{12}	Spherical Aberration
Z_{13}	Secondary Astigmatism at 45°
Z_{14}	Quadrafoil at 45°

Table 2.1: A list of the first 15 Zernike terms along with their names.

Importantly, these equations apply for when the value of $(n-m)$ is even, for $(n-m)$ odd $z_n^m(\rho,\theta)=0$. These expressions can be normalised

$$Z_n^m(\rho,\theta)=\begin{cases} z_n^m(\rho,\theta)\sqrt{(n+1)} & for\ m=0 \\ z_n^m(\rho,\theta)\sqrt{2(n+1)} & else, \end{cases} \tag{2.4}$$

so that an arbitrary wavefront $W(\rho,\theta)=NZ(\rho,\theta)$ represents N waves RMS of the specified Zernike term. It is also possible to express the polynomial with just a single index term, i, from which the individual m and n terms can be extracted from

$$n=\text{int}(\sqrt{2i+1}+0.5)-1, \tag{2.5}$$

$$m=2i-n(n+2), \tag{2.6}$$

where the 'int' function returns the nearest integer of the argument.

It is possible by analysing the expanded Zernike terms to obtain expres-

Seidel Term	Aberration	Magnitude	Angle
$W_{0,0}$	Piston	$Z_0 + Z_4 + Z_{12}$	-
$W_{1,1}$	Distortion	$\sqrt{(Z_2 - 2Z_8)^2 + (Z_1 - 2Z_7)^2}$	$\arctan(\frac{Z_1 - 2Z_7}{Z_2 - 2Z_8})$
$W_{2,0}$	Field Curvature	$2Z_4 - 6Z_{12}$	-
$W_{2,2}$	Astigmatism	$\sqrt{Z_5^2 + Z_3^2}$	$\frac{1}{2}\arctan(\frac{Z_3}{Z_5})$
$W_{3,1}$	Coma	$3\sqrt{Z_8^2 + Z_7^2}$	$\arctan(\frac{Z_7}{Z_8})$
$W_{4,0}$	Sph Abb	$6Z_{12}$	-

Table 2.2: Magnitude and angle with respect to the x axis of primary Seidel aberrations written in terms of Zernike coefficients up to Z_{12}. Terms of Z_{11} and Z_{13} have been omitted as they comprise a first order correction to astigmatism.

sions for the first and third order wavefront aberrations, i.e. standard Seidel Aberrations [16]. By writing the wavefront $W(\rho, \theta)$ as

$$W(\rho,\theta) = W_{0,0} + W_{1,1}\rho\cos\theta + W_{2,0}\rho^2 + W_{2,2}\rho^2\cos^2\theta + W_{3,1}\rho^3\cos\theta + W_{4,0}\rho^4, \quad (2.7)$$

we can compare the coefficients relating to similar terms in ρ and θ from

$$\begin{aligned} W(\rho,\theta) &= Z_0 + Z_1\rho\sin\theta + Z_2\rho\cos\theta + Z_3\rho^2\sin 2\theta + Z_4(2\rho^2 - 1) \\ &+ Z_5\rho^2\cos 2\theta + Z_6\rho^3\sin 3\theta + Z_7(3\rho^2 - 2)\rho\sin\theta + Z_8(3\rho^2 - 2)\rho\cos\theta \\ &+ Z_9\rho^3\cos 3\theta + Z_{10}\rho^4\sin 4\theta + Z_{11}(4\rho^4 - 3\rho^2)\sin 2\theta \\ &+ Z_{12}(6\rho^4 - 6\rho^2 + 1) + Z_{13}(4\rho^4 - 3\rho^2)\cos 2\theta + Z_{14}\rho^4\cos 4\theta. \end{aligned} \quad (2.8)$$

By using the identity $\alpha\cos\theta + \beta\sin\theta = \sqrt{\alpha^2 + \beta^2}\cos(\theta - \arctan(\beta/\alpha))$ it is possible to express the magnitude and the angle with respect to the x axis of all 6 Seidel aberrations as seen in table 2.2.

Despite their usefulness in modelling wavefront aberrations, they carry the major limitation of being defined over a circular area. For astronomical adaptive

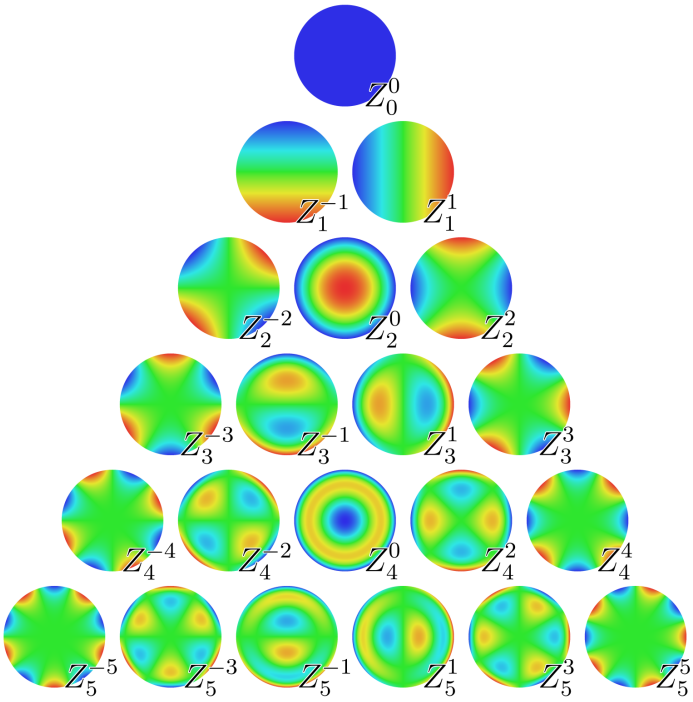


Figure 2.1: Aberration shape and ordering of first 21 Zernike Polynomials which can be trivially reproduced through programming software. Single index notation here begins at the top, and moving down goes from left to right along rows. Each new level in the pyramid designates an increase in the radial order of the polynomial[17].

optics this is not necessarily a major drawback, as often a central circular region is obscured and it is possible to use a modified annular Zernike set[18] however for interferometric applications this is not the case. Zernike analysis is not possible when looking at a region containing the edge of an optic, as it is not possible to define a circular region over which to calculate the orthogonal basis set. This necessitates that a Zernike modal deconstruction can only be applied to a stitched image of a full circular optic, or over a specified circular region of interest on the test surface.

2.3 Classical tests

2.3.1 Foucault's Knife Edge Test

There are many techniques which have been developed for the measurement of wavefronts, a number of which shall be discussed here. Different methods often measure different aspects of the wavefront under test, for example the Foucault test measures the curvature of the beam being investigated through the blocking off of the beam at varying points along the optic axis. Shack Hartmann techniques however measure the slope of the wavefront under test, by detecting the lateral shift in focussed spot patterns from expected positions. It is also possible to measure the height of a surface through profilometric techniques, or use interferometry to measure relative phase differences over the wavefront.

The Knife Edge test described by Foucault in 1858 [19] has proven to be a highly effective yet simple technique for analysing the surface form of optics. The technique allows for the detection of transverse aberrations by using a sharp edge to partially block the beam in order to create shadowed regions in the image of the test optic. It is also a useful tool for detecting the focal point of a lens or spherical reflector as can be seen in figure 2.2. As the knife is inserted into different positions along the optic axis, it blocks varying parts of a wave moving from left to right, which gives rise to characteristic patterns when propagated to the image plane. When the knife is inserted before the focal point the shadow moves across the surface in the opposite direction to the movement of the knife and vice versa for when the knife is inserted after the focal point. Should the knife block the beam at exactly the focal point the shadow will fill the whole image almost instantaneously without appearing to have moved from one side to the other.

It is possible to consider the Foucault test as a form of null test, where a result devoid of features corresponds to an ideal surface, as an irregularly

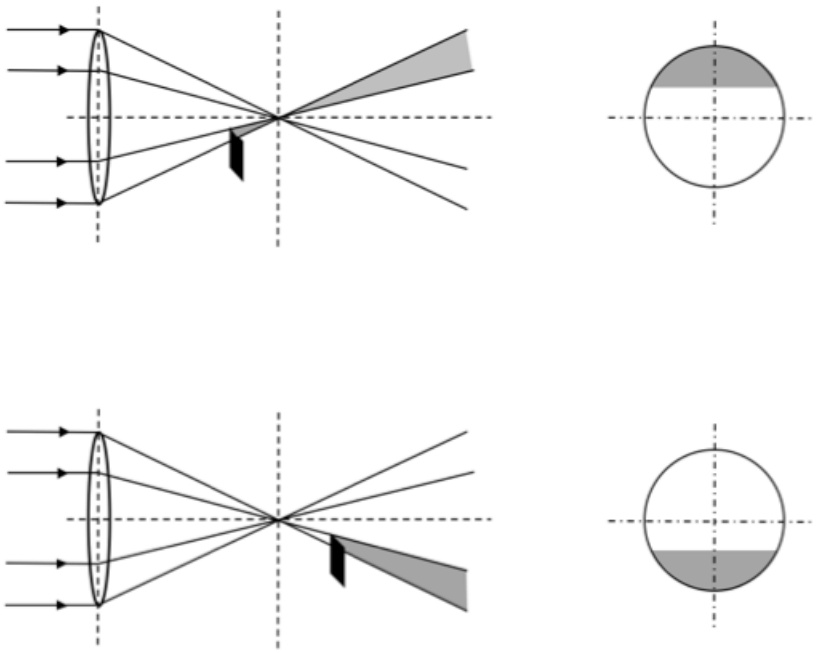


Figure 2.2: Knife edge test on an ideal lens for use in determining focal point. By inserting the knife edge before the focal point of the lens we see a darkened region at the top of the image which moves down as the knife approaches the focal point, and once the knife has passed through it shows the darkened region on the bottom. Deviation from an ideal lens will be observed through deviations in the light-dark boundary from a straight chord, or even through the introduction of secondary darkened regions.

shaped concave optic will have regions possessing different radii of curvature. This means that when the knife is introduced, the shadow pattern will posses corresponding light and dark regions. Should the optic be a perfect spherical reflector then a Foucault test at the focal length will show a uniform shadowed result.

The geometrical theory behind the Foucault test is elegantly simple. Consider a knife placed across the (X, Y) plane as seen in figure 2.3, at an arbitrary position between the mirror and the camera, using the convention that ϕ, the angle with respect to vertical the knife edge blade makes, is positive for when

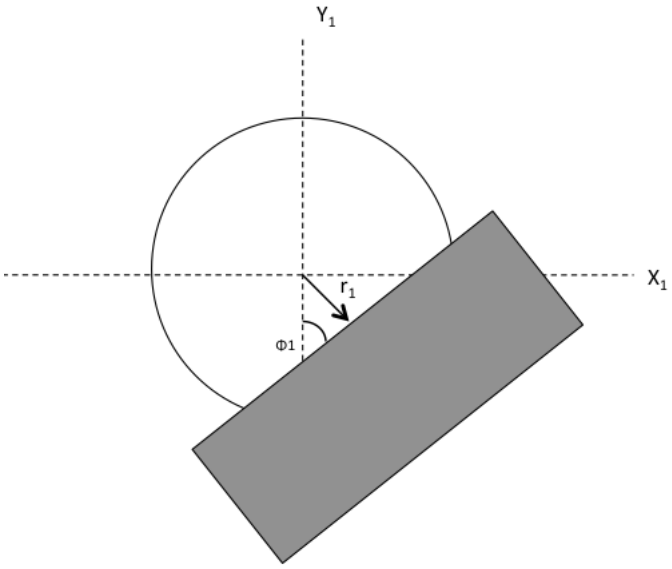


Figure 2.3: Basic geometry of the knife edge test. A sharp edge, frequently a razor, is inserted into the beam with closest approach distance r_1 from the beam centre at an angle ϕ_1 from the vertical. Knowing these two quantities allows for the analysis of the light and dark regions in the image frame to reveal the nature of the aberrations imbued upon the beam by the optic used.

the slope of the knife edge is positive and r is the closest distance from the knife to the origin. The transmission function, $T(X, Y)$, for that plane is given by

$$T(X,Y) = \begin{cases} 1 & if\ X\cos\phi - Y\sin\phi < r \\ 0 & if\ X\cos\phi - Y\sin\phi \geq r. \end{cases} \tag{2.9}$$

The wavefront $W(x_1, y_1)$ can be described using:

$$(x_1, y_1) = \left(-R\frac{\partial W}{\partial x}, -R\frac{\partial W}{\partial y} \right), \tag{2.10}$$

where R is the distance between the (x_1, y_1) plane in which the shadow pattern is being viewed and the parallel (X, Y) plane [20].

By combining equation 2.9 and equation 2.10 we can describe the effect of the transmission function of the knife on the wavefront

$$T\left(\frac{\partial W}{\partial x}, \frac{\partial W}{\partial y}\right) = \begin{cases} 1 & if\ -\frac{\partial W}{\partial x}\cos\phi_1 + \frac{\partial W}{\partial y}\sin\phi_1\ <\ \frac{r_1}{R} \\ 0 & if\ -\frac{\partial W}{\partial x}\cos\phi_1 + \frac{\partial W}{\partial y}\sin\phi_1\ \geq\ \frac{r_1}{R}. \end{cases} \tag{2.11}$$

Equation 2.11 can now be used to analyse the the shadow pattern produced by examining a Foucault Knife Edge test on a reflective concave mirror with focus and spherical aberration present. By inserting the knife at a point before or after the focal point of the mirror as in 2.2 we know that there is a degree of defocus present in the transmitted wavefront from the position of the knife. The aberration function in such a scenario can be given by:

$$W(x, y) = A(x^2 + y^2)^2 + D(x^2 + y^2). \tag{2.12}$$

Also, by noting that spherical aberration and focus both are entirely rotationally invariant we can simplify equation 2.11 by choosing any axis as a reference point to specify the position of the knife, in this case the X axis so that we get $\phi_1 = 90°$ giving:

$$T\left(\frac{\partial W}{\partial y}\right) = \begin{cases} 1 & if \frac{\partial W}{\partial y}\ <\ \frac{r_1}{R} \\ 0 & if \frac{\partial W}{\partial y}\ \geq\ \frac{r_1}{R}. \end{cases} \tag{2.13}$$

Using equation 2.13 on equation 2.12 leaves us with the equation for the border of the shadow pattern:

$$y^3 + \left(\frac{D}{2A} + x^2\right)y - \frac{r_1}{4AR} = 0. \tag{2.14}$$

In order to gain an appreciation of the rough structure of the shadowed regions,

consider only the boundaries of the shadow pattern on the Y axis, letting $x = 0$. The discriminant of this cubic equation is:

$$\Delta = \left(\frac{r_1}{8AR}\right)^2 + \left(\frac{D}{6A}\right)^3. \tag{2.15}$$

So now we have three known cases from the nature of the roots of any cubic equation [1]:

1. If $\Delta < 0$, there are three real and different roots.
2. If $\Delta = 0$, there are three real roots, at least two of which are equal.
3. If $\Delta > 0$, there are two imaginary roots, and only one real root.

Now we can see that the shadow pattern will have more than one dark region if conditions 1 or 2 are met, which is so when the sign of the spherical aberration and that of the focus terms are different. When performing such a test therefore with a knife inserted into the system before the focal point where the focus term will be negative, a shadow pattern containing two darkened regions can be interpreted as being due to a positive spherical aberration term at the point where the knife is.

It is also straightforward however to analyse aberrations without rotational symmetry, for example coma. The aberration function for coma is given by:

$$W(x, y) = By(x^2 + y^2) + D(x^2 + y^2). \tag{2.16}$$

By considering a knife inserted parallel to the Y axis, $\phi^\circ = 0$, we get from equation 2.11:

$$2Bxy + 2Dx = \frac{-r_1}{R}, \tag{2.17}$$

[1] In the general analysis of cubic roots the cases for 1 and 3 are reversed, but here the discriminant has a prefactor of $-\frac{1}{108}$ omitted for simplicity

and

$$\left(y + \frac{D}{B}\right)x = \frac{-r_1}{2RB}. \tag{2.18}$$

Equation 2.18 is clearly that of a hyperbola centred on $[0, -D/B]$.

Next consider the knife being inserted parallel to the X axis, we get from equation 2.11:

$$Bx^2 + 3By^2 + 2yD = \frac{r_1}{R}, \tag{2.19}$$

$$x^2 + 3\left(y + \frac{D}{3B}\right)^2 = \frac{r_1}{RB} + \frac{1}{3}\left(\frac{D}{B}\right)^2. \tag{2.20}$$

Equation 2.20 describes the shape of an ellipse centred on $[0, -(D/3B)]$. By inserting a knife into the paraxial plane of the system at the focal point so $D = 0$, and then placing a scale over the shadow pattern and measuring the properties of the ellipse or hyperbola the amount of aberration B present can easily be measured.

2.3.2 Ronchi Test

Another classical metrology method, one which is possible to elegantly express purely in the language of geometrical optics, is the Ronchi Test [21]. In principle the test is based on the observation that when a grating is introduced close to the centre of curvature of a mirror, an observer looking through the grating from the far side, as in figure 2.4[22], will see a superposition of the grating over the image of the grating on the mirror. These are called combination fringes and it is the interpretation of these that allow for form measurements to be made. The Ronchi test is similar to the Knife-edge test in that what we see is the effect of transverse aberration on the image of the grating lines. As the reflected ray

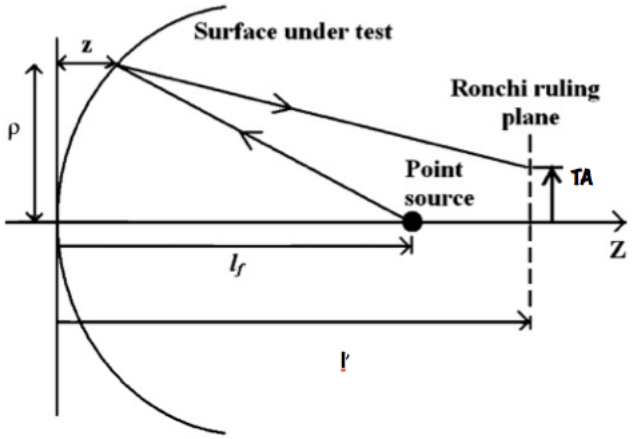


Figure 2.4: Basic set up for a Ronchi test. When viewing in the image plane, the deviation from straightness of the grating lines exposes information on the aberrations present in the test mirror. These lines can then be used to isolate the presence of specific aberrations in a similar way to the results of a knife edge test.

moves towards the right of 2.4 it focusses at the mirrors focal point, and as a grating is moved from inside the radius of curvature towards the focal point we will see a reduction in the number of fringes and an increase in their size. This is because the increasingly small beam diameter is illuminating progressively less of the grating. As we move through the focal point and beyond, the expanding beam will illuminate more of the grating, so we see an increase in fringe density. Deviations in the mirror from spherical results in a bowing in the observed fringe pattern, the shape of this will depend on the type of aberration seen. The accuracy of the Ronchi test is dependent on considerations of the grating density used. As is logical a higher fringe density would allow for increased sensitivity of the test, however the effects of diffraction become more prevalent here and cause a reduction in the contrast of observed fringes as well as blurring the fringe boundaries. A trade off between the two is therefore necessary.

By approximating the wave aberration function given by [20] as:

$$\frac{\partial W}{\partial x} = -\frac{TA_x}{l'}; \frac{\partial W}{\partial y} = -\frac{TA_y}{l'}, \tag{2.21}$$

where l' is the distance from the exit pupil to the Ronchi grating. We may then write for the general case that:

$$\frac{\partial W}{\partial x}\cos\phi - \frac{\partial W}{\partial y}\sin\phi = -\frac{md}{l'}, \tag{2.22}$$

where ϕ is the angle the ruling of the grating makes with respect to the y axis and m is the order of fringe corresponding to a point (x, y) on the grating.

By following the methodology in [23] and noting the equation for primary aberrations :

$$W = A(x^2+y^2)^2 + By(x^2+y^2) + C(x^2+3y^2) + D(x^2+y^2), \tag{2.23}$$

and noting how the defocus term $D = r/2l'^2$ with r being the distance from the grating to the paraxial centre of curvature we can combine with equations 2.21 with 2.22 to obtain:

$$\begin{aligned} &4A(x^2+y^2)(x\cos\phi - y\sin\phi) + B[2xy\cos\phi - (3y^2+x^2)\sin\phi] + \\ &2C(x\cos\phi - 3y\sin\phi) + 2D(x\cos\phi - y\sin\phi) = -\frac{md}{l'}. \end{aligned} \tag{2.24}$$

From this we can now obtain the descriptions for various ronchigrams, aided by the relations:

$$x = \eta\cos\psi + \zeta\sin\psi, \tag{2.25}$$

$$y = -\eta\sin\psi + \zeta\cos\psi. \tag{2.26}$$

Considering first of all the simple case of moving the grating along z we can

trivially obtain expressions for the expected Ronchigrams for the case of pure defocus:

$$2D(x\cos\phi - y\sin\phi) = -\frac{md}{l'}, \tag{2.27}$$

$$2D(\eta\cos\psi^2 + \zeta\cos\psi\sin\psi + \eta\sin\psi^2 - \zeta\cos\psi\sin\psi) = -\frac{md}{l'}, \tag{2.28}$$

$$2D\eta = -\frac{md}{l'}. \tag{2.29}$$

These clearly will result in Ronchigrams consistent of straight bars parallel to the grating ruling with a separation of:

$$\Delta\eta = \frac{d}{2Dl'}. \tag{2.30}$$

From this we can also see that in the special case where the grating is placed at the centre of curvature there will be an infinite distance between the observed fringes; a uniform intensity distribution.

Now considering the more complicated case of a test optic displaying astigmatism we start with a slightly more complex aberration term:

$$2C(x\cos\phi - 3y\sin\phi) + 2D(x\cos\phi - y\sin\phi) = -\frac{md}{l'}, \tag{2.31}$$

$$\begin{aligned}&2C(\eta\cos\psi^2 + \zeta\cos\psi\sin\psi + 3\eta\sin\psi^2 - 3\zeta\cos\psi\sin\psi) + \\ &2D(\eta\cos\psi^2 + \zeta\cos\psi\sin\psi + \eta\sin\psi^2 - \zeta\cos\psi\sin\psi) = -\frac{md}{l'},\end{aligned} \tag{2.32}$$

$$2C[\eta + 2\eta\sin\psi^2 - 2\zeta\cos\psi\sin\psi] + 2D\eta = -\frac{md}{l'}, \tag{2.33}$$

$$2C[\eta(2 - \cos 2\psi) - \zeta\sin 2\psi] + 2D\eta = -\frac{md}{l'}. \tag{2.34}$$

From 2.34 we can see that once again the observed fringes will be parallel lines with constant separation. Unlike with the case of pure defocus however the alignment of these bars is dependant on the rotation of the Ronchi grating ψ according to:

$$\tan\Theta = \frac{C\sin 2\psi}{D + C(2 - \cos 2\psi}). \tag{2.35}$$

We can see that the inclination of the Ronchi grating is not the only cause of fringe rotation; moving the grating from the tangential focus through to the sagittal focus will also cause a rotation of the combination pattern.

2.3.3 Platzeck - Gaviola Caustic Test

The Platzeck-Gaviola test [24] is based on the observation that the centre of curvature for an off axis segment of a purely parabolic surface does not lie on the optic axis but rather on the caustic curve, a phenomena where light reflected by a curved surface gives a pattern with a notable curved line, called the caustic, tangential to the direction of the reflected light. By placing a mask with two elliptical windows over a parabolic mirror they observed that the image was sharpest not where the two beams cross the optic axis but rather beyond it. This is shown in figure 2.5 where we see that rays being focussed from the outer zones of the mirror do not focus to a point on the optic axis, as we consider zones moving towards the centre of the mirror the deviation of the focal point from the optic axis diminishes.

Once again, considering the geometrical analysis presented in [23] we define the co-ordinates for the centre of curvature of the caustic (η, ξ) then from figure 2.5we can see that:

$$\frac{\xi}{2(\eta + Kz)} = \frac{S}{(1/c) - (K+1)z}. \tag{2.36}$$

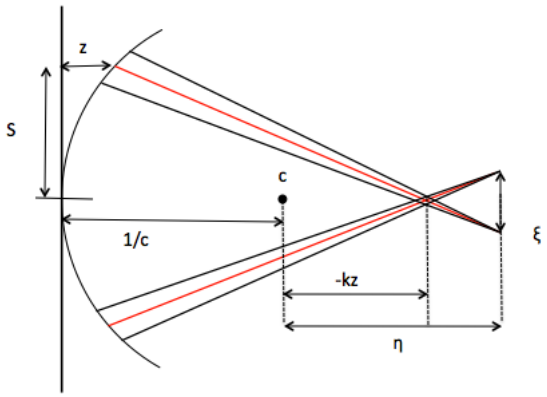


Figure 2.5: Geometry of the caustic where c is the inverse of the radius of curvature of the central zone. Significantly where the two rays cross the optic axis is not the focal point of the segment of the mirror from whence they originate.

We can also consider the distance R between a point on the surface of the mirror and the centre of curvature from figure 2.5 as:

$$R^2 = \left(S + \frac{\xi}{2}\right)^2 + \left(\frac{1}{c} + \eta - z\right)^2. \tag{2.37}$$

The first step towards obtaining expressions for ξ and η is to note that R is the radius of curvature at that point on the mirror, also obtainable from the expression:

$$R = \frac{[1 + (z'(S))^2]^{3/2}}{z''(S)}. \tag{2.38}$$

One then needs to obtain an expression for the behaviour of the surface under test. Using $S^2 = x^2 + y^2$ with z as the optic axis we can describe it as:

$$z = \frac{cS^2}{1 + [1 - (K+1)c^2S^2]^{1/2}} + A_1S^4 + A_2S^6 + A_3S^8 + A_4S^20. \tag{2.39}$$

The case where all the A_x coefficients are zero corresponds to a conic surface of

revolution. This then provides:

$$\frac{dz}{dS} = cS[1-(K+1)c^2S^2]^{-1/2}. \tag{2.40}$$

$$\frac{d^2z}{dS^2} = c[1-(K+1)c^2S^2]^{-3/2}. \tag{2.41}$$

Next we acquire a pair of simultaneous equations by first substituting 2.40 and 2.41 into 2.38 and secondly substituting $\xi/2$ from 2.36 into 2.37 to get:

$$R = \frac{1}{c}(1-Kc^2S^2)^{3/2}, \tag{2.42}$$

$$R^2 = \left(\frac{1}{c}+\eta-z\right)^2\left\{1+\frac{S^2}{[(1/c)-(K+1)z]^2}\right\}, \tag{2.43}$$

which combine to give us an expression for η:

$$\eta = z - \frac{1}{c} + \frac{(1/c)(1-Kc^2S^2)^{3/2}}{\left\{1+\dfrac{S^2}{[(1/c)-(K+1)z]^2}\right\}^{1/2}}, \tag{2.44}$$

Substituting 2.39 into 2.44 to finally reach:

$$\eta = -Kz\{3+cz(K+1)[cz(K+1)-3]\}. \tag{2.45}$$

This can be then substituted back into 2.36 to acquire ξ and thus a provide a description of the co-ordinates for the centre of curvature of the conic of revolution.

$$\xi = 2ScKz\left\{\frac{2+cz(K+1)[cz(K+1)-3]}{1-cz(K+1)}\right\}. \tag{2.46}$$

From these equations it is possible to calculate the theoretical caustic co-ordinates

for a known mask geometry. Comparison of experimental and theoretical values allow for determination of conic constant K of the mirror under test.

2.4 Profilometry

2.4.1 Contact profilometry

Contact profilometers are a basic tool for providing surface form measurements, and fall into one of two different varieties; Stylus profilometers and scanning probe microscopes. Unlike the classical methods outlined above, this method allows for the direct measurement of the height of the surface under test rather than the slope or curvature properties of the reflected wavefront.

In stylus profilometry a small tipped probe is placed in contact with the optic under test, and then one of the two is moved whilst vertical displacements of the stylus are recorded to produce a height map. One method of reading the vertical motion of the stylus is by using a linear variable differential transformer (LVDT) which outputs information on exact height measurements. The measurements are taken with respect to a reference surface known as the datum, which can be achieved through placing the test part on three gage blocks prior to measurement. A measurement of the position of the gage blocks using the profilometer provides three data points from which a plane can be defined which forms the basis for an absolute measurement of the test surface profile. The accuracy of the vertical height measurements can be verified by using a length measuring Michelson interferometer to incorporate scaleability into the technique, allowing for experimental results to be confidently expressed as absolute distance measurements.

It is of major importance to consider the characteristics of the stylus tip when performing contact profilometry. As the tip is in contact with the surface

at all times it is important to obtain as much information about the stylus as possible. Clearly one vitally important property of the stylus is the dimensions of the tip, as a smaller tip, that is one where the point of the stylus is described by a small radius of curvature in the region of 0.1μm, will be able to resolve much smaller features of the optical surface. This comes with a drawback however a smaller tip imposes a larger surface pressure onto the optic than a larger one, thus increasing the chance of scratching. Not only does a tip with a larger radius of curvature lessen the risk of damage to the surface it increases the speed at which a measurement can be taken. Plastic deformations damaging the surface are not the only risk associated with fine tipped styluses either as elastic deformations in the surface, whilst non damaging to the test optic, provide inaccurate measurements. The shape of the stylus also warrants attention as the measurements taken via this technique are convolutions of the stylus shape and size on the test surface. A tip which tapers at a large angle will clearly be unable to measure a narrow deep trench in the surface as shown in figure 2.6.

Further considerations that need to be taken into account are the hardness of the stylus material and the contact force applied to the surface under test. A harder stylus material will resist wear more than a soft material which is crucial as any wearing of the tip would lead to errors in surface height measurements, but further increases the possibility of scratching. In general the stylus tip is made from diamond in order to reduce the effect of wear on measurements. Finally the speed at which the stylus is traced over the optic (or vice versa) must be considered; too fast and the stylus may skip over surface features and lose contact with the optic however the speed must not be so slow as to lead to low measurement efficiency.

The very nature of stylus profilometry as a contact metrology method gives it one unique advantage over other non-contact methods. Since techniques such as

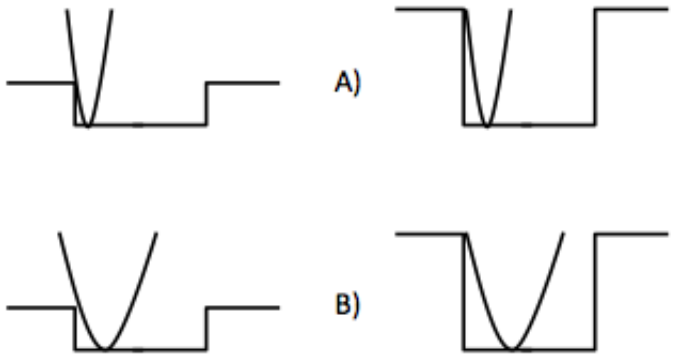


Figure 2.6: The size of the stylus needs to be suitable to the size of the feature it will be measuring. It can be seen in A) that the stylus is capable of measuring both shallow trenches with high accuracy and deep ones too, whereas in B) the larger stylus does not detect the bottom of the shallow trench as soon as the smaller one, an effect which worsens for deeper trenches where the measured width of the base will be significantly smaller than that made with a smaller stylus.[25]

interferometry and Shack-Hartmann wavefront sensing image the optic surface directly, they are prone to errors from contaminants on the test surface. If a significant number of these contaminants accumulate or are large enough then they may lead to significant deformations of the imaged wavefront resulting in incorrect measurements. A stylus in contact with the surface however, will remain unaffected by the the presence of dust particles etc, providing they do not strongly adhere to the surface and therefore display an increased suitability for use in dirty environments. They are also well suited for the measurement of surface roughness as the stylus can be moved with respect to a reference pattern describing the form of the test optic.

The point measurement nature of stylus profilometry measurements mean that they can take a long time to build up a reading of the whole test shape.

Naturally this dramatically increases with the size of the test piece making it unfeasible for large optic testing. It is possible however to measure surfaces with large form variations through moving the stylus over a reference pattern using a swing arm system, so that height deviations from a pre-programmed surface are recorded [26, 27, 28].

Another contact profilometry method is the scanning probe microscope,[29] which allows for the detection of features on the angstrom scale. Similarly to stylus profilometry this method consists of a tip which is moved over a surface to provide a point by point surface map. However where stylus profilometry measures vertical height displacements of the tip, the scanning probe microscope measures tunnelling currents or the atomic force emanating from the test piece at a distance of angstroms from the surface depending on the type of probe used.

With a scanning tunnelling microscope a metal tip is moved in proximity to an electrically conductive surface and upon application of a potential difference between the two a tunnelling current is measured. The current increases exponentially with proximity to the surface allowing for highly resolved atomic scale measurements. This technique is capable of running in constant height or constant current modes. In the constant current mode a feedback loop is used to maintain the current value between tip and surface at all points, and constant height mode measures the tunnelling current at a fixed distance from the sample. Constant height mode allows for faster measurements to be taken as it does not involve the need for a feedback loop to maintain the current, however due to the exponential growth of the current with proximity to surface of the tip it has a smaller dynamic range than the constant current mode. Scanning tunnelling microscopes however have a clear limitation in their need for a conducting surface in order to obtain the tunnelling current that therefore

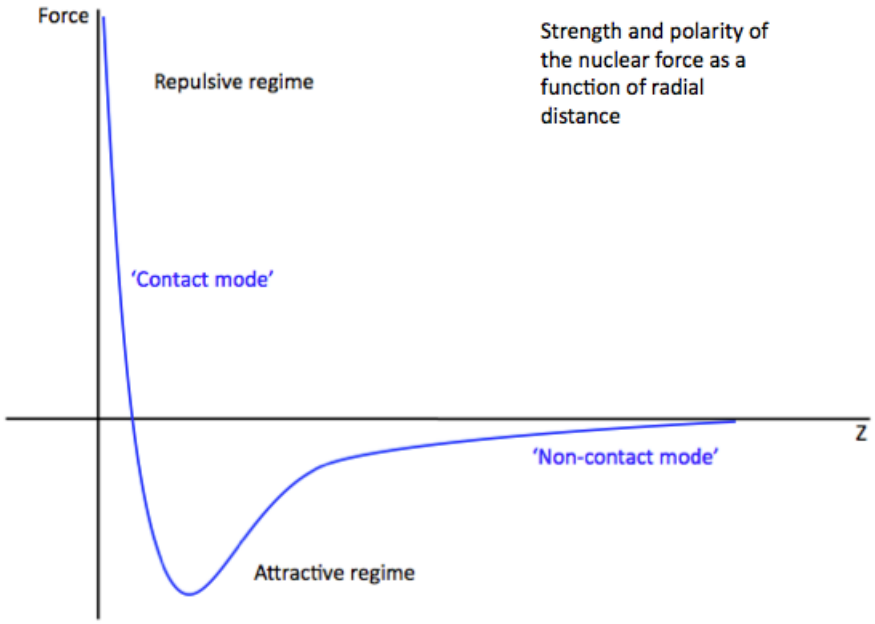


Figure 2.7: The behaviour of the nuclear force at small scales. At distances of roughly 0.7 fm it acts repulsively to maintain separation of nucleons as a result of the Pauli exclusion principle, however it has a peak maximal attractive value at 0.9 fm. It continues to provide an attractive force between two nucleons up to a distance of approximately 2 fm at which point it is overcome by the repulsive strength of the electromagnetic force. These two separate regimes allow for two different methods of microscope operation where measurements are taken on the influence of either the repulsive or attractive tendency of the force depending on the length scales used.

precludes the technique from use in measuring non conductive substances such as glass.

The Atomic Force microscope[30] is a variation on the scanning tunnelling microscope that allows for the metrology of non conductive surfaces by measuring both the repulsive and attractive behaviour of the nuclear force, as seen in figure 2.7. When being used to measure the repulsive force, known as the contact method, the probe head is mounted on the free end of a highly sensitive cantilever and kept in contact with the surface in an equivalent manner to a

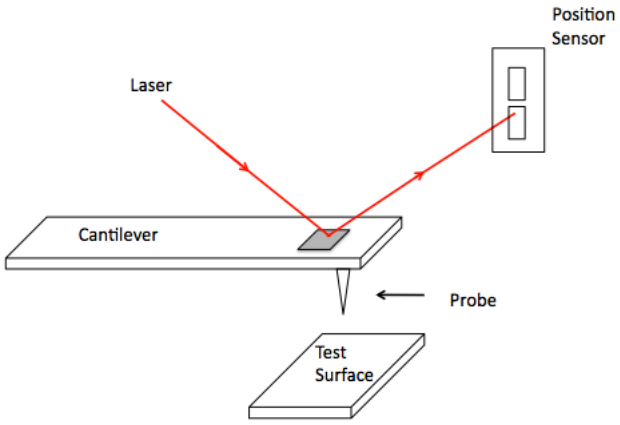


Figure 2.8: General schematic for an atomic force microscope using a laser deflection based method of reading cantilever flexure. A position sensor is used to measure the deviation of the cantilever motion as a result of interacting with the nuclear force.

record player. The repulsive force felt by the tip bends the cantilever and the degree of flexure can be measured by reflecting a laser beam off the back of the cantilever and measuring the deflection of the spot from it's equilibrium position as in figure 2.8[31]. Using this technique it is possible to measure vertical displacements of the cantilever down to the angstrom scale. As scanning at a constant height risks the probe crashing into the potentially delicate surface under test a feedback mechanism is generally incorporated to keep constant force between surface and probe.

In the non-contact operating mode the probe is oscillated in proximity to the surface at a frequency very close to that of the cantilever resonance frequency, a technique called the Tapping Method [32]. Any deviations in the recorded oscillatory motion are ascribed to gains or losses in energy through interaction with atomic forces from the surface [33]. A fundamental consideration in the application of this technique is the amplitude of the cantilever oscillation, too

small and the tip may stick to a sample with liquid coating due to meniscus forces however too large and the tip can impact upon the surface potentially damaging both the test piece as well as the metrology tool. This technique is capable of not only measuring the surface profile of a test piece but also the magnetic and electric field profiles as the collected data is indicative of variations in the energy of the oscillations, furthermore characteristics such as the viscosity of samples and the adhesive properties of a surface can also be evaluated.

2.4.2 Non-contact profilometry

Rather than using contact methods to profile a surface, it is possible to use non-contact systems to characterise the topography optically. These still provide a point by point analysis of the surface under test however do so by interferometric analysis or the sensing of the best focus point via vertical translations of a lens.

The latter, known as optical focus sensors [34] consists of the test beam being focussed to a point on the surface under test using an objective capable of vertical translation. The returning beam can be split using a prism with each half then being directed to a split detector [35] as shown in figure 2.9. When the objective is either too low or too high then the beams will be registered as larger signals on either the inner or outer detector, and this feedback allows the objective to move to the best focus position and thus provide height information on the surface.

Confocal microscopy[36], as seen in figure 2.10 is another technique that can be employed in the determination of surface profiles, and has similarities to optical focus sensors. Where confocal microscopy offers a significant advantage however is that the signature use of an aperture/spatial filter in the method drastically reduces the amount of stray light that fall upon the detector as the aperture only allows through light focussed at the plan conjugate to it. Phase

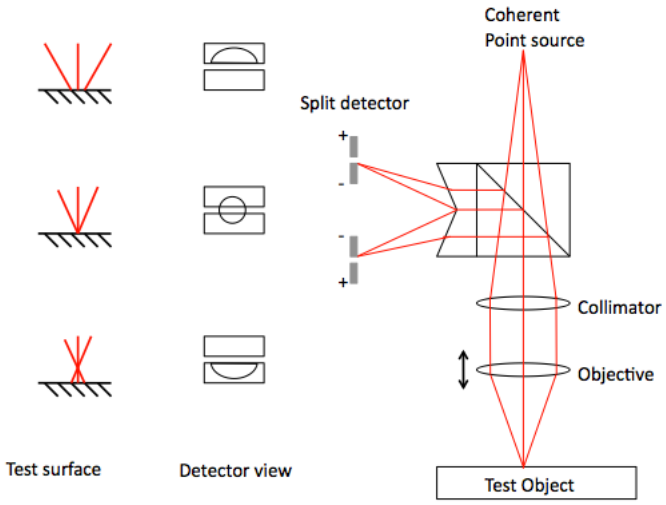


Figure 2.9: Schematic of a basic profilometry focus sensor. The response measured on the split detector indicates the position of the beam focus relative to the test surface, and vertical translation of the objective lens allows for determination of surface profile. When the lens is positioned too high the beam is focussed in front of the detector giving a larger signal on the inside half, and when it is focussed behind the detectors the outside detector registers the larger signal.

maps are built up of point-by-point scans across a test optic, with a vertical translation of the surface applied at each point. The detected irradiance over a series of vertically translated images, or axial PSF, peaks when the surface under test is at the focal point of the system, and drops off as distance from that point increases. This can further be improved by fitting a function to the axial PSF rather than simply determining the best focus position from the point of maximum irradiance. Clearly a narrower axial response function allows for a more accurate measurement of the best focus point, and for smooth surfaces this technique can allow for nanometre sensitivity. Confocal microscopy techniques also allow for precise thickness measurements of transmissive test pieces, as it is possible to obtain an axial PSF for the focus of back reflections originating from

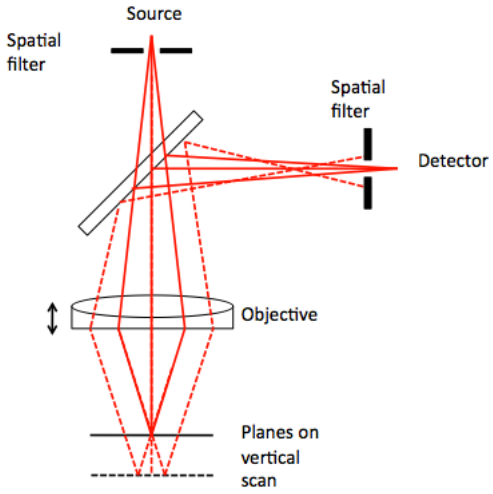


Figure 2.10: Schematic for a basic scanning confocal microscope. The test surface is raster scanned laterally and vertical translation of the objective lens allows for build up of the axial response function.

the interfaces of the test optic. The ability of confocal microscopy to allow for thickness measurements has led to its popularity in biological imaging, as it can be combined with fluorescence techniques to take images of cells in transparent media.

A predictable downside of this technique for metrological applications due to its nature as a profilometer based system is that the data needs to be raster scanned to obtain a full image. However due to its popularity as a measurement tool there has been no shortage of research into improving the efficiency of its data acquisition. Some proposed improvements include the addition of a spinning microlens array [37] to increase irradience measurements and thus reduce integration time. Another suggestion is to replace the confocal pinhole immediately before the camera with a slit and increase the number of detectors along it [38]. This allows for multiple measurements of data points at the same time and can completely remove the necessity of scanning in the direction parallel to

the orientation of the slit.

Yet another optical profiling technique is to use the properties of white light interference. When illuminating a sample with broadband light one can still obtain interference fringes when the path length is small as can be seen from figure 2.11. As each wavelength will interfere with the corresponding reference wave to create many overlapping fringe patterns which when measured with a monochromatic detector will combine to produce a white light interference pattern. Fringe patterns for different wavelengths of light naturally have different periods, however at the point of zero optical path difference they all share a common maxima. As the path difference increases they become less synchronised giving lower irradiances on the detector until they are so incoherent that the fringe modulation disappears entirely. This is what is meant when the fringe pattern is said to be localised. Similar to the two previous techniques, this means that a maximum irradiance value will be registered when the sample is at the best focus position of the objective upon performing a vertical scan. Digital filtering of the observed axial irradiance pattern then allows for the recovery of the envelope and from that the surface shape can be obtained.

2.5 Shack-Hartmann wavefront sensing

The Hartmann test [39] is yet another tool used in metrology that allows for the measurement of aberration based off the measurement of a detected wavefront slope. The test involves placing a mask with an array of holes at the entrance or exit pupil of a test system. In the most simple version of the test the mask consists of a square array of holes placed at regular intervals which allows for a regular sampling of the surface. Such a mask placed over a perfectly flat surface would result in the observation of a perfectly uniform array of spots, however any deviations in surface shape such as defocus would cause a shift in the spot

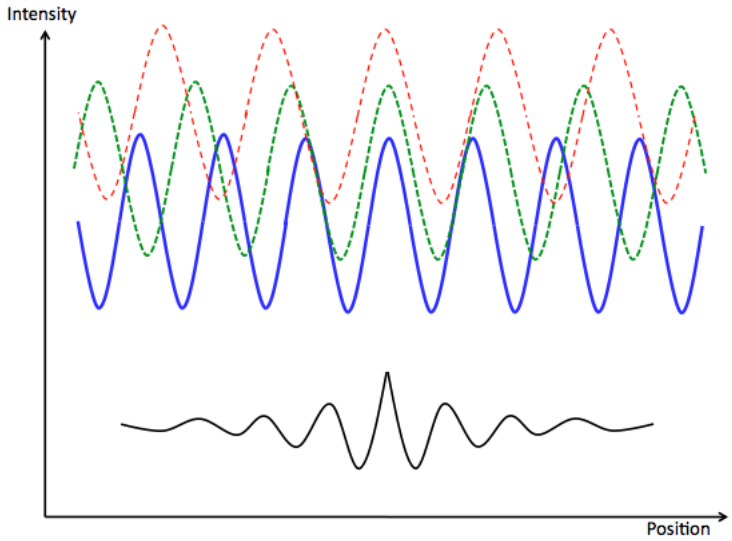


Figure 2.11: As a basic detector records the intensity of all incident wavelengths, a fringe pattern develops as the contributions to the bolometric intensity from the fringe pattern of many different wavelengths is summed up similarly to standard interference. When the two arms of a white light interferometer are perfectly matched, all wavelengths register a maxima and thus we get a bright fringe. As the path length increases we see that wavelength maxima become desynchronised and the modulation of the bolometric intensity drops off.

pattern. This observed transverse aberration can then be used to understand the shape of the wavefront via:

$$\frac{\partial W(x,y)}{\partial x} = -\frac{TA_x(x,y)}{r}, \frac{\partial W(x,y)}{\partial y} = -\frac{TA_y(x,y)}{r}, \tag{2.47}$$

where $W(x,y)$ describes the wavefront, r the focal length of the lens and $TA(x,y)$ the transverse aberration present. A more commonly used variant of this technique, the Hartmann-Shack test [40][41] replaces the regular array of holes and instead uses a lenslet array to focus the light into a series of smaller spots as seen in figure 2.12. Here we see that as a ray bundle travels left to right, it becomes focussed to a spot at the focal plane of the relevant lenslet. The overall wavefront slope of this bundle causes a shift in the position of the focussed spot,

the magnitude of which is proportional to the size of the wavefront slope. The light in this case is focussed into a much tighter area than in the Hartmann test, allowing for more accurate measurement and furthermore is more suitable for lower beam intensities as the created spot is brighter due to the use of lensing optics. The datum with respect these measurements are made is taken from the expected spot pattern from a plane wave encountering the lenslet array.

For use in testing of focussing elements, it is important to note that the image of the spot pattern needs to be taken outside of the focal plane of the tested objective in order to allow for the spot pattern to be measured. The important quantity to be measured from the spot patterns is the deviation of the observed spot from its expected position in the case of a ideal wavefront. As this deviation is a measure of the transverse aberration it corresponds from equation 2.47 that integrating this slope can allow for the build up of the wavefront:

$$W(x,y) = \frac{1}{r}\int_0^x TA_x(x,y)dx, \tag{2.48}$$

$$W(x,y) = \frac{1}{r}\int_0^y TA_y(x,y)dy. \tag{2.49}$$

Which when implementing the trapezoidal integration technique gives a result of:

$$W_{n,m} = W(n-1,m-1) + \frac{\Delta d}{2r}[TA_x(n-1,m-1) + TA_x(n,m-1) + TA_y(n,m-1) + TA_y(n,m)]. \tag{2.50}$$

The dynamical range of the Shack-Hartmann wavefront sensor is predominantly determined by the size of the lenslets used. This is clear to see, since the focal point deviation of the recorded spot is dependant on the average wavefront slope over the focussing element, no features smaller than the lenslet can be re-

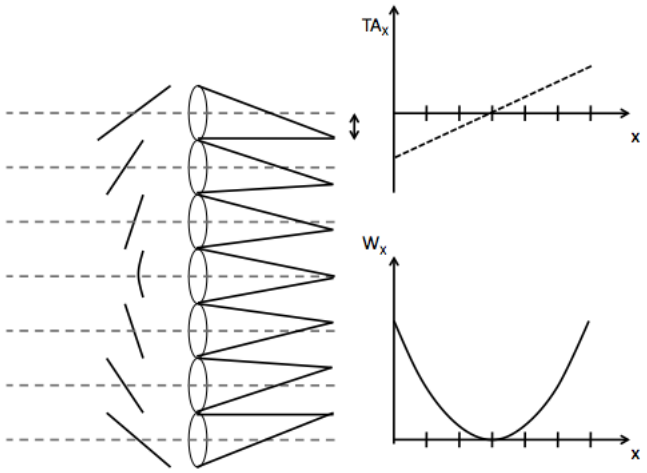


Figure 2.12: The spot pattern registered by the Shack-Hartmann test is representative of the average wavefront slope over the area of each lenslet. The deviation from the expected spot position is then measured and integrated over to recover the phase profile of the incident wave. In this example a wavefront representative of sag is characterised by a linear profile of displaced spot patterns which integrates into a recovered wavefront with a quadratic dependancy on x, as expected.

solved as per $f\theta = d$. Furthermore in order for this technique to be reliable there has to be no possibility of overlapping spots, therefore the maximum wavefront slope that can be measured is equal to half the distance between lenslet centres minus the radius of the spot size:

$$\theta_{max} = \frac{d/2 - \frac{1.22\lambda f}{d}}{f}. \tag{2.51}$$

This precludes the use of Shack-Hartmann screen tests with anything other than the densest microlens arrays from measuring wavefronts with a high degree of spatial variation, such as roughness, due to the inability to reach the Nyquist sampling regime.

One critical obstacle to overcome when using this technique is that measuring the location of the detected spot is only trivial for small wavefront deviations.

Highly aberrated phase profiles will cause the focussed spot to cease appearing as a point spot but rather stretched out corresponding to the shape of the aberration PSF. Much effort therefore is spent determining the optimal centroiding technique for measuring the spot position accurately so that it can be applied to real world metrology. As well as being heavily dependant on a robust reconstruction algorithm, the technique is less sensitive than many other other optical metrology techniques and shows a significantly reduced spatial resolution over stylus profilometry. Despite this however, the wavefront reconstruction procedure is much faster than that capable through profilometry allowing for fast control over deformable mirrors in wavefront shaping. This coupled with its lack of dependance on a coherent light source as well as being a relatively low cost technique has led to it seeing extensive use in the field of adaptive optics.

2.6 Interferometry

A highly useful property of monochromatic coherent light is its ability to interfere constructively or destructively. This allows for one of the most useful and widely used optical testing techniques: Interferometry. Interferometry allows us to compare the wavefront from a known reference surface with one which has interacted with an unknown test piece. The combination of the two gives us a fringe pattern which describes, at each point, the difference between the known and unknown wavefronts. This allows us to measure the phase difference, and hence the height differences between the two surfaces with respect to one another.

There are many different designs of interferometer, some of which will be covered below, and whilst many of the techniques already discussed merely obtain aberration measurements through slope detection or the build up of raster scanned images, interferometry allows for wide field and accurate imaging

of an optic under test. Below is a brief overview of some of the most common and basic types of interferometer, and which is followed in the next chapter by a look at supplemental techniques to improve their versatility.

2.6.1 Types of interferometer

One of the simplest and most easily imagined examples of an interferometric system is that of the Newton Interferometer, commonly used in undergraduate physics experiments to observe Newtons rings as an introduction to interference theory. The principle behind this system is that a transmissive is placed in front of a mirror, and then the interference between the primary reflection and the internal reflection from the transmissive give a characteristic interference pattern. As the interference fringes correspond to the size of the air gap between mirror and optic back surface this technique has been used to measure the shape of transmissive optics in addition to reflective ones. Consider a transmissive optic with sag $x^2/2R$, where R is the radius of curvature and x measured from the centre of symmetry of the optic, almost in contact with a plane mirror. The path difference between the ray corresponding to the back reflection and the ray from the mirror will be x^2/R and taking into account the π phase change in one ray due to reflection the first dark fringe will be found when:

$$\frac{x^2}{R} = n\lambda. \tag{2.52}$$

One distinguishing feature about the Newton interferometer is that the transmissive and reflective elements are generally touching, or close to, in order to provide the smallest possible air gap for measurement. This means that the required illumination need only be monochromatic, the small path difference alleviates the necessity for high coherence lengths. Because of this, this test was very widely used before the laser became commonly available as illumination

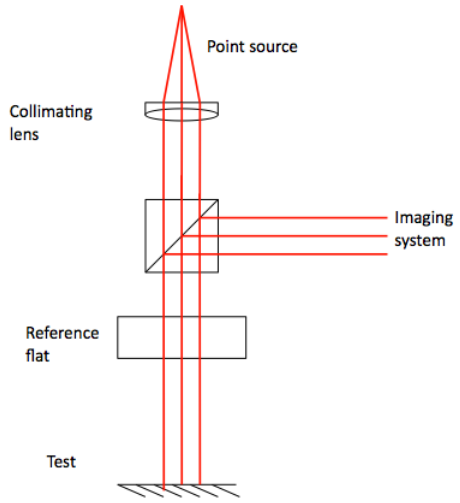


Figure 2.13: Geometry for simple testing of a nominally flat test piece using a Fizeau interferometer illuminated by a monochromatic point source collimated using a concave mirror.

sources such as sodium vapour lamps were sufficient to complete the test. For the measurement of highly curved objects where there would be a large air gap, a reference shape other than a flat needs to be incorporated. The reference shape would need to also be highly accurate however and thus makes the technique less than ideal for the measurement of aspheric optics or those in which an accurate reference flat would be troublesome to produce.

It is sometimes desirable however to measure optical elements so that the air gap between test and reference surfaces is significantly larger than that required in Newton interferometry. The Fizeau interferometer, figure 2.13, is a variant of the Newton interferometer where the two surfaces are not considered to be touching, and instead are separated by more than the coherence length of the sources commonly used for the latter. The illumination source for such a design is therefore required to have an appreciable coherence length, making lasers the ideal source. The prospect of having a longer coherence length source, and thus

reducing the need for a compact system, allows for more flexibility in the design of the interferometer. It is possible to illuminate the system using a point source now, as a collimating lens can be introduced into the system or even a concave collimating mirror.

One limitation of the Fizeau interferometer comes from the fact that it is a common path interferometer. A high reflectivity test piece or reference flat will cause a vast difference in the intensity between the two interfering waves and thus causes a poor contrast in the observed fringes. Furthermore, if the reflective test piece has a poor reflectivity or efficiency then the returned beam will have a severely attenuated intensity, a problem only compounded in the testing of transmissive pieces as they must necessarily be placed in a double pass configuration. As the reference beam is entirely common to the test beam there is very little scope for adjusting the intensity of it and therefore limitations are placed on the flexibility of the system for adjusting to compensate for poor quality fringes.

Conversely however the common path nature of the Fizeau interferometer can also be a blessing as there is less scope for vibrational errors as seen in more complicated interferometric systems. The high sensitivity of an interferometer means that any high frequency variations in path length between the two arms can cause error measurements in the recovered observations. In unequal path interferometers there is vast scope for the origin and propagation of these errors. Perhaps one of the most significant of these comes from the presence of air turbulence in the experiment. Air flow currents and heat variations in the lab will lead to a high level of turbulence which can vastly complicate measurement taking, and in keeping a predominantly common path design Fizeau interferometers limit the path length travelled by the test beam and thus minimise the effect of turbulent air currents in contaminating results.

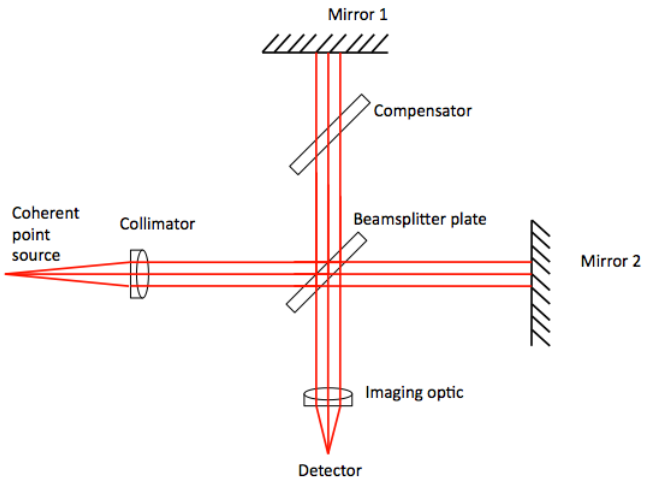


Figure 2.14: Basic layout for a simple Twyman-Green interferometer. The schematic shown here is known as a compensated interferometer as the path difference induced by use of a beamsplitter plate is accounted for by the presence of a compensating optic on one of the arms.

In both the Newton and Fizeau interferometers there is a degree of common path shared by both the reference beam and the test beam. As mentioned this somewhat limits the flexibility of beam modification and interferometer design possible using such methods. A variant of the Michelson interferometer using a point source and a collimator designed by Twyman and Green [42] known as the Twyman-Green interferometer provides incredible flexibility for the testing of optical components by splitting the beam into two separate paths as seen in figure 2.14.

As mentioned in figure 2.14, this example is known as a compensated interferometer. The beamsplitter plate used is only coated on one side and therefore creates a path difference between the two arms of the interferometer which necessitates the addition of a compensator optic in one arm. Commonly however, many Twyman-Green interferometers use beamsplitting cubes rather than plates and are therefore inherently compensated as both beams travel through

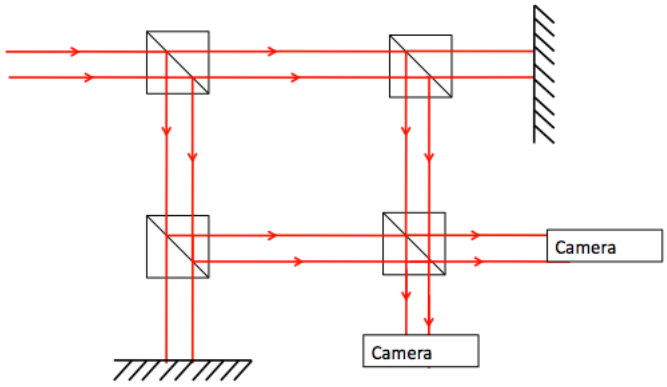


Figure 2.15: A Mach-Zehnder interferometer. This configuration allows for both the double pass applications of the standard Twyman-Green as well as the inclusion of single pass optical systems. Tighter control of beam intensity and polarization state is also permitted as a result of being able to use single pass optics.

equal thicknesses of glass. A more complicated variation of the Twyman-Green is the Mach-Zehnder interferometer, figure 2.15, which provides for a greater splitting of the path suitable for the inclusion of optics or optical configurations that require single-transmission use. A limitation of the Mach-Zehnder interferometer however is that the vibrational issues seen in Twyman-Green configurations are further worsened due to the increased separate path length and larger number of optical elements present. One way to counter this is to use equipment which can be connected with a rigid cage mount system, vibrations are then no longer isolated to one optical element but damped out faster or partially transferred to the other components to lessen the relative translation.

The imaging of the surface as a whole, rather than acquiring information on a point-by-point basis, provides a significant speed advantage over profilometry techniques although at a cost of not being able to detect angstrom scale structure, such as that revealed by atomic force microscopy. Unlike the profilometer or Shack-Hartmann wavefront sensor however, interferometry is almost entirely

reliant upon having a coherent monochromatic source and, particularly in cases where a large path difference is present, the coherence length of the illumination can prevent high quality fringes from being formed. It is also particularly unsuitable in environments where there may be large vibrations, temperature fluctuations or air turbulence due to it's high sensitivity, a problem which is exacerbated further in the testing of large optics through the need for beam projection over a long path length.

One of the primary limiting factors in the dynamic range of the interferometer is the size of the interferograms in the image plane and the camera used to record them. Measuring a wavefront with marked deviation from the reference surface will result in a high fringe density and information can be lost through insufficient sampling in the recorded image. The various configurations of interferometer allow for a range of techniques by which this problem can be overcome however, through the introduction of optics in the reference arm to correct for certain wavefront shapes, such as sphericity, which allows an interferometric record of the deviations from a sphere. The following chapter shall explore in detail how the use of holography with interferometry can expand the dynamic range of the technique.

2.7 Errors in measurements

When measuring anything, be it surface or quantity it is important to consider the errors in the measurement technique used. In most metrology methods there are two fundamental and separate types of error; random errors and systematic ones. The uncertainty in the overall measurement is the combination of all random and systematic errors that may be encountered during the taking and processing of results.

Random errors are those which are unpredictable and will vary throughout

repeated measurements of the quantity under test. Examples of these are random shot noise in CCD cameras where incident photon flux from a uniform flat field varies between pixels and follows a poissonian distribution. Another example which is of interest in interferometry is the effect of vibration and air turbulence in the environment of testing. Air turbulence will cause random variations in the local refractive index both over time, and between points on separate arms of the interferometer. These can be highly disruptive to the measurement technique and often must be reduced by placing the system in a box. Similarly vibration also causes random errors and necessitates interferometers be mounted on vibration isolating supports such as air legs. The source of random errors must wherever possible be determined, as it is crucial to minimise their effect on experimental accuracy.

Systematic errors are those which do not vary between repeated experiments, and can be thought of as a bias or offset in the measured results. Whereas often random errors in a system can be removed or mitigated by averaging over very large data sets this is not the case for systematic ones. An interferometric example of this type of error can be easily imagined when one considers an experiment designed to measure the flatness of a plane mirror. A possible source of systematic error here can be one from the misalignment of a collimating optic, or irregularities in the beam splitting element used. Inhomogeneities in the refractive index of a lens, for example, will remain constant throughout measurements of the plane surface and give the appearance of aberrations when there are none. This example also leads to the need to consider relative vs absolute measurements.

The previous example of systematic errors discussed is an example of a relative measurement, as the test surface is compared to a known reference surface. It is unknown whether any observed error corresponds to errors in

the test piece, errors in the reference piece, or a combination of the two. It is only possible to determine this by removing elements from the interferometer and comparing results. If, for example, the test piece is a refractive optic, then the results from the measurement of the part compared to the measured results in the absence of the part give an indication of the source of the observed irregularities. Another method of compensating for this is to use an independent absolute measurement technique, such as profilometry.

2.8 Comparison of surface metrology techniques

With the wide variety of different techniques available for measuring surfaces it is important to ensure the ideal method is employed for the desired circumstance. Interferometry for example provides high resolution, full field imaging of a surface part but is highly sensitive to environmental conditions. This complicates the implementation of it in situations with large degrees of vibration or air turbulence, such as factory environments where a more suitable method may be employed. The properties of the part under test may also dictate which technique is ideal, for example a non-reflective part cannot be analysed using optical techniques and hence could best be measured using profilometry. Whilst scanning probe microscopy is capable of highly accurate measurements down the angstrom scale, it requires a conductive surface in order to operate.

Contact profilometry holds a significant advantage over optical measurement techniques as it directly measures height variations over a test piece. Factors such as correct conjugation or diffraction effects which may induce wavefront aberrations in interferometry are not a concern, and as previously mentioned the presence of contaminants on the surface which would block out incident light may not affect the accuracy of direct measurements through profilometry. Contact methods do hold a unique disadvantage compared to other techniques

however in that they can potentially damage the part being measured. Careful considerations on contact force and measurement speed need to be made in order to efficiently measure the part, as a slow scanning speed will minimise the risk of damage but at the cost of taking a long time to achieve results.

Shack-Hartmann wavefront sensing is capable of obtaining results significantly faster than profilometry techniques, by imaging the full field at once rather than raster scanning over a series of points. This method is highly versatile and is not hindered by the need to eliminate damage potential as in profilometry or limited by the requirement of a coherent source or path length matching that interferometry requires. Stylus profilometers are also limited in that they only measure the height variations of a surface, whereas optical methods are capable of measuring total wavefront deformations induced from sources such as refractive index inhomogeneities in the optic, allowing measurements to be taken of transmissive elements. Shack-Hartmann sensor's major limitation is in the comparatively low spatial resolution they are able to achieve. The resolution of this technique is dependant on the size of the lenses being used to obtain the spot pattern, and cannot resolve features which have a size significantly smaller that this element.

Whilst interferometry requires tight environmental controls and careful alignment of reference piece it is capable of achieving very high resolution and fast full-field images of test parts. Through the use of nulling elements which will be covered in more detail in the following chapter, it is also capable of measuring a wide variety of wavefront shapes through by modifying the reference wavefront shape allowing a very high dynamic range to be reached. Unlike profilometry however which measures the absolute surface height variations over an area interferometric techniques only record the relative differences from a known surface. This requires the careful calibration of interferometric systems in order

to minimise the systematic error on results present. Interferometry also is not a universally successful wavefront measurement technique due to the need to unwrap the recovered modulo 2π, phase unwrapping is a complex problem and no software currently exists that is capable of successfully unwrapping these phase maps in all cases. Whilst Shack-Hartmann wavefront sensing also relies on post exposure algorithm implementation in order to obtain a wavefront phase map these algorithms are generally highly successful at obtaining a recovered wavefront profile. Despite these problems interferometry is a highly useful technique, and once calibrated it is possible to insert both transmissive and reflective parts into the system for fast, accurate metrology.

Chapter 3

Advanced interferometric techniques

3.1 Introduction

This chapter will introduce two significant advances to the technique of experimental interferometry; phase shifting interferometry and holographic interferometry, as well as the background behind them. First it will cover the theory behind holography and birefringence in liquid crystals as these provide the basis for discussion of holographic interferometry at the end of the chapter. Phase shifting interferometry will also be covered as it allows for extensive analysis of interferometric fringe patterns and is one of the most useful tools in interferometry as well as a critical part of the experimental method implemented in chapters 4, 5 and 6.

3.2 Holography

Standard imaging techniques have been become one of the most vital tools in the field of experimental physics. They not only provide a record of an experimental result but also allow the transferring of data into computers without which progression of virtually all research would have been stymied. Nevertheless conventional imaging techniques only record half of the information available to us by providing a map of the intensity measurements of recorded light waves whilst completely ignoring any three dimensional phase information from the incident wave. The Nobel Prize for Physics in 1971 was awarded to Denis Gabor for his work in developing a new type of imaging which allowed for the reconstruction of three dimensional images using a two step process involving a type of optic he referred to as a 'hologram' [43]. The first stage in this process involves the recording of a fringe pattern formed by the interference of a plane reference wave with the wave resulting from the scattering pattern formed from interacting with a small object. If this fringe pattern is recorded on a thin transmissive film then by providing an illumination source corresponding to the original reference beam, a virtual image is created of the original source of the scattering behaviour. The advancement of the field of holography was not appreciable however until the advent of the laser, the degree of coherence given by laser light allowed for the furthering of the technique[44].

Consider the interference of two plane waves as seen in figure 3.1. Where the maximum from one plane wave coincides with the maximum from the tilted wave we get a bright patch due to the constructive interference between the two maxima. Correspondingly we get a very dark patch at the point where the crest from one wave destructively interferes with the trough of the other. As these waves progress forward they remain in phase and thus the resulting fringe pattern is static. Now if the fringe pattern here is recorded on a transmissive

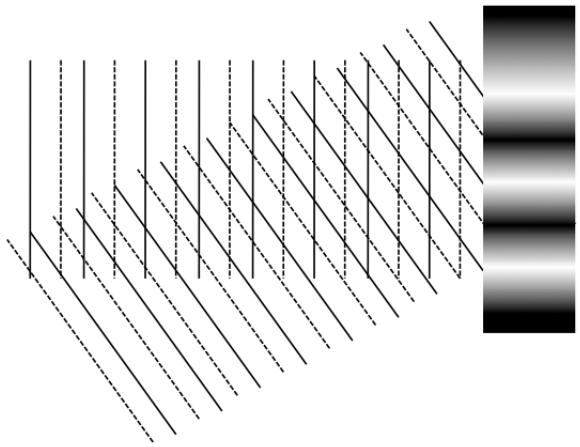


Figure 3.1: Construction of a simple sinusoidal hologram by the recording of fringes on a transmissive film caused by the interference of two plane waves with a tilt angle between them.

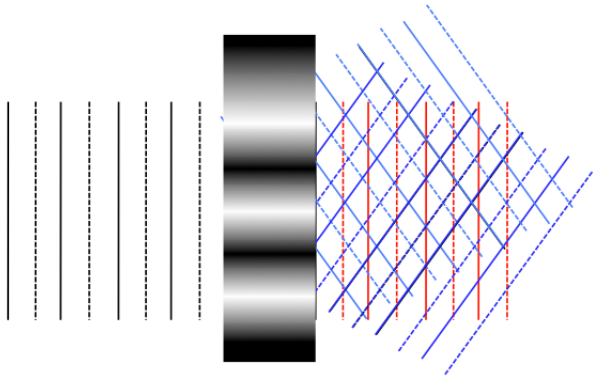


Figure 3.2: Reconstruction of a plane wave by illumination of a transmissive sinusoidal hologram with the same reference wave as used in the recording procedure. This results in the propagation of three beams; one zeroth order beam and two first order beams.

film and then re-illuminated with the original reference wave, the thin film hologram gives us three diffracted waves as seen in figure 3.2, one corresponding to the zeroth order unaffected beam and two first order ones. A corresponding mathematical description can be considered simply as follows. The interference of a background plane reference wave, E_B, with a tilted object beam, E_O, is described as

$$E_B = E_{B0}e^{i(\phi_B)},$$

$$E_O = E_{O0}e^{i(\phi_O)},$$

$$I = |E_B + E_O|^2 = E_{B0}^2 + E_{O0}^2 + E_{B0}E_{O0}(e^{i(\phi_B - \phi_O)} + e^{i(\phi_O - \phi_B)}), \quad (3.1)$$

where E_B and E_O are the electric fields of the background and object waves respectively, ϕ_B and ϕ_O the phase of the background and object waves and I is the recorded intensity pattern. This intensity pattern is the recorded hologram, and reconstruction of the original wave can be obtained by illuminating the hologram with a reconstruction beam, E_R gives:

$$E = E_{R0}(E_{B0}^2 + E_{O0}^2)e^{\phi_R} + E_{R0}E_{B0}E_{O0}(e^{i(\phi_B - \phi_O + \phi_R)} + e^{i(\phi_O - \phi_B + \phi_R)}), \quad (3.2)$$

When the reconstruction illumination is identical to the original background wave:

$$E = (E_{B0}^3 + E_{B0}E_{O0}^2))e^{\phi_B} + E_{B0}^2E_{O0}(e^{i(2\phi_B - \phi_O)} + e^{i(\phi_O)}), \quad (3.3)$$

The first term in equation 3.3 is the zeroth order term, which can be generally ignored as it contains no information regarding the object wavefront ϕ_O. The second term encompasses the $+/-$ first order diffracted wavefronts, one

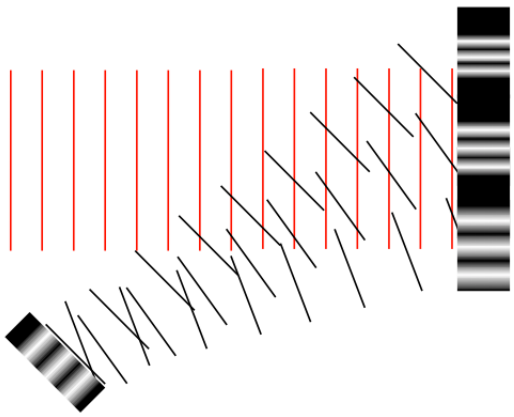


Figure 3.3: The recording of a more complicated hologram, using the interference of a plane wave with a wave corresponding to the diffraction pattern from a sinusoidal hologram.

containing the original object wave information and the other containing the negative along with a tilt term ϕ_B.

Now consider a slightly more complicated example, as in figure 3.3, where instead of a tilted plane wave we have light coming from an optical element with some feature, say a sinusoidal interference pattern like the one resultant from the previous example. Plane wave illumination of this will give us the three diffracted waves as before, and each will interfere with the plane wave in the detector plane to give an interference pattern with three different spatial frequencies. Reconstructing the image of this interference pattern with the original plane wave will now result in a series of diffracted waves, and a virtual reconstruction of the original sinusoidal grating can be observed at the point where the first order from each separate interference pattern coincide as seen in figure 3.4.

As intimated previously, the first holographic plates were created by recording the two-dimensional fringe pattern formed by the interference of the object and reference wavefronts on a thin transmissive element[45], however there are

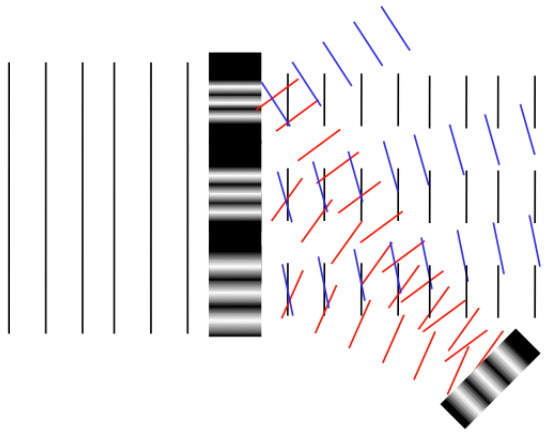


Figure 3.4: Reimaging of a hologram caused by the interference of a structured wave with a plane reference wave. In this case the positive first order diffractive beam gives a reimaging of the original sinusoidal grating at the point where the first order beams from all three differing spatial frequency sinusoidal elements combine.

various other methods of creating them. Volume phase holograms [46] are those where the thickness of the plate used is larger than the fringe spacing or finest detail on the interference pattern seen. Whereas with thin film holograms only the fringe pattern on the surface is recorded, with volume phase holograms the fringe pattern is recorded throughout the medium. Such a hologram can allow for significantly higher throughput than those recorded on thin plates as they allow for the full diffraction of the incident light into desired wavefront. The downside to using such a hologram is that the light must be steered correctly as the scattering effect follows Bragg's Law and therefore different wavelengths of light will be diffracted at different angles.

Computer Generated Holograms (CGHs) are holograms that are not created through the direct recording of an interference pattern [47][48][49]. Instead the pattern is created computationally before being written onto an optical element and inserted into the test system. For testing of accurate surface shapes this

method can offer a significant advantage over classical holography due to the lack of need for the reference object to actually exist. Instead the reference wavefront can be created from design specifications of the test piece and then inserted into the system allowing for the measurement of complex surface shapes with reference to their designed shape rather than the shape of a master optic thus eliminating errors in the production of the master. One method of calculating the hologram is to determine the far field light distribution desired and then Fourier transform this back to the reconstructed lens plane. Another method of designing a CGH is by considering every point on the hologram as a point source emitter, similar to the case of a Fresnel Zone plate.

One widely used modern approach for the writing of CGHs is to use electron beam lithography to plot the diffraction pattern on to a thin piece of photoresist. Whilst this method allows for high accuracy holograms to be produced, e-beam machines only accept standard sized substrates, which is inconvenient for creating holograms used in metrology as they are needed in a wide variety of shapes and sizes. CGHs can also be written using a thermochemical writing technique [50, 51], where a glass substrate is given a thin coating of chromium before having the holographic design written onto it with a laser. The laser oxidises the chromium coating before the substrate is placed in an acid bath, and as the unoxidised chromium reacts more rapidly with the caustic bath after a short immersion time the only the pattern of oxidised chromium written onto the substrate by the laser remains. A drawback of both these processes is that they create static holograms suitable for the testing of only the specific optic that they are designed for. As shall be described later however it is also possible to create holograms on spatial light modulators and therefore be inherently reprogrammable.

3.3 Liquid crystals

Whilst simple crystalline structures refract transmitted light simply with no unusual effects, there are also many crystals which exhibit the property of birefringence, where two orthogonal polarization states are refracted differently. Inhomogeneities in the structure of these materials cause the molecular bonding forces between layers of the crystal to be different in different directions, and since the propagation of light in the medium is through the interaction of the wave with the molecular structure the disparity in bond strength results in orthogonal polarizations of light being absorbed and re-emitted in different ways. The anisotropy in binding forces within the crystal directly leads to an anisotropy in the refractive index of the medium, with the result that a linearly polarized wave transmitted through the medium can be split into two orthogonally polarized beams. A uniaxially anisotropic crystal is defined as one where the optical properties of the crystal are identical in directions perpendicular to a single axis, known as the optical axis. Linear polarization states parallel and perpendicular to this axis will have different refractive indices and thus will be transmitted differently. A linear wavefront polarized perpendicular to the optical axis will emerge unmodified from that predicted using Snell's law and thus is termed the ordinary ray (o-ray). A wave polarised with a linear component parallel to the optical axis will be deflected and emerge at an angle dependant on the relative angle between polarization state and optical axis, and is termed the extraordinary ray (e-ray).

In the special case of the incident light traveling in a direction orthogonal to the optical axis, the deflection of the e- and o-rays is identical so there is no splitting of the beam, however the e- wave will still propagate through the medium at a different speed to the o-ray, and the resultant wave exiting the crystal will be a superposition of the two. As the e-ray lags with respect to the

o-ray the effect of this is to introduce a phase retardation, $\delta\phi$, to the outgoing wave of the amount

$$\delta\phi = \frac{2\pi}{\lambda} d(n_o - n_e), \tag{3.4}$$

where λ is the wavelength of light, d the thickness of the medium and n_o and n_e are the refractive indices of the material for the o- and e- rays. This is the principle behind the operation of waveplates, where the thickness of the crystal determines the extent of the phase modification of the incident beam. With a full wave plate the lag of the e-ray compared to the o-ray is 2π, and correspondingly for a half wave plate the lag is π which is equivalent to rotating the polarisation state of a linearly polarised beam through π. Whilst these preserve the linearity of the polarization, in the case of a quarter wave plate this is not so. This corresponds to a relative lag in the e-ray of $\pi/2$ and allows for conversion of linearly polarized light into elliptically polarised light. In the case of linearly polarised light the e-ray and the o-ray oscillate in phase with each other, when the phase value of one peaks, so does the other. By adding a $\pi/2$ phase shift into the e-ray we cause it to oscillate out of phase with the o-ray, so that the combined polarisation vector progresses with constant amplitude and varying angle over time.It is apparent from equation 3.4 that waveplates are chromatic optics, only functioning correctly when used with light of the correct wavelength. Passing a linearly polarised beam through a half wave plate designed for use with a different wavelength will result in the outgoing wave having elliptical polarization rather than the desired linear.

Birefringence in crystals is not limited to solid crystalline structures[52][53]. Liquid crystals, containing numerous crystalline molecules which have varying degrees of positional and rotational freedom depending on temperature and molecular size, also show birefringent properties. One of the most common and widely used type of liquid crystal are those in the nematic state. In this

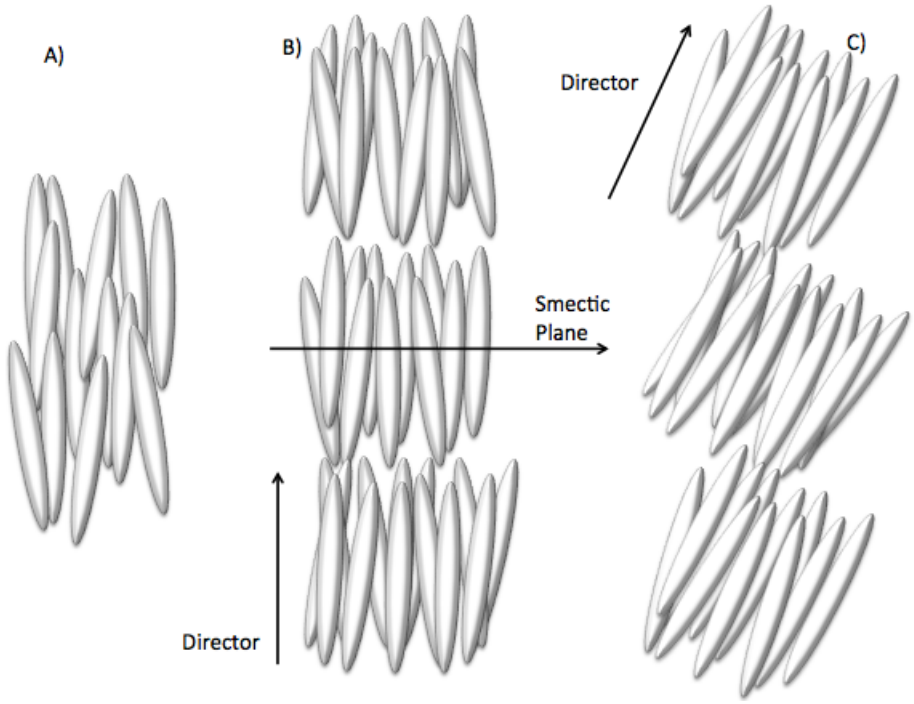


Figure 3.5: Configuration of liquid crystal molecules in three commonly used LC mesophases; A) In the nematic phase the molecules display no positional configuration but all possess a similar degree of rotational orientation. B) Smectic A, or smA, phase is one in which the molecules are arranged in a series of layers known as the smectic plane, and the director is orientated perpendicular to that plane. C) The Smectic C phase is similar to the smA state but here the director is inclined at a tilt with respect to the smectic plane.

state the molecules observed are long and thin with an approximately random positional distribution throughout the sample as seen in figure 3.5a. What they do posess however is a high degree of rotational alignment, with the long axes of the molecules all aligned in a particular direction. Whilst there is a degree of slight anisotropy in the angular alignment, the application of an external electric field can cause them to become aligned parallel and in this state they display the characteristics of a uniaxial crystal.

As the word “nematic” comes from the greek “nema” which means thread, other liquid crystal states are named from their descriptive properties. The

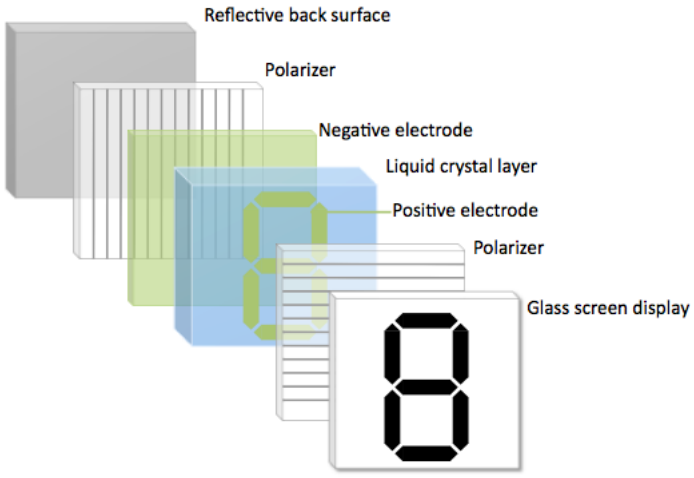


Figure 3.6: Schematic of a nematic liquid crystal display unit. Where there is an electric field acting upon the liquid crystal it fails to rotate the polarization of the light and therefore it is blocked by the cross polarizers giving dark patches on the front of the display.

smectic liquid crystal state is found at a lower temperature than the nematic state and is so named from the greek for soap, as the molecules now contain a degree of positional alignment as well as rotational, forming layers which slide over each other smoothly. As there is a high degree of rotational alignment in LC states it is convenient to define the director, a dimensionless vector which points along the direction of preferred orientation of the LC molecules. The smectic A state is defined as the state where there the director is aligned perpendicular to the plane of the smectic layers, and the smectic C (or smC) state is defined as the state where the director is aligned with a tilt with respect to the smectic plane. A visual representation of these states is shown in figure 3.5b and 3.5c.

One of the most commonly used types of LC display (LCD) is based upon the optical behaviour of twisted nematic crystals. A nematic LC layer is sandwiched between two electrodes, with a polarizer on the top layer to allow through incident light, and a mirror on the bottom layer in order to reflect it back

through the device. The liquid crystal is orientated via grooves on one side of one of the polarisers, and grooves running perpendicularly to this on a second polariser placed on the other side of the liquid crystal layer, as seen in figure 3.6. The nematic properties of the crystal which cause a high degree of rotational alignment will cause a gradual twisting of the crystal layer between the two polarisers. When linearly polarised light is incident upon the liquid crystal layer the twisted structure will rotate the polarization by 45°, before reflecting off the back surface, passing through the crystal a second time to realign it to the same linearly polarised state as when it was incident. When an electric field is passed through the nematic crystal however it untwists and ceases to add the 45° rotation to the wavefront causing it to be blocked by the second polariser and the cell in question shows up on the display as black. Liquid crystal devices have also been demonstrated to be capable of being used as retardance devices, where a pixellated display of LC cells can retard the phase of incident light to create an arbitrary phase profile to a high degree of accuracy [54, 55, 56]. This method of wavefront shaping requires a full understanding of the dynamic range of the device as well as the voltage response of its retardance properties, similar to the need for a characterisation of the hysteresis properties of deformable mirrors. Deformable mirrors are a popular device used in wavefront manipulation and aberration correction[57], where multiple actuators below the surface are used to modify the form of the surface shape. They can allow for a rapid reconfiguration of the wavefront profile however one of their major limitations is that they are unable to create highly sloped or discontinuous profiles as this would damage the surface coating. Crucially however, whereas deformable mirrors have these very tight limitations on the magnitude of the stroke that they are capable of producing, LCs allow for much steeper phase gradients to be produced, to the extent that it is possible using a binary device to create phase maps containing

discontinuous phase jumps. The limitation these devices have in severity of gradient that they can display accurately is that they need to allow for adequate sampling of the wavefront.

It is also possible to create an LCD using ferroelectric liquid crystals (FLCs) rather than twisted nematic ones. In this case the state used is a variant of the smC state called the smC* state, where the director processes around an axis perpendicular to the smectic plane and thus shows a degree of chirality giving rise to an overall molecular polarisation of the crystal. Naturally if the crystal is allowed to form freely the helical structure will cause there to be no ferroelectric effect as the net polarization will average out, however [11] showed that this can be suppressed and thus it is possible to create a ferroelectric crystal with each layer having identical orientation which can be influenced. Upon the application of an electric field the orientation of the director will be changed as the polarization of each molecule is directly coupled to its orientation. For use in LCD technology this can be implemented similarly to how twisted nematic crystals are used, using two cross polarizers on either side of the LC layer and then electrically switching the LC to allow one state to pass and one state to be blocked.

Another use of FLC devices are as waveplates, the orientation of molecules of the LC layer induces a phase retardation the same as in nematic LCs, but the switching speed between the states is vastly increased allowing a much faster refresh rate. Furthermore, whilst nematic LC devices allow for the production of a continuous phase profile this is not possible in FLC devices. The response of nematic LC devices, the degree of phase modulation with respect to the applied voltage, must be fully characterised whereas due to the binary nature of the FLC there is a significantly reduced need to understand perfectly the greyscale response of the device. Whilst they are unable to produce a continuous phase

pattern, it is possible to use FLC devices to create an analogue phase pattern [10] as will be discussed towards the end of this chapter.

3.4 Phase shifting techniques

Metrology methods mentioned in the previous chapter which involved the interpretation of fringe patterns, were in the past limited by the lack of accuracy in their examination, as measurements would be done by hand on photographic plates or films. Now however all data tends to be processed electronically, and thus we can achieve a level of precision far surpassing that gained at the time such metrology techniques were developed. The primary piece of information given to us through interferometric fringes is the distance between two points corresponding to an optical path difference (OPD) of half a wavelength, however in order to accurately determine the distance between two fringe maxima the detection of the fringe centre must be obtained with correspondingly high accuracy. The measurement of small fringes naturally leaves more room for error, as any inaccuracy corresponds to a much greater OPD than that which would be obtained from a much larger fringe. This leads to a trade-off between accurate fringe detection and the accompanying lack of information which is associated with images that contain only few maxima.

A second problem encountered with this technique is that detection of fringe maxima is vastly complicated by any intensity variations in the beam. These can arise from beam inhomogeneities, marks in optical elements used, or irregularities in the detector. Yet another critical limitation of static fringe analysis techniques is that they are incapable of determining the orientation of the phase pattern being measured, concave and convex features appear indistinguishable. Phase Shifting Interferometry (PSI) however does not suffer from these limitations. A temporally varying element is added into one of the interferometer

arms giving the visual effect of the fringes shifting across the image. As multiple measurements are taken for varying phase offsets the effect of static intensity variations are nullified, and furthermore allows for image processing of very large or partial fringes which would be unavailable in static fringe analysis techniques.

One of the most straightforward ways in which to apply a phase shift into an interferometer is to insert a mirror, mounted on a piezoelectric stage (PZS), in one arm of the system which changes position according to an applied electric field and therefore allows for programmable translation of the mirror. A gradual variation in the applied electric field allows for a programmable phase shift function to be created. One major potential source of error that arises from using a PZS is that the response from the applied electric field may not be linear, so a thorough characterisation of the device is required in order to provide the regular phase shifted steps needed for the recovery of the wavefront shape.

An alternate method of applying a phase step in an interferometer is to use a retardation plate [58] to modify the polarization state of linearly polarized light. Rotation of a $\lambda/2$ wave plate in one arm of the interferometer will give a phase step equal to twice the rotation of the plate, i.e. a rotation of the plate by $\pi/4$ will cause a $\pi/2$ phase shift. A limitation of using a wave plate instead of a PZS however is that a well characterised PZS device allows for a significantly higher level of accuracy in the phase step being introduced by the device. Another weakness of using a half wave plate is that the optical element requires a high level of manufacturing accuracy, deviations in the homogeneity of the element will induce an aberration as it is rotated that varies between phase shifted frames and hence cause significant errors in the resultant phase map.

The interference of two monochromatic coherent beams, a reference wave Ψ_r

and test wave Ψ_t , which can be expressed as

$$\Psi_r(x,y) = E_r(x,y)e^{i(\phi_r(x,y)-\delta(t))},$$

and

$$\Psi_t(x,y) = E_t(x,y)e^{i(\phi_t(x,y))},$$

is given as:

$$\begin{aligned} I &= E_r^2 + E_t^2 + 2E_rE_t\cos(\phi_t - \phi_r + \delta(t)), \\ I &= I_o + I_m\cos(\phi_t - \phi_r + \delta(t)), \end{aligned} \tag{3.5}$$

where ϕ_r and ϕ_t are the reference and test phases, E the electric field and $\delta(t)$ is the time dependant frequency offset introduced. If the frequency shift term is linear with respect to time, we can see that the term I_o is the offset to the intensity term and I_m is a modulation to the observed fringe intensity.

Interferograms can be recorded in one of two ways when implementing PSI. The first, phase stepping, simply involves applying a discrete phase shift before recording the fringe pattern, then repeating as many times as needed to obtain the requisite number of images. This can be slow however, it is necessary to allow the system to stabilise after the phase shift has been implemented in order to allow the degradation of vibrations induced by the method of the shift. Manual methods of phase shifting via waveplate rotation will cause large vibrations in the system and even mechanical methods such as the use of a PZS may still result in small scale oscillations in the optical elements. An alternative method for collecting data, integrated bucket PSI, is to apply a linearly varying phase shift into the interferometer and collecting data whilst it progresses [59], with the visual effect of the fringes moving across the detector's field of view. Whilst the phase shift varies over the period of the detector's integration, the speed

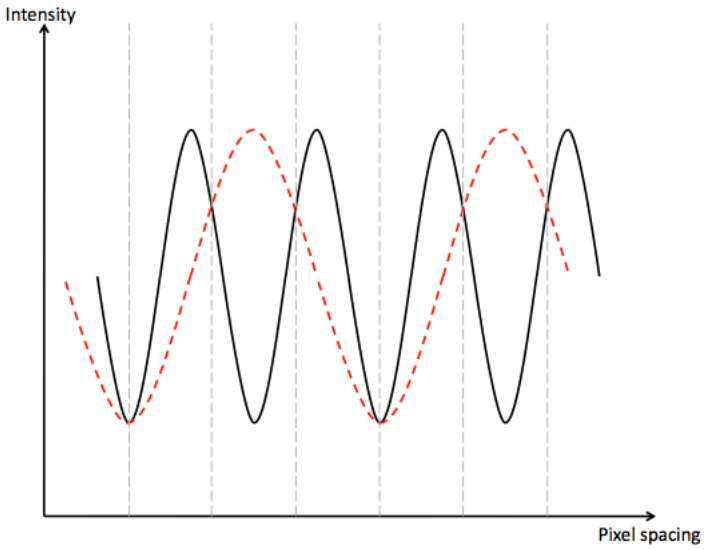


Figure 3.7: When the main signal, in black, is sampled at a frequency more than half the signal period, then the recorded intensity values can be aliased to appear as if they were registering a lower frequency signal as seen in red.

of the the linear shift can be varied both to account for variables such as the cameras specific running speed, and also to allow for longer integration times thus lessening the need for a high gain and minimising the influence of detector response nonlinearity. Regardless of whether the interferograms are recorded using phase stepping or integrated bucket PSI, it is important that the signal be sufficiently well sampled so as not to observe aliasing effects caused by the detection of a high frequency signal being indeterminate from that of a lower frequency signal, as seen in Figure 3.7. For this case we can consider the intensity pattern incident on the detector to be:

$$I(x, y, t) = I_o(x, y) + I_m(x, y) \cos(\Delta\phi(x, y) + \theta(t)), \tag{3.6}$$

where $\theta(t)$ is the time varying phase shift, and $\Delta\phi$ is the phase map. The intensity recorded for a period corresponding to a change in phase shift Θ is

then:

$$I_n = \frac{1}{\Theta}\int_{\theta_n-\Theta/2}^{\theta_n+\Theta/2}[I_o + I_m\cos(\Delta\phi+\theta)]d\theta, \tag{3.7}$$

$$I_n = I_o + \frac{I_m}{\Theta}[\sin(\Delta\phi+\theta_n+\Theta/2) - \sin(\Delta\phi+\theta_n-\Theta/2)], \tag{3.8}$$

$$I_n = I_o + \frac{I_m}{\Theta/2}\sin(\Theta/2)\cos(\Delta\phi+\theta_n), \tag{3.9}$$

$$I_n = I_o + I_m sinc(\Theta/2)\cos(\Delta\phi+\theta_n), \tag{3.10}$$

where θ_n is the phase shift in the middle of the integration period. From equation 3.10 we can see that the effect of integrating over a varying phase shift is to introduce an extra modulation term to the fringe pattern. Equation 3.10 can be considered a general equation for PSI, as in the specific case where Θ tends to zero it corresponds to phase stepping interferometry. Over a period of $\Theta = 2\pi$ the fringe modulation tends to zero, as seen in figure 3.8, as we have just integrated over the whole period of the phase shift, however for smaller values such as $\Theta = \pi/2, \pi/4$ then a high level of fringe contrast can still be maintained.

There are many algorithms for retrieving the resultant phase distribution from the combination of the two wavefronts, $\phi_t - \phi_r = \Delta\phi$, some examples of which shall be discussed here. The first that shall be mentioned is the simple four step technique which relies on the capturing of four intensity measurements with a $\pi/2$ phase step between them,

$$I_1 = I_o + I_m\cos(\Delta\phi), \tag{3.11}$$

$$I_2 = I_o + I_m\cos(\Delta\phi+\pi/2) \qquad = I_o - I_m\sin(\Delta\phi), \tag{3.12}$$

$$I_3 = I_o + I_m\cos(\Delta\phi+\pi) \qquad = I_o - I_m\cos(\Delta\phi), \tag{3.13}$$

$$I_4 = I_o + I_m\cos(\Delta\phi+3\pi/2) \qquad = I_o + I_m\sin(\Delta\phi), \tag{3.14}$$

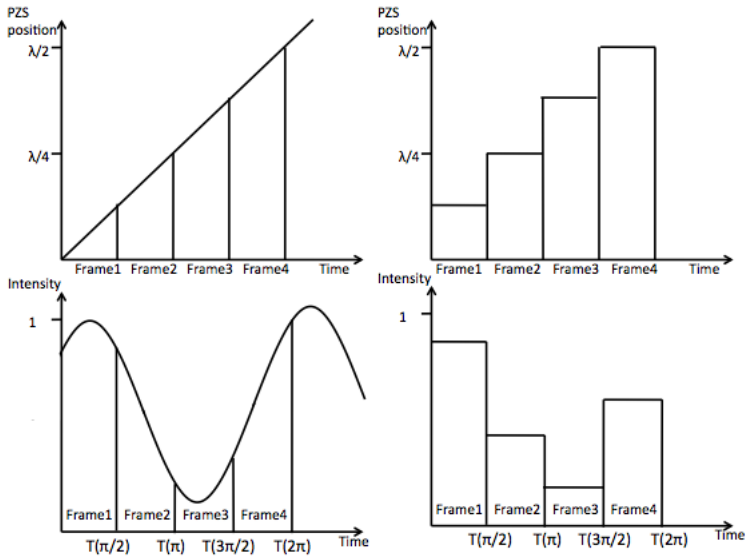


Figure 3.8: Recorded intensity of phase shifted frames and how they differ with phase shifting method. On the left is a continually varying phase shift where the camera records over a varying intensity level, and on the right the case where the phase is shifted separately for each individual frame so the camera records a constant intensity value. When the phase shift varies over time, the camera records the average intensity over the integration period and therefore we see a reduction in observed fringe modulation.

From here we can combine equation 3.11 with equation 3.13, and equation 3.12 with equation 3.14:

$$I_1 - I_3 = 2I_m \cos(\Delta\phi), \tag{3.15}$$

$$I_4 - I_2 = 2I_m \sin(\Delta\phi). \tag{3.16}$$

Allowing us to recover the phase distribution, $\delta\phi$, from the ratio of equation 3.16 to equation 3.15:

$$\Delta\phi = \tan^{-1}\left(\frac{I_4 - I_2}{I_1 - I_3}\right). \tag{3.17}$$

It is also possible to obtain an expression for the modulation of the fringe

intensity $\gamma_0 = I_m/I_o$ trivially from equations 3.11 - 3.16:

$$\gamma_0 = \frac{2\sqrt{(I_1 - I_3)^2 + (I_4 - I_2)^2}}{I_1 + I_2 + I_3 + I_4}. \tag{3.18}$$

As there are three unknowns in the equation for the recovered interferogram; I_o, I_m and $\Delta\phi$, it follows that the minimum number of frames required to recover one of these is three. One such algorithm involves the capture of frames with $\pi/2$ phase shifts between them [60]:

$$I_1 = I_o + I_m \cos(\Delta\phi + \pi/4), \tag{3.19}$$

$$I_2 = I_o + I_m \cos(\Delta\phi + 3\pi/4), \tag{3.20}$$

$$I_3 = I_o + I_m \cos(\Delta\phi + 5\pi/4), \tag{3.21}$$

From which we can obtain the resultant phase map via:

$$I_3 - I_2 = 2I_m \sin(\Delta\phi)\sin(3\pi/4), \tag{3.22}$$

$$I_1 - I_2 = 2I_m \cos(\Delta\phi)\sin(3\pi/4), \tag{3.23}$$

$$\Delta\phi = \tan^{-1}\left(\frac{I_3 - I_2}{I_1 - I_2}\right). \tag{3.24}$$

Alternatively, this method can be employed using three phase shifted frames with a $2\pi/3$ phase step between each image:

$$I_1 = I_o + I_m \cos(\Delta\phi - 2\pi/3), \tag{3.25}$$

$$I_2 = I_o + I_m \cos(\Delta\phi), \tag{3.26}$$

$$I_3 = I_o + I_m \cos(\Delta\phi + 2\pi/3). \tag{3.27}$$

Allowing for a reconstructed phase map to be built using:

$$\delta\phi = \tan^{-1}\left(\sqrt{3}\frac{I_3 - I_2}{2I_1 - I_2 - I_3}\right). \tag{3.28}$$

As previously mentioned, this technique is almost entirely unaffected by errors in static intensity variation from beam non-uniformities and detector read out. This can now be seem mathematically by consideration of equation 3.5. Errors in measurement corresponding to a non-intensity dependant variation in detector sensitivity are equivalent to a spatially dependant intensity term acting upon I_o and I_m. As the resultant phase map is independent of both of these values as seen in equation 3.17 they minimally affect the accuracy. Correspondingly as the data modulation term γ_0 in equation 3.18 is determined from I_m/I_o it also is independent of static detector sensitivity variation. The second source of error from beam non-uniformity corresponds to a spatial variation in the offset intensity I_o and therefore also minimally impacts upon recovered phase map accuracy. Nevertheless irregularities in beam uniformity and detector signal response should still be considered as they could still affect the signal to noise ratio of the system.

The three- and four-step phase recovery algorithms mentioned above are used when the phase shift between frames is known, however when using a PZS there is the potential for errors in the linearity of the response that result in the phase steps being of an unknown value. The Carre algorithm [61] is an example of a phase recovery method which is designed for use in the presence of an unknown yet constant phase shift, α, and requires the use of four shifted

frames:

$$I_1 = I_o + I_m \cos(\Delta\phi - \frac{3\alpha}{2}), \tag{3.29}$$

$$I_2 = I_o + I_m \cos(\Delta\phi - \frac{\alpha}{2}), \tag{3.30}$$

$$I_3 = I_o + I_m \cos(\Delta\phi + \frac{\alpha}{2}), \tag{3.31}$$

$$I_4 = I_o + I_m \cos(\Delta\phi + \frac{3\alpha}{2}). \tag{3.32}$$

The phase shift can be calculated from these by using:

$$\tan\left(\frac{\alpha}{2}\right) = \sqrt{\frac{3(I_2 - I_3) - (I_1 - I_4)}{(I_2 - I_3) + (I_1 - I_4)}}, \tag{3.33}$$

which can be used in recovering the overall phase map:

$$\tan(\phi) = \tan\left(\frac{\alpha}{2}\right)\frac{(I_2 - I_3) + (I_1 - I_4)}{(I_2 + I_3) - (I_1 + I_4)}. \tag{3.34}$$

Before considering further analysis on the phase maps produced using these phase shifting techniques it is important to note that, in their simplest form, the three- and four- step algorithms appear to only allow for $\lambda/2$ phase recovery due to the tan function being defined over the space $-\pi/2$ to $\pi/2$. It is simple enough however to account for this by considering the sign of the numerator and denominators given in equations 3.17, 3.24 and 3.28 and comparing them to table 3.1 in order to correctly determine the range in $0 - 2\pi$ for which each recovered phase value belongs. This then allows the phase map to be unwrapped and transformed into a value for the surface height profile.

Whilst, as mentioned earlier, linear irregularities in detector response can be ignored so long as they do not result in a severely degraded signal to noise ratio, the same cannot be said for higher order variations. Consider for example

Sin	Cos	Correction to phase	Phase range
+	+	ϕ	$0 - \pi/2$
+	-	$\phi + \pi$	$\pi/2 - \pi$
+	0	$\pi/2$	$\pi/2$
-	-	$\phi + \pi$	$\pi - 3\pi/2$
-	+	$\phi + 2\pi$	$3\pi/2 - 2\pi$
-	0	$3\pi/2$	$3\pi/2$
0	+	0	0
0	-	π	π

Table 3.1: Look-up table for corrections to phase to account for the tan function being defined from $-\pi/2$ to $\pi/2$ in three- and four-step interferometry[60].

the case of a second order error to detected intensity:

$$I' = I + \gamma I^2. \tag{3.35}$$

The presence of such an error requires a modification of the equation of the interferogram to become:

$$\begin{aligned} I &= I_o + I_m \cos(\Delta\phi + \theta) + \gamma(I_o + I_m \cos(\Delta\phi + \theta))^2, \\ &= I_o + I_m \cos(\Delta\phi + \theta) + \gamma I_o^2 + 2\gamma I_o I_m \cos(\Delta\phi + \theta) + \gamma I_m^2 \cos(\Delta\phi + \theta)^2, \end{aligned} \tag{3.36}$$

$$= (I_o + \gamma I_o^2) + I_m \cos(\Delta\phi + \theta) + 2\gamma I_o I_m \cos(\Delta\phi + \theta) + \frac{\gamma I_m^2}{2}[\cos(2(\Delta\phi + \theta)) + 1], \tag{3.37}$$

$$= I_o' + I_m(1 + 2\gamma I_m)\cos(\Delta\phi + \theta) + \frac{\gamma I_m^2}{2}[\cos(2(\Delta\phi + \theta)) + 1]. \tag{3.38}$$

By substituting equation 3.38 into equations 3.15 and 3.16 we can see that the phase shift in the final term is now π and therefore eliminated when recovering the final phase map. As the recovered phase pattern is completely independent of the modified phase term in equation 3.38 the only effect a second order nonlinearity has is to reduce the modulation of the fringes further. A further

extension of this is that when using four-step PSI any even power non-linear terms are eliminated from the final result, however all odd power terms are still present. Correspondingly three-step PSI allows for the removal of all non-linearities that are of an order which is a multiple of three, and therefore in order to remove a source of possible error when using PSI the response of the detector needs to be evaluated and the algorithm to retrieve the phase map chosen accordingly.

It is also possible to modify the phase shifting system so that all of the required interferograms are processed simultaneously. One method of accomplishing this is by passing the common path output arm of a Twyman Green interferometer through a quarter wave plate and placing a pixellated mask over the camera [62]. This mask is designed so that it introduces a phase shift on the incoming beam, with four different orientations of the grating being superposed over neighbouring camera elements. This allows for four step interferometry algorithms to reconstruct the wavefront without the need for four separate interferograms. Another method of implementing instantaneous PSI is to use a common path interferometer based around a 4-f system[63]. Two windows are aligned so that when imaged overlap the Fourier plane, and quarter wave plates placed in front of them so the wave from each window is circularly polarised with opposite handedness. By placing a phase grating in the Fourier plane four different images can be re-imaged, each of which corresponds to the overlap of the two windows however with a different phase difference between the two. From this a standard four step PSI algorithm recovered the phase map.

Instantaneous phase shifting methods provide a significant advantage over the stepping methods as they drastically reduce the extent to which vibrational and atmospheric effects reduce accuracy. They also provide a speed advantage due to the lack of need to take multiple measurements. Where they show a dis-

tinct disadvantage is in the recording method used to obtain the images. This imposes a reduction in accuracy over the full separate imaging of individually shifted frames. Either they require the imaging of four fringe patterns onto a single camera which lowers the resolution of the system, or they are recorded on different cameras which drastically lowers the throughput as well as complicating the optical design. Integrated bucket PSI is potentially the most suitable method for low light circumstances as it allows for a protracted image acquisition, however this comes at a cost of low speed. Phase stepping interferometry bridges the gap by allowing for the ability to record interferograms at moderately high speeds with no loss of resolution due to the need to record multiple phase stepped frames at once.

3.5 Holographic interferometry

One of the primary concerns regarding the use of interferometry as a tool in metrology is the dynamic range of the system. When considering other techniques such as stylus profilometry the dynamic range merely limits the maximum range of surface deviation that the the system can measure. In interferometry however one must also consider the reduction in accuracy of a high dynamical range interferogram. Consider for example an interferogram formed by the interference of a plane wave with a highly aspheric one. The observed fringe pattern will be a series of concentric rings with very small spacings between the fringes, and the limit of the dynamical range of this would be when the fringes are no longer distinguishable from one another and therefore no meaningful phase recovery can occur. One of the problems in this scenario however is that the large amounts of spherical aberration dominate the fringe pattern seen, and the presence of small amplitude aberrations can be lost as they result in small fringe pattern features which are too small to be resolved in the pres-

ence of significant levels of other aberrations. One way to bypass this when a piece is being measured with a high and known degree of deformation in it, is to introduce an element with an identical and negative aberration in it to cancel it out and thus reveal other smaller scale aberrations. This is analogous to using a swing arm profilometry technique to improve the dynamic range of a stylus profilometer by introducing a known aberration into the system and measuring an optic with respect to that. A simple example of this technique was described in [64] where a concave test surface is being measured in a Fizeau interferometer and rather than introducing a separate concave reference surface into the system, a beamsplitter cube is used with a concave edge built into it, called a Shack Cube. When the test surface has a form perfectly matched by the curve of the Shack cube, an interferogram devoid of fringes will be produced, and is called a "null test". A Shack cube is described here as a "null compensator" as it perfectly nulls the test wavefront. Null tests, whilst ideal, are often inconvenient to implement due to the need for a specific null compensator for each wavefront under test. A consequence of this is that the path travelled through the interferometer by the reference surface is no longer identical to that travelled by the test wavefront. This naturally gives rise to an error term which appears indistinguishable from the measurement of the test wavefront and is known as re-trace error. It has been shown however that by ray tracing the experimental configuration in use, it is possible to lessen the debilitating influence of these errors on the final wavefront measurement[65].

Null compensators are often used when testing aspheric surfaces, removing the spherical deformation from the reference wave so the interferometer is simply measuring the degree of asphericity in the test piece. It is not always the case however that the aberration needing to be nulled is simple spherical aberration so it would be convenient to be able to obtain nulling optics suitable for a wide

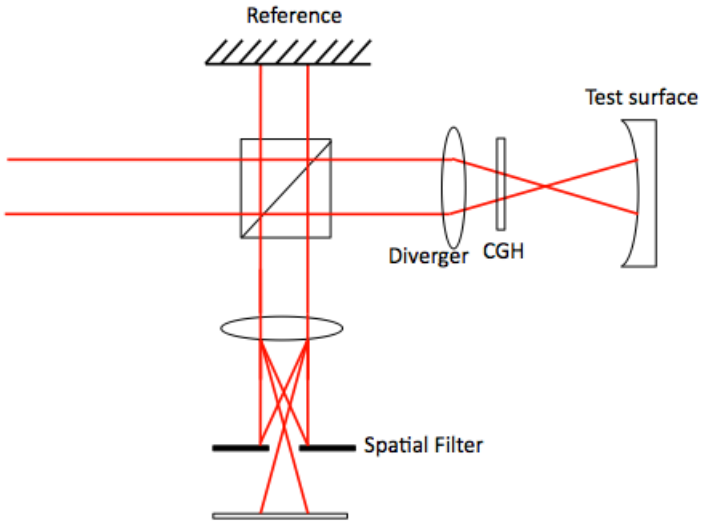


Figure 3.9: Commonly used configuration for a transmissive CGH. Here it is placed in the test arm of the interferometer and acts as part of the nulling system.

variety of wavefront shapes. As it transpires, it is indeed possible to obtain nulling optics which are essentially capable of recreating any arbitrary phase profile via the technique of holography. This can be simply implemented in cases where a large number of identical optics need to be tested, such as in industrial manufacturing processes. An interferometer can be set up using a master surface from which all other components will be copied. A holographic recording of the interference of the master surface with a known reference wavefront can be made, which, after processing, will be replaced precisely back into the system. After this the master is removed and replaced with test pieces with the result that any interference fringes seen will be due to differences in the test piece from the shape of the desired master. The test piece is then removed from the system, polished to eliminate the detected aberration before being retested, as many times as required.

It is of more use however in interferometry to use Computer Generated Holograms (CGHs) as they can be produced without the need for a high quality master reference[51] and can work with, or instead of, a null compensator [66]. They are highly flexible and capable of being used in a wide variety of interferometers and in various positions throughout the optical layout. One such configuration is seen in figure 3.9 where the CGH is placed in the test arm of the interferometer along with a nulling optic. The nulling optic partially compensates for the spherical deviation of the concave test surface and the CGH helps to accommodate for any residual curvature not eliminated by the null corrector as well as accounting for any other phase irregularities which may need removing, often due to errors in nulling optic production. The hologram must have a high efficiency when used in this configuration as it is used in double pass mode, furthermore it must be made to a high quality to avoid introducing substantial error into the test beam from optical inhomogeneities.

It has also been demonstrated [67][68] that the hologram can be effectively inserted into a Fizeau interferometer before the beam is split by the reference surface, with the reference beam being defined as the beam which comes from the zeroth order of the CGH which reflects off the back edge of the reference surface as normal, and the first diffracted order becomes the test beam as seen in figure 3.10. The CGH is designed so that the first order diffracted beam contains an identical degree of asphericity as expected in the concave test surface, with the intent of cancelling it out upon interaction so that the sphericity of the test beam matches that of the reference surface after reflection from the reference element. It is also possible to write the CGH onto the surface of a known spherical reference shape so that the reference wavefront is identical to one that was reflected from an ideal asphere [69]. One major advantage to using the CGH in this manner compared to its use in figure 3.9 is that inhomogeneities

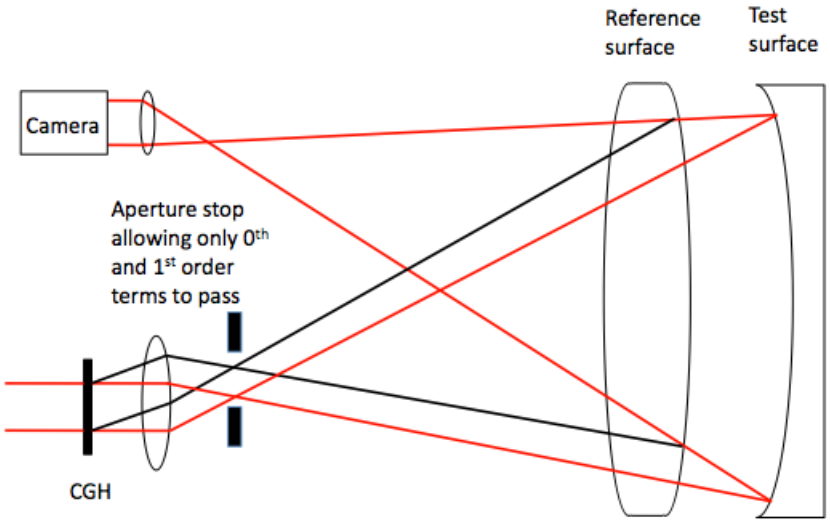


Figure 3.10: CGH implemented in a Fizeau interferometer. The Zeroth order spot from the hologram is the reference wave reflected off the back of the reference sphere and the 1st order wave constitutes the test beam and reflects off the aspherical mirror.

in the structure have little effect on the obtained result as A) the device is only used in single pass mode and B) both arms of the interferometer pass through it in almost identical paths. Furthermore aberrations seen in common path optics such as the beamsplitter are eliminated as both arms pass through them.

One major shortfall of the use of standard CGH plates in holographic interferometry is that they are designed for use only with the testing of a specific wavefront shape. In the case of the testing of telescope primary mirror segments, this means a new holographic plate must be produced for each family of segment shape, which adds to the cost of the technique. Furthermore it is often the case in the production and testing of surfaces like these that the size of the element needs to be taken into consideration. Either one regular sized holographic plate needs to be created for the whole optic; meaning the test beam must be expanded requiring a high degree of accuracy and detail to be

etched into the plate or a very large hologram must be created, both of these can be expensive and time consuming to calculate and produce. It would be vastly preferable therefore if it were possible to use a reconfigurable hologram which could be combined with stitching techniques to measure a large surface using a series of stitched sub-aperture measurements. Spatial light modulators using liquid crystals are a prime candidate for this application as they allow for an in-situ reconfiguration of the displayed hologram. Whilst due to their pixellated design and nature as an emerging technology they currently cannot match holographic plates for detail resolution, nevertheless they save on cost and production speed significantly due to their ability to be modified as the situation demands.

Replacing the reference mirror of a Twyman-Green or Mach-Zhender with a liquid crystal device has been shown to allow the accurate testing of optical surfaces and transmissives [70]. A further advantage of using LC SLMs in interferometric systems is that it removes the need for piezoelectric translators for the implementation of phase shifting methods. It is straightforward to include the phase shifted quantity directly into the design of a series of holographic displays and then display them in sequence within the system.

As previously mentioned it is still possible to create an analogue phase profile from a binary device [1]. Consider desired wavefront $\psi(x,y) = \exp[i(\phi(x,y) + \tau(x,y))]$ where $\phi(x,y)$ is the desired wavefront shape and $\tau(x,y)$ is an applied tilt term. By binarizing this wave according to:

$$\Phi(x,y) = \begin{cases} \pi & \text{if } \cos(\phi(x,y) + \tau(x,y)) \geq 0 \\ 0 & \text{else,} \end{cases} \tag{3.39}$$

we obtain a wavefront with a square wave profile. Upon passing this wavefront through a lens we acquire a diffraction pattern corresponding to the well known

Fourier series of a square wave:

$$\begin{aligned}\Psi(x,y) =& \frac{2}{\pi}\Big\{ \exp[i(\phi+\tau)] + \exp[-i(\phi+\tau)] \\ & - \frac{1}{3}\exp[3i(\phi+\tau)] - \frac{1}{3}\exp[-3i(\phi+\tau)] \\ & + \frac{1}{5}\exp[5i(\phi+\tau)] + \frac{1}{5}\exp[-5i(\phi+\tau)]\dots\Big\}. \end{aligned} \tag{3.40}$$

By using a spatial filter to isolate the first diffracted order from this pattern and then re-imaging, we are able to produce an analogue wavefront with the desired phase profile. The contents of the following chapters will be results based on the application of this simple concept to a FLC device and it's ability to accurately produce the wavefronts designed. The tilt term seen in equation 3.40 splits the component parts of the square wave expansion in the Fourier plane, allowing for the isolation of the desired first order term. It is important to ensure that as few higher order terms as possible be allowed through the pinhole in order to eliminate any contaminating aberration in the resultant wavefront, as shall be explored later.

Chapter 4

Simulated reconfigurable binary holograms

4.1 Introduction

As mentioned in the previous chapter, it is possible for the production of analogue wavefronts from the display of a binary hologram through the principle that a square wave can be deconstructed into an infinite sum of alternating positive and negative sine terms. In physical optics this can be achieved through the focussing of a wavefront through a lens and then isolating the desired diffracted order in the Fourier plane with a spatial filter before re-imaging. The contents of this chapter shall include an overview of the device used to display the holographic pattern as well as the design of the interferometer used in order to obtain the desired wavefront. In order to understand the basic behaviour expected from the device a simulation was made to determine the degree to which binary holograms could be successfully converted to analogue phase ones. The results of this simulation will be given starting with the basic ability to produce

simple Zernike terms and then covering considerations that need to be made which could severely affect the production of wavefronts, such as spatial filter size and how the binary pattern is encoded onto the device.

4.2 Forth SXGA 3DM FLCOS SLM

The SLM used for testing in this thesis is a Forth SXGA 3DM Ferroelectric-Liquid-Crystal-On-Silicon Spatial-Light-Modulator (FLCOS SLM). The device consists of a 1280 by 1024 pixel layout with a pixel width of 13.633 μm and interpixel gap of 0.54 μm. Each pixel is intended to act as a quarter wave plate, and used in reflection mode so a net effect of a half wave plate is achieved. For an ideal linearly polarised wave incident parallel to the short edge of the SLM display the effect of the device would be to add a π phase shift in the pixels considered "on" and to leave the polarization state incident on the "off" pixels unaffected. In reality the switching angle of the device is not 45° as intended but rather 33.5° at optimum running temperature which is given to be 40°C. For running temperatures different from this value, the short edge of the display unit no longer corresponds to one of the switched LC angles so setting the system up to obtain the desired polarisation modulation is a time consuming activity and shall be discussed in the following chapter. In order to achieve pure phase modulation the output light must be analyzed with a linear polariser as described in figure 4.1 and described in the following section.

The device is controlled via a USB interface using Metrocon software provided with the SLM and the display is pre-programmed into a sequence of images known as repertoires. Each repertoire is a list of the desired display images, combined with timing information called a sequence, which contains instructions for the displaying of frames on the SLM. This is necessary as due to the FLC nature of the device, for every application of an electric field across

the device necessary to orient the crystals, an equal and opposite magnitude field must also be applied in order to prevent electrochemical degradation of the liquid crystals. The maximum time an image can be displayed before the negative has to be shown is given to be 20 ms, however for the purposes of this thesis significantly smaller display times than this shall be used. The device allows for the display of images as rapidly as 100 μs however this is not the limiting speed of the device as there must be a 100 μs display of the negative plus a 369 μs warm up and warm down time on each sequence. Wave generators are used to synchronise the display of these positive frames with the integration time of the camera to ensure that no negative consequences of recording two different holograms appears in the data analysis.

4.3 Experimental interferometer design

As the response of the SLM is dependant on the polarization of the incident wavefront, that must be the first consideration in experimental design. A 633 nm HeNe laser is expanded using a spatial filter attached to the front of the laser and re-collimated with a f = 120 mm lens into a 36 mm diameter beam. Care must be taken to align the spatial filter so that as much beam uniformity is attained as possible in order to reduce limitations on the dynamical range of the camera later in the experiment. Crucially this light must be polarized with a known orientation before it becomes incident on the SLM. The reason for this can be seen in figure 4.1, which shows how in order to obtain pure phase modulation the wave reflected from the SLM must be passed through an analyzer oriented orthogonal to the incident polarization, so that equal amplitudes of S and P polarized light are passed. Any variation from this passes amplitude modulation in the wavefront through the interferometer which shows up on the final fringe pattern as an image of the diffraction grating from the SLM.

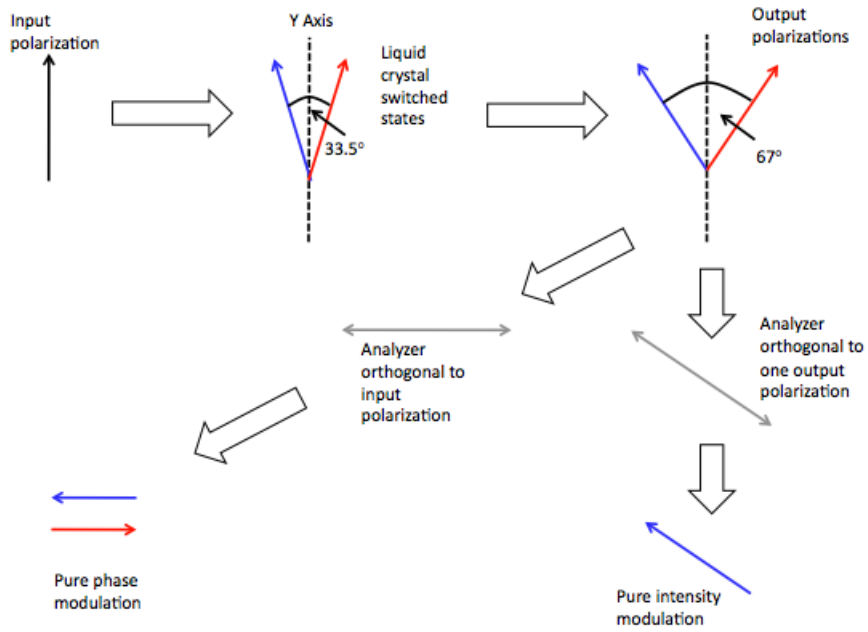


Figure 4.1: Basic principle of operation for using the FLC SLM. By observing the SLM with an analyzer oriented perpendicular to the incident polarization, light from both switched states is passed equally giving purely phase modulation in the resulting beam. By orienting the analyzer perpendicular to twice the switching angle of the device however it is possible to entirely block the contribution from one state giving intensity modulation in the beam.

A Mach-Zehnder interferometer configuration is used in the experiment, see figure 4.2, as it provides a high degree of flexibility with placement of optics throughout, and easily allows the arm containing the SLM to include the analyzer and reimaging optics required to produce the desired wavefront. Using the Mach-Zehnder configuration also gives the interferometer the capability to easily include both reflective and transmissive optics into the system when using the SLM to create holograms corresponding to aberrations seen in measurements. Furthermore, as seen to be required in later chapters, the large distance of path separation allows for the inclusion of motorised translation stages in single pass

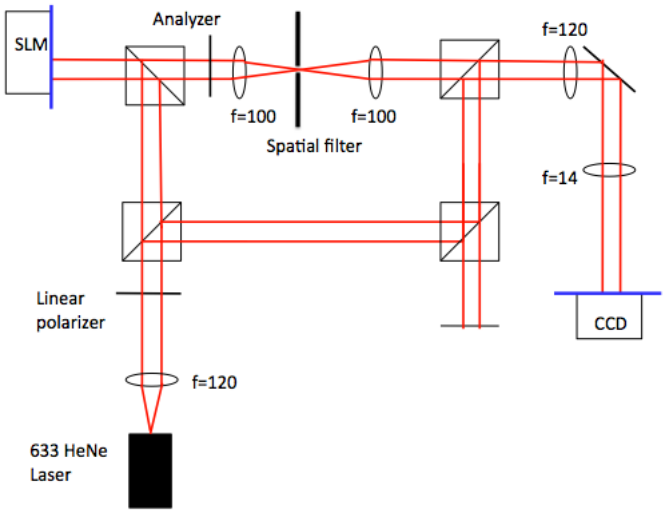


Figure 4.2: The basic layout of the experimental design using a Mach-Zehnder interferometer. The two blue lines denote conjugate planes. In this chapter, as the SLM is the object being evaluated, the arm containing it shall be referred to as the test arm, and correspondingly the other shall be known as the reference arm.

mode without mounting mechanisms infringing upon the optical path of the other arm. Unlike many instances of phase shifting interferometry no separate phase shifting components are included in the experimental design, as the phase increments are directly incorporated into the design of the holograms. In later experiments the addition of a half wave plate and polarizer in the reference arm was made, to allow for control over the intensity and polarisation state in order to obtain higher quality fringe modulation.

Ensuring the hologram is positioned in the conjugate plane to the camera is achieved via two methods. The camera is placed on a translation stage which allows for movement in the direction of beam propagation, and initially placed at a rough estimate of the conjugated position. This initial position is found by inserting a knife edge into the beam, the image of the edge of the knife

will be clear and sharp when the knife is placed at the conjugate plane to the hologram and blurry elsewhere allowing a rough gauge of the ideal point to place the camera. The effect of incorrect conjugation would be the observation of deviations in the shape of the wavefront from circular in the presence of aberrations, as a circular hologram is used. To find the ideal position to place the camera a large aberration, usually 9 waves of Zernike astigmatism Z(2,2) is put on the SLM, although the degree of accuracy in aberration production at this stage is mostly irrelevant as all that is required is the presence of a large Zernike term, and the camera is moved through various positions until the resultant image recorded is circular. This is checked by using DS9 software to impose a circular mask over the intensity pattern and iteratively comparing multiple positions of the camera until the closest match is achieved.

4.4 Simulation and simulated results

In order to obtain an initial indication for the ability of the SLM to produce analogue phase profiles from binary holograms a simulation was written in Python to propagate a hologram through the system used in the experiment. A circular hologram with a diameter of 800 pixels was modelled which matched the size of the hologram used during the majority of the experimental procedure. This size of hologram would not fully fill the SLM display but was used as larger holograms proved highly problematic to process in simulation due to computational restrictions. A Zernike polynomial was calculated and added to the 60 waves of Zernike tip and tilt used to split the individual diffracted orders in the Fourier plane, before being split into three different frames with 0, $2\pi/3$ and $4\pi/3$ respective phase shifts added to it in order to use three step interferometry

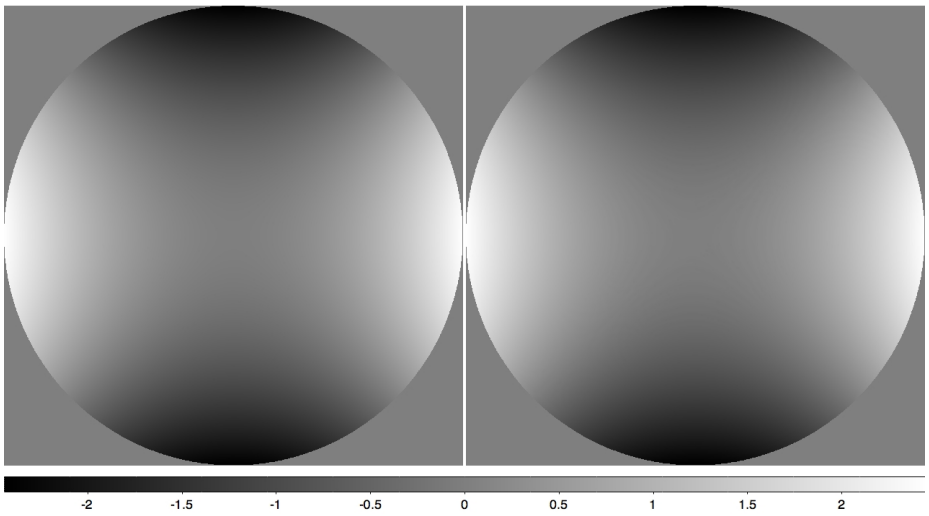


Figure 4.3: A comparison between the unwrapped phase profiles of 1 wave of Zernike astigmatism (Z5 or Z(2,2)) on left, and on right is the phase map obtained through using a binary hologram of 1 wave astigmatism and spatially filtering the first order diffracted term. Z scale on image corresponds to the wavefront phase value given in radians.

to process the result. This phase pattern was then binarized according to

$$\Phi(x,y) = \begin{cases} \pi \text{ if} \cos(\phi) \geq 0, \\ 0 \text{ else.} \end{cases} \tag{4.1}$$

where Φ is the phase of the hologram and ϕ is the phase of the pre-binarized phase map for the specific hologram, in order to create a binary hologram. A fast Fourier transform was applied to the final binary hologram, and a mask of size $R = 150$ pixels was imposed on that result to spatially filter only the first order wavefront before being inverse Fourier transformed to recover the image. Throughout, the size of the spatial filter will be referred to in units of pixels so the size in the Fourier plane can be compared to the size of the hologram. A three step phase shifting algorithm was used to process this result before being phase unwrapped to obtain a simulated estimate of the analogue hologram which

would be produced. Figure 4.3 shows the comparison between desired wavefront shape placed on the hologram and that obtained from the simulation showing an astigmatism component of 0.9998 waves RMS and overall wavefront RMS of 0.9999 waves.

First simulation results were obtained using three step interferometry, however it is worth considering another method for comparison purposes. The simulation ran using four step instead of three step also gave a result of 0.99809 waves of astigmatism, with an overall RMS of 0.99989 waves for a hologram corresponding to 1 wave RMS Zernike astigmatism. Three step interferometry was chosen to be used as the analysis method for the simulation predominantly down to eliminating the effects of higher aliased orders. If higher order diffraction patterns are allowed to pass through the system they will contaminate the interferogram however certain phase shifting algorithms allow for the elimination of some of these. By considering the Fourier expansion of a square wave, we see that all diffracted orders are odd so the primary advantage of using a four step method is eliminated, namely the elimination of even numbered higher order harmonics. Three step however allows for the removal of all orders which are a multiple of three, as the phase step $(3N)2\pi/3$ is identical for every frame and thus gives no influence over the recovered phase distribution.

Simulated results were also obtained for the propagation of Zernike Trefoil (Z(3,3,1) or Z9) showing a result of 0.99794 waves RMS trefoil from an intended 1 wave, however an overall wavefront RMS of 0.99985 waves is seen, for a comparison see figure 4.4. For the propagation of a combination of Z(2,2,1) and Z(3,3,1) a result of 0.99984 waves RMS astigmatism and 0.99984 waves RMS trefoil is obtained with an overall wavefront RMS of 1.41423 waves compared to the expected wavefront RMS of 1.4141 waves, with a visual comparison being given in figure 4.5.

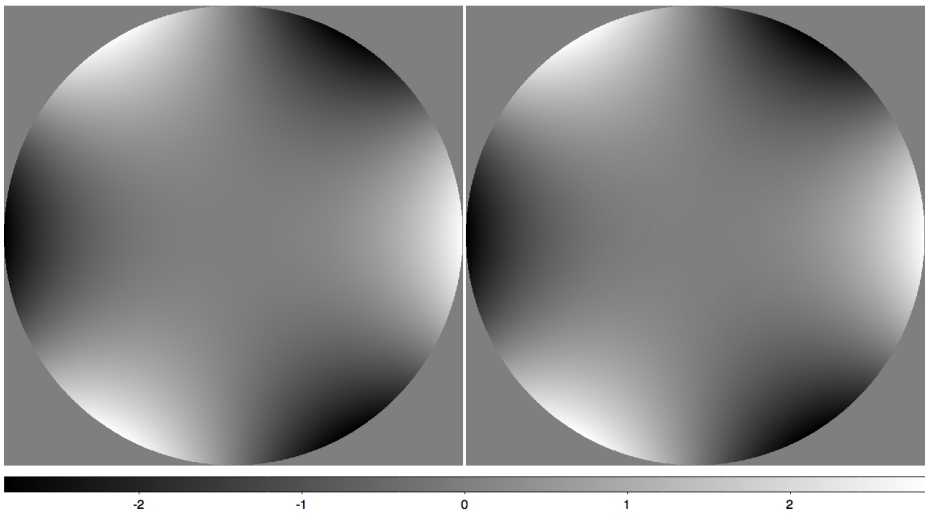


Figure 4.4: A comparison of the perfect and simulated wavefronts for 1 wave RMS Zernike Trefoil, with Zscale in radians.

n	1	2	3	4	5	.	.	.	9
Measured	0.9998	1.9996	2.9994	3.9992	4.9991	.	.	.	8.8340
Tot RMS	0.9999	1.9998	2.9996	3.9995	4.9994	.	.	.	8.8480

Table 4.1: The simulated ability of the SLM to produce of n waves of Zernike astigmatism for a fixed spatial filter size along with the total RMS of the produced wavefront.

Table 4.1 shows the simulated growth of the produced astigmatism terms along with the overall RMS of the final wavefront up to 9 waves of applied aberration. Up until this value the measured quantity of desired aberration follows a steady growth until the applied aberration approaches 9 waves RMS where the PSF becomes significantly larger than the spatial filter being used cutting off the first order and leading to a reduction in measured aberration. This suggests that a high degree of precision in controlling the desired aberration is possible through this technique.

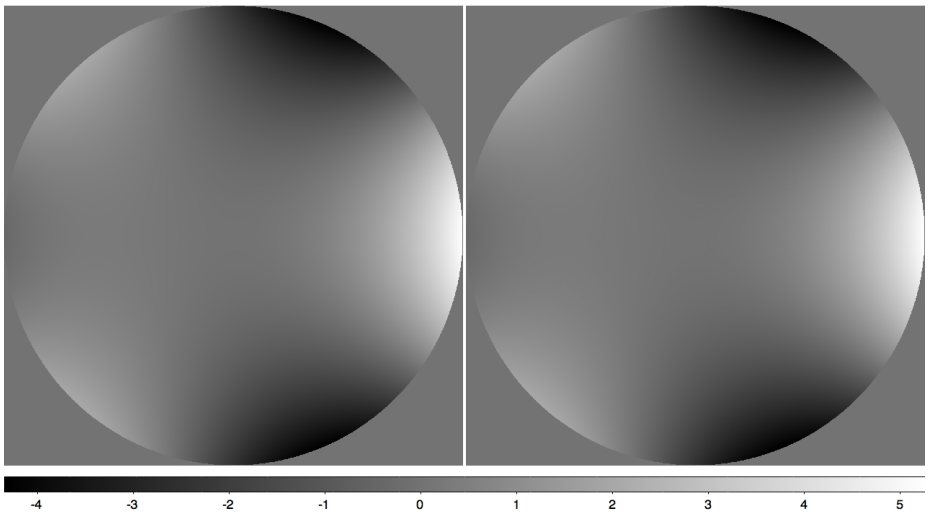
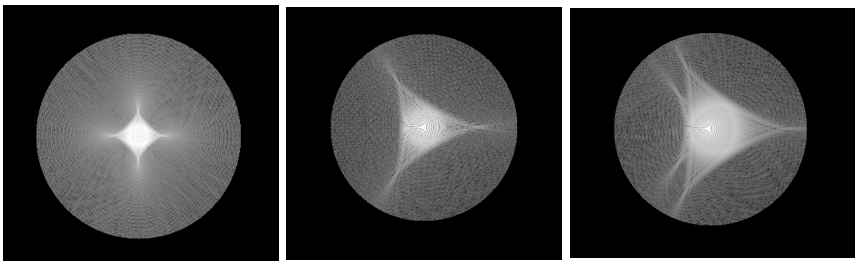


Figure 4.5: Comparison of ideal and simulated wavefronts for a combination of 1 wave astigmatism and 1 wave trefoil, again with scale in radians.

4.5 Varying spatial filter size

It is apparent from the result in table 4.1 that the size of the spatial filter used to isolate the desired diffraction order must be considered as it can affect the accuracy of the desired wavefront. Too large a filter will allow through higher order harmonics and contaminate the beam whereas too small a filter will cut off the PSF and therefore not allow the full propagation of the beam. Furthermore it is seen that larger aberrations will require a larger spatial filter than smaller ones, and certain modes with more spatially extended PSFs will also. Regardless of the magnitude of the desired aberration however, the locations of the PSF for each harmonic will remain fixed as their positions are determined by the constant 60 waves tip and tilt applied with the intent of separating the orders. Figure 4.6 shows the PSF along with spatial filter for 1 wave astigmatism, 1 wave trefoil and the sum of both. It is apparent that the PSF for trefoil is significantly more spatially extended than for astigmatism alone, and even more so for the sum of the two. The spatial filter size therefore needs to be considered for each



(a) (b) (c)

Figure 4.6: A comparison of the spatial filter and spatial extent of the PSF in the Fourier plane for an aberration of a) 1 wave astigmatism, b) 1 wave trefoil and c) 1 wave astigmatism + 1 wave trefoil.

aberration type being produced, especially for the case where multiple Zernike aberrations are being combined.

Figure 4.7 shows the growth of the recovered Z5 astigmatism component for a hologram of 5 waves RMS displayed on the SLM along with the total RMS of the produced wavefront. For a small spatial filter size, insufficient wavefront is passed through the system to produce the intended quantity of aberration until the spatial filter reaches approximately 100 pixels in size at which time the magnitude of the recovered aberration appears to plateau along with the RMS of the system. Further observation however shows that as the spatial filter continues to grow, more unwanted light is passed through the system resulting in growing artifacts in the recovered phase profile. These manifest themselves as jagged fringe edges in the observed interferogram as seen in figure 4.8. As one of the main purposes for using the SLM to create arbitrary phase profiles is use in nulling interferometry it is desirable that these artifacts be kept to a minimum as in a successful null test they would dominate the residual RMS of the wavefront.

It is not enough to merely consider the size of the aberration that the SLM is able to produce however. As mentioned above, one of the driving reasons for the consideration of such a device in wavefront shaping is the use of it in

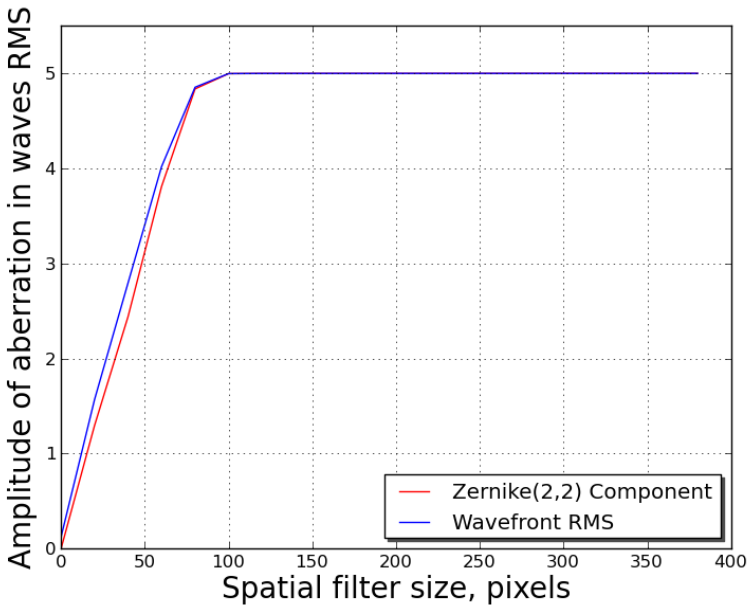


Figure 4.7: Growth of recovered Z5 astigmatism component for a hologram of 5 waves RMS Z5 as a function of increasing spatial filter size. Until R=100 pixels the filter is too small to pass the wavefront fully through the system and above this value the recovered aberration quantity plateaus as errors caused by the contamination of higher orders passing through the filter are too small to dominate the RMS in the presence of the significantly larger astigmatism term.

nulling interferometers. In such a case the created aberrations will cancel out existing aberrations to create an ideally flat interference pattern, however any small scale irregularities in wavefront production will show clearly and could potentially cause significant error by distorting this null pattern. Figure 4.9 shows the growth of the residual RMS error of the simulated wavefront with increasing spatial filter size where the SLM is displaying 5 waves astigmatism and then used to null out the same amount of aberration in the other arm of the interferometer. The high RMS seen at a filter size of under 100 pixels is due to insufficient PSF being passed through the system, however this residual drops sharply to roughly $\lambda/100$ once the PSF is entirely transmitted. Above this filter

size however a gradual increase in the residual RMS of the system is seen as extra light makes it through the increasingly large pinhole. This light serves no purpose in the reconstruction of the desired wavefront and merely contaminates it with higher order structure which during the null testing procedure dominates the fringe pattern. It should be noted that there is no flat bottom to figure 4.9 which tells us that there is a unique optimum spatial filter size for the specific aberration being considered, rather than a range of values which provide a minimum of higher order contamination. We expect that eventually the need for a large spatial filter for increasingly aspheric wavefronts will limit the possible production of highly aberrated profiles as we will begin to pass light from the zeroth order term. By increasing the carrier tilt we can further separate the distance between each diffracted order, however this then may give rise to insufficient sampling of the wavefront on the hologram display.

Figure 4.10 shows the null interference of 5 waves astigmatism with the wavefront produced from a hologram of the same aberration created with a spatial filter of 300 pixels in size, i.e. significantly larger than needed and giving a high residual RMS as seen in figure 4.9. It is clear from this that there is a large astigmatic artefact present in the image which dominates the phase profile. This contamination would mask small scale phase variations in a real experimental null test in the area of the artefact, an effect which can be minimised by reducing the spatial filter size. Notably it would appear this artefact can only be minimised through spatial filter variations , and not eliminated entirely.

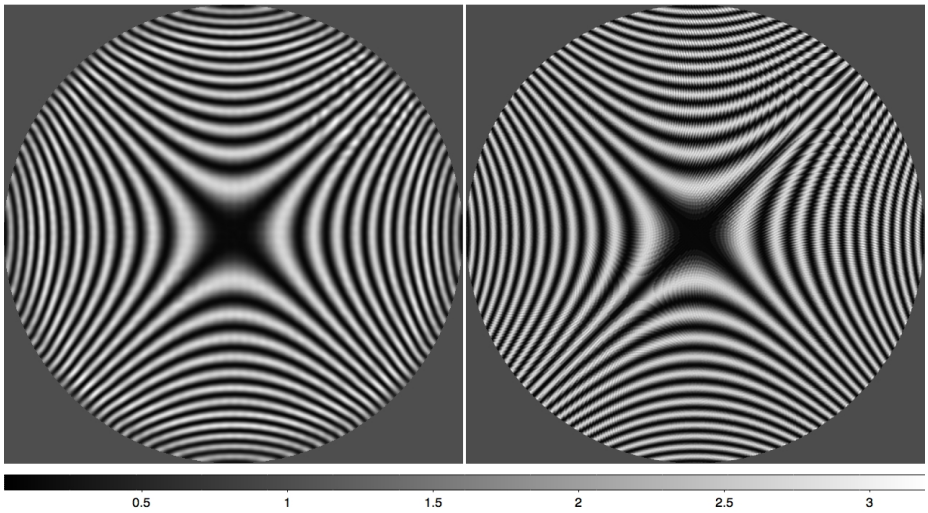


Figure 4.8: Comparison of interferograms for 5 waves of Z5 astigmatism for two different spatial filter sizes. Left: Spatial filter radius=100 pixels, the first value for which the recovered wavefront approaches 5 waves. Right: Spatial filter R=380 which is beyond sufficient for full passage of the wavefront.

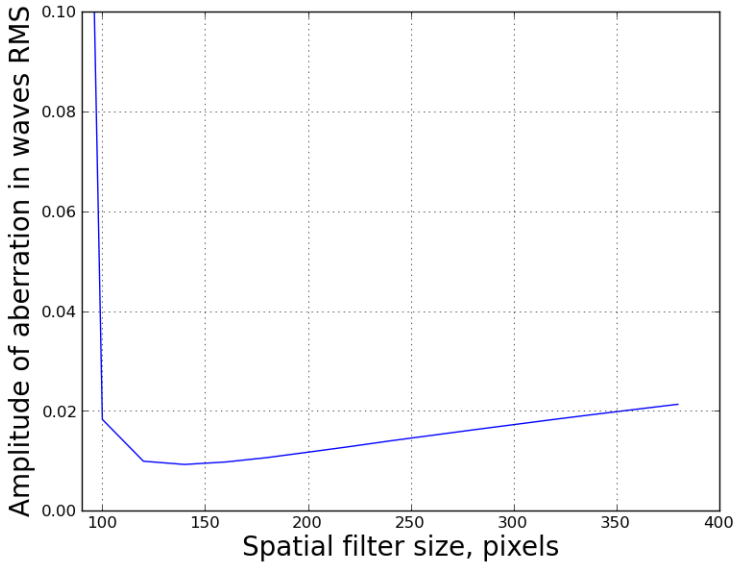


Figure 4.9: Growth of the residual error in recovered wavefront after a direct subtraction of 5 waves Z5 astigmatism for a varying spatial filter size.

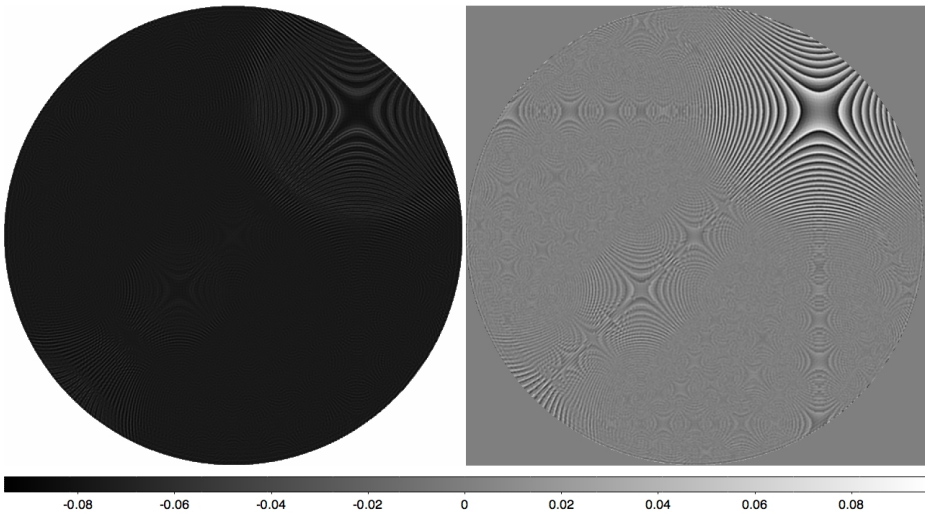


Figure 4.10: Left: Interferogram of a null test between a perfect 5 waves of Zernike astigmatism and the wavefront produced by displaying 5 waves on the SLM and using an F=300 pixel spatial filter. Right: Unwrapped phase profile of interferogram on left showing a residual RMS of 0.017 waves. Scale on key in fractions of a wave.

4.6 Residual growth with increasing aberration size

As previously mentioned, increasing aberration size results in an increasingly large PSF in the Fourier plane of the lens and therefore requires a larger spatial filter to fully pass the aberration than a smaller aberration would. Experimentally however it may not be straightforward or practical to continually calculate and accurately change the spatial filter size for every new aberration displayed on a hologram. Because of this it is worth considering how using a fixed sized spatial filter will affect the accuracy of the produced wavefront for increasing sizes of aberration placed on the hologram. Figure 4.11 shows how the magnitude of the desired Zernike term will grow with increasing hologram aberration size for a fixed spatial filter width of 150 pixels. For small aberration sizes the

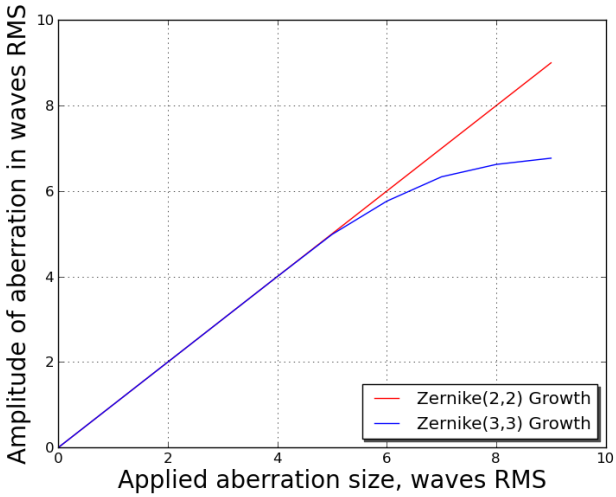


Figure 4.11: This graph shows the increased in produced Zernike term when a hologram of an increasingly large corresponding term is shown on the SLM for a fixed spatial filter size F=150 pixels. Astigmatism continues to be produced to a high degree of accuracy as seen already in table 4.1 up until an applied aberration of 9 waves RMS, however the produced trefoil term drops off sharply well before this.

production of both astigmatism and trefoil is as desired, n waves applied gives n waves of aberration in the resulting phase pattern, however as seen in table 4.1 this begins to drop off when the applied astigmatism component is at 9 waves RMS. As would be expected, the production of trefoil begins to falter at smaller applied aberrations than astigmatism, which is a direct result of having a more spatially extended PSF. For a spatial filter twice the size of this however both aberrations are produced highly accurately, to the extent where a graphical representation like figure 4.11 would be redundant as both data sets would overlap on a near perfect straight line and further analysis would be impossible.

Instead of looking at the production of the desired Zernike term to evaluate the accuracy of producing increasingly large aberrations, consider the residual RMS of the null interferogram produced. Figure 4.12 shows the increase in

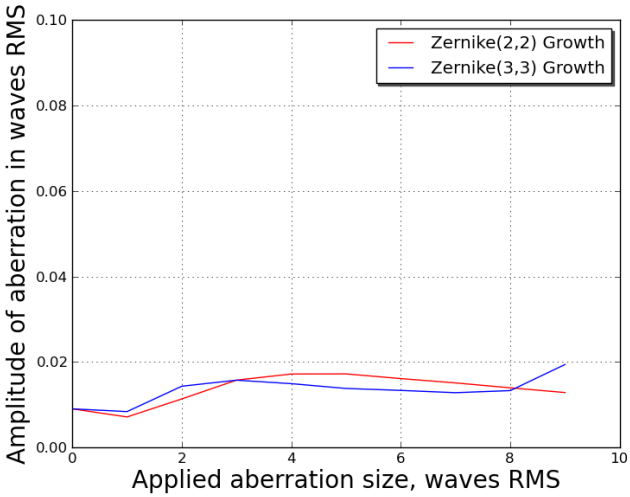


Figure 4.12: The growth of the residual RMS wavefront error with increasing magnitude of astigmatism and trefoil present in the wavefront for a fixed spatial filter size of 300 pixels.

residual RMS of the recovered phase map produced from null interferograms of increasingly large astigmatism and trefoil for a fixed F=300 pixel spatial filter. Considering the growth of astigmatism we see that for small sized aberrations the effect of higher order contaminating terms grows along with the desired aberration, in this case the spatial filter is significantly more extended than needed to pass through the whole PSF of the aberration. Comparison of the residual for 5 waves applied astigmatism can be made with figure 4.7 where we see that F=300 is significantly larger than optimal for the passing of the full aberration and for this value there are significant errors introduced due to higher order terms. Above this value the residual RMS drops which can be associated with a reduction in the severity of higher order terms being passed through, as the optimum spatial filter size becomes closer to F=300. Trefoil follows a similar pattern with an expected reduction in residual RMS terms being observed for smaller magnitude aberrations, again associated with the increased size of the

PSF with respect to the astigmatic one. At 9 waves RMS applied trefoil an increase in the residual RMS of the wavefront is observed due to the size of the filter becoming insufficient to fully pass and propagate the desired wavefront shape.

It is also worth considering the effect increasing aberration size has on the magnitude of other produced Zernike terms. It would clearly pose a problem if during the production of a specified Zernike aberration other undesired terms arose that contaminate the beam. Figure 4.13 shows a surface map of the magnitude of the first 20 Zernike terms and their growth with increasing amounts of Zernike astigmatism being displayed on the hologram. Using a spatial filter size of R=150 pixels which we already know gives highly accurate production of the astigmatic term in the wavefront up to a value of 9 waves RMS, we can see that the predicted size of the other Zernike terms remains generally minimal. When the spatial filter stymies the production of the largest aberration simulated there is also an increase in the Z13 term, however this vanishes when the spatial filter size is further increased. This indicates that the technique used to create arbitrary wavefront shape is able to preserve the orthogonality property of Zernike polynomials in the presence of a sufficiently large spatial filter.

4.7 Optimal binarization algorithm

All of the above results have been considered using the binarization method from equation 4.1. This gives a standard square wave pattern which in the Fourier plane consists of very structured regular higher order terms. When these terms are partially or wholly passed through the spatial filter they cause the presence of similarly well structured residual aberrations which greatly contaminate the null interferogram. Depending on the spatial filter size these can reduce the residual RMS of the null interferogram to worse than $\lambda/50$ RMS. Whilst this

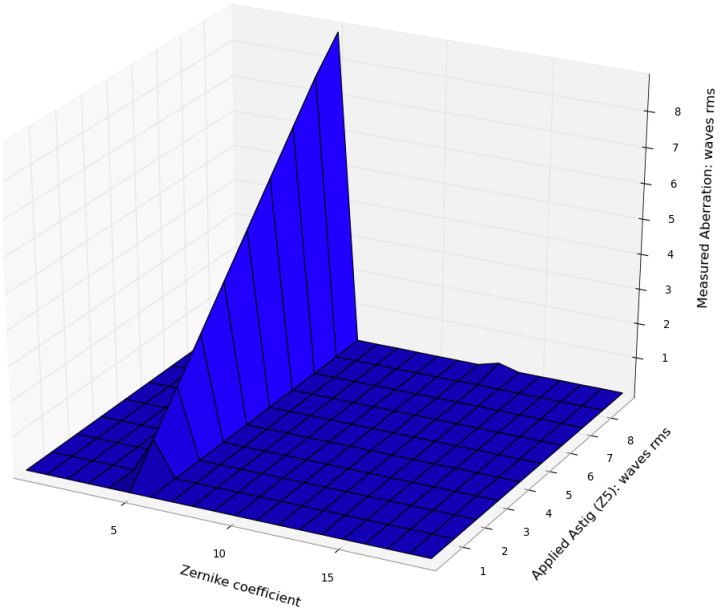


Figure 4.13: Surface plot of the size of the first 20 Zernike terms and how they grow with increasing amount of astigmatism being displayed on the hologram using a spatial filter of F=150.

is not an especially poor level of accuracy it is worth looking into whether or not this effect can be reduced so that spatial filter size considerations can be somewhat more relaxed.

An alternate method of programming the hologram is to use a random binarization algorithm,

$$\Phi(x,y) = \begin{cases} \pi \text{ if} \cos(\phi) \geq \text{Random}-1:1, \\ 0 \text{ else.} \end{cases} \tag{4.2}$$

The effect on the hologram that this binarization algorithm has can be seen in figure 4.14. Completely absent is the presence of the neat diffraction grating seen in a standard binarization however this pattern still results in a first order diffracted term corresponding to the desired aberration. Where the standard

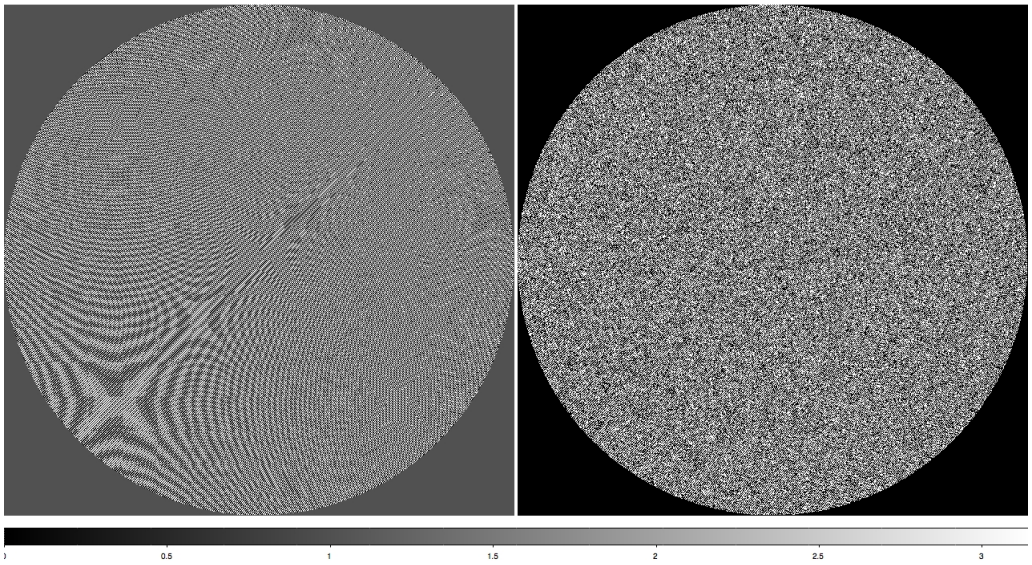


Figure 4.14: Two holograms of the same aberration binarized using the standard normal method given in equation 4.1 on left, and using a random binarization given in equation 4.2

binarization technique gives a series of well structured higher order diffracted terms however the random technique does not. Instead it introduces a degree of background noise into the PSF which is devoid of regular structure so that passing this through a filter will introduce a random error component rather than structured aberrations. As the dominant source of error in the null interferograms simulated above was due to contamination by structured harmonics it is worth quantifying the accuracy of a technique of producing holograms which may minimise the presence of these artifacts.

Figure 4.15 shows how the growth of the spatial filter size affects the production of 5 waves of Zernike astigmatism. Included in this is an intermediary binarization technique denoted as quasi-random where for each pixel, during the phase comparison step of the binarization it is either compared against a value of 0 as in equation 4.1 or a random value as per equation 4.2. The determination of this is once again decided using a random number generation function. This

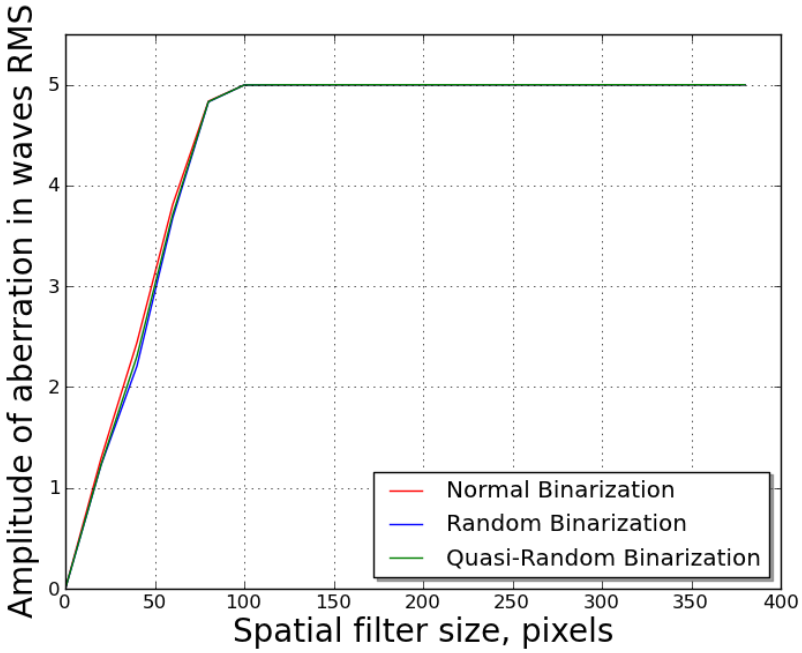


Figure 4.15: The growth of 5 waves Zernike Astigmatism with increasing spatial filter size for three different binarization techniques.

quasi-random technique was investigated as it produces holograms that contain a higher degree of structure than a fully random technique and therefore lessens the degree of background noise however also disrupts the structure of higher order harmonics. Figure 4.15 shows that the growth of the astigmatic term is slightly larger for a standard binarization technique, however all three methods plateau at a similar spatial filter size.

As mentioned previously however it is more important to consider the effect of the residual errors in hologram production in the case of null interferograms. Figure 4.16 shows the growth of the residual RMS wavefront error with increasing spatial filter size in the case of a null interferogram, with the aberration being nulled again being 5 waves Zernike astigmatism similar to figure 4.9. For all values of spatial filter size the effect of higher order structured terms pro-

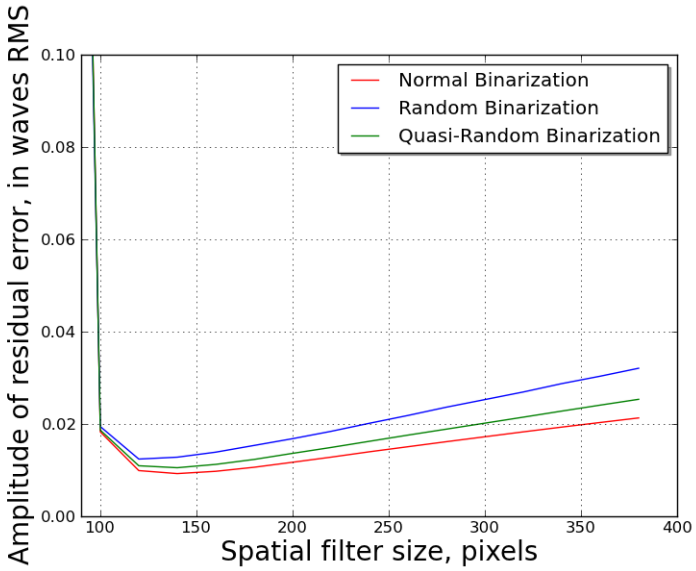


Figure 4.16: The growth in residual error with increasing spatial filter size of a null interferogram involving the production of 5 waves Zernike astigmatism for three binarization methods.

duces a smaller error than the random background noise introduced through randomization. We also see from this figure that there is a slight increase in the growth of the randomised residual RMS as a function of increasing spatial filter size compared to the standard technique. In the case of the latter, a larger spatial filter passes through a greater amount of the higher structured harmonics, however for the randomized technique this increase will be a result of a larger spatial filter allowing a greater amount of randomised background noise through the system. It can be inferred from this that the errors induced from introducing a degree of random background noise causes a larger error growth than the allowing of structured orders to partially or fully pass does.

Figure 4.17 shows the growth of the null interferogram residual with increasing aberration size for a fixed spatial filter of 300 pixels, which corresponds to the results seen in figure 4.12. In the latter we see that increasing aberration

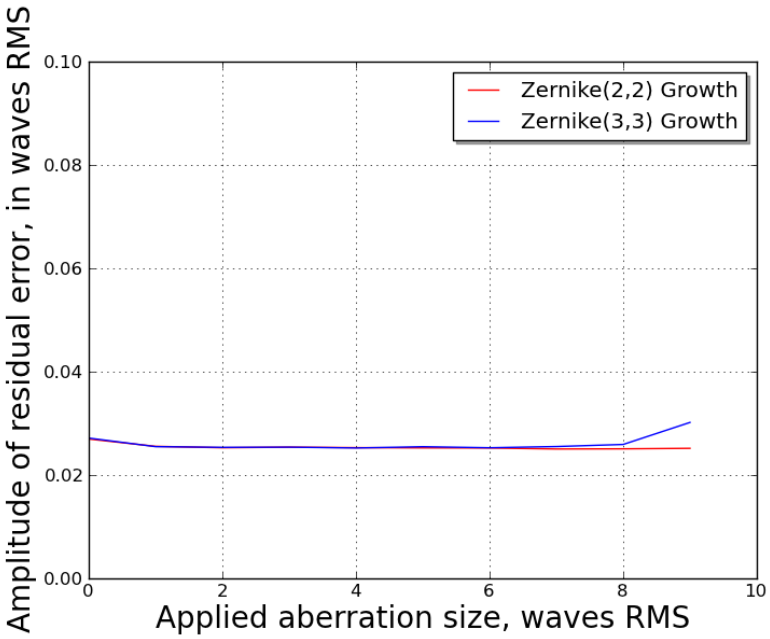


Figure 4.17: Growth of residual RMS error in null interferograms for the random binarization technique with increasingly large aberrations on the SLM with a fixed spatial filter size of 300 pixels.

size causes variations in the residual RMS measurement which is understood to be an effect of varying amounts of higher order harmonics being passed through the filter contaminating the image. In the case of the random binarization technique we see that the there is a roughly constant level of residual error for all but the largest sized aberrations simulated, at which point an increase is observed for the production of trefoil at the point where the PSF is no longer being fully transmitted through the Fourier plane, something also seen in figure 4.12. The lack of increase in these residuals with increasing aberration is again ascribed to the lack of higher order contaminants being passed through the filter.

Looking at the nulled phase maps for the fully random binarization technique compared to standard, figure 4.18, helps to confirm that the erroneous aberrations we see in the case of figure 4.10 are a result of the presence of higher

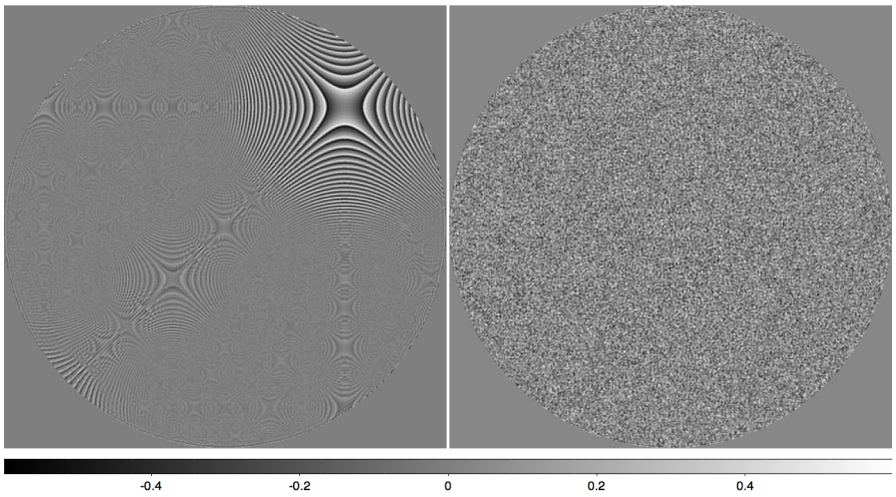


Figure 4.18: Residual phase maps recovered from a simulated nulled interferogram of 5 waves astigmatism with a spatial filter size of F=300 for Left) standard binarization and Right) random binarization. Whilst the highly structured elements are eliminated by using a randomised technique the background noise introduced fluctuates on a similar scale to the height of the well defined astigmatic artifacts.

order structure transmitted through the spatial filter as they do not appear here when this structure is disrupted. It would appear from these results that using a non-random binarization technique to encode the hologram is preferable however and should be used experimentally. Initial results found experimentally should be obtained without serious consideration to the spatial filter size, as residual errors even for the largest spatial filter sizes simulated are still in the area of 0.05 waves RMS and provisionally acceptable. After an experimental indication of SLM accuracy is obtained the effect of spatial filter size can be considered when attempting to reduce errors in the experiment.

4.8 Conclusions

We have initially seen through a simulation that the basic premise of creating an arbitrary wavefront shape from a binary hologram has promise, with deviations from the desired wavefront shape being less than one hundredth of a wave providing that the entire PSF is propagated through the Fourier plane. Whilst there is little difference between the results obtained through using three- or four-step phase shifting techniques the decision to use the former was made primarily because of the increased number of higher order harmonics it eliminates as these may contribute towards residual wavefront error. We saw that the effect of passing these terms through the system in simulation is to give the fringes a jagged appearance and after phase shifting analysis leaves highly structured features corresponding to the aberration being created. These features become more prominent as the size of the spatial filter increases, however both larger magnitude aberrations and higher order Zernike terms have more extended PSFs and so there must exist an optimum size filter for each possible hologram. These higher order harmonics are the result of using a highly structured hologram and so the possibility of introducing a random component into the binarization of the hologram was investigated, and whilst this successfully eliminated the presence of unwanted higher structure in the residual phase pattern the inclusion of random background noise leads to a higher residual RMS error than seen in the case of a standard binarization and so is deemed less suitable. It should be noted that using multiple randomised images per subframe may be worth investigating as this could lead to a reduction of the residual random error, however with only 1 image per phase step the standard binarization technique is preferred as it leads to a smaller overall residual error.

Chapter 5

Experimental characterisation

5.1 Introduction

This chapter shall present the results of the experimental quantification of the ability of the FLCOS device to produce arbitrary wavefronts, which were simulated in the previous chapter. Firstly a description of some of the specific experimental details shall be given, followed by the need for polarisation control over the beam due to the nature of the FLCOS device. The first experimental results on the ability of the device to produce Zernike astigmatism are described, along with a comparison of the binarization techniques described in the previous chapter. Following this, a discussion on the nature of the residual errors seen in the recovered nulled phase maps is given as well as a comparison of phase unwrapping techniques and their associated shortcomings.

5.2 Interferometer set-up

As described in the previous chapter, a Mach-Zehnder interferometer was chosen for use in the evaluation of the SLMs ability to produce custom wavefronts. The two separate path arms in this configuration allow for the easy insertion of the imaging system needed to isolate the first diffracted order produced from the hologram, as well as the use of an analyzer in single pass to reduce the beam attenuation. The re-imaging lens after the spatial filter was placed on a translation stage to ease collimation, which was obtained using a combination of shear plate analysis and software image analysis to ensure the beam output from the imaging system was the same size and shape as the beam recorded without it. As mentioned in the previous chapter, DS9 software was used to fit a circular mask over recorded hologram images to ensure their correct conjugation, with the parameters of this mask being recorded and used to characterise the area over which data processing would be done. Furthermore when aligning the two arms of the interferometer to give a fringe pattern, any differences in collimation would be observed to give a defocus interferogram and therefore an absence of this would suggest both beams were collimated identically. A 633nm HeNe laser was used with a detected output of 2mW CW as an illumination source for the experiment, and it was necessary to allow sufficient time for the laser cavity to stabilise otherwise the observed interferogram would contain a slow phase drift. Other inclusions into the optical design were a reference mirror on a translatable mount for added flexibility in conjugation upon the addition of any optics into the reference beam as well as a half wave plate in the reference arm to increase fringe contrast in recorded interferograms.

Also mentioned in the chapter 4 section 2 was the need for DC balancing of the SLM to prevent electrochemical degradation of the liquid crystals. This is achieved by displaying the negative of the desired image for an equal time

as the original after it has been displayed. To ensure that images were taken during the time where the hologram was displaying the desired grating pattern, as opposed to the inverse image required for balancing, two wave generators were used to synchronise the camera and the SLM. A 1KHz pulse was used to trigger the activation of the SLM display sequences which was also sent to the second wave generator which then waited a duration of 369 μs before sending a 100 μs square wave to a 1024x768 pixel Fire-I XGA Pro CCD camera defining the integration time. The timing of these pulses was taken from documentation provided by the manufacturers of the SLM and associated software and chosen as they corresponded to the shortest possible display time of the hologram on the SLM in an attempt to minimise the effects of any mechanical vibration present in the set up.

5.2.1 Polarization optimization

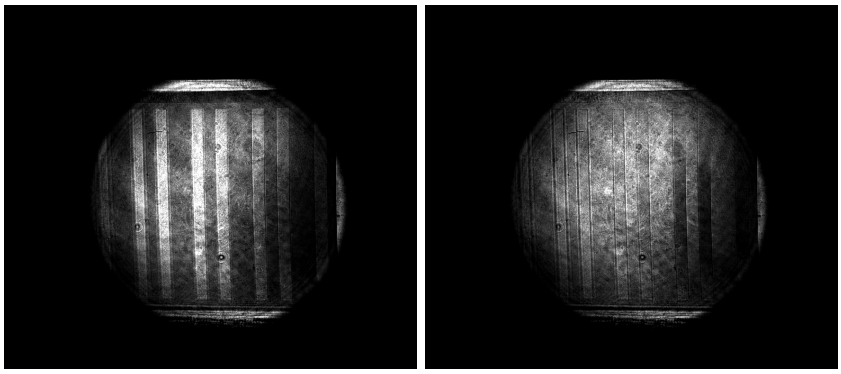
In the previous chapter it was mentioned that it was possible to obtain pure phase modulation from the SLM using an analyzer perpendicular to the polarization of linear light incident upon the SLM. In order to achieve this however it is necessary to illuminate the SLM with linearly polarized light aligned at an angle midway between the orientations of the two different liquid crystal axes. If the incident angle of polarization is not halfway between these two states then the resultant wavefront will contain a mixture of phase and amplitude modulation and can lead to artifacts in the recorded interferograms.

One method of finding the ideal polarization alignments for the incident light is to use the test mode programmed into the SLM which displays a series of parallel bars on the microdisplay. When directly imaging the SLM in this mode intensity modulation can be obtained by adjusting the analyzer until maximum contrast between the 'on' and 'off' bars is observed. Theoretically we know that

in order to get pure phase modulation, the analyzer and the input polarisation need be orthogonal to each other, and therefore by rotating both elements equally we should find that the intensity modulation maximises when the input polarisation is aligned with one of the liquid crystal switched states. To find the ideal phase modulation configuration therefore we just have to rotate both elements 33.5°, the switching angle of the LC, from that orientation to obtain the desired wavefront state. The value of 33.5° is given from the manufacturers of the SLM as the switching angle between the two liquid crystal states when the device is running at the optimum temperature of 40°C. In reality the ideal configuration for pure phase modulation gives too little light through the system to reliably record results with the experimental laser used, causing regular errors in phase unwrapping. This could be corrected with a more powerful source however and is not an inherent limitation to the technique. To compensate for this it is necessary to allow a certain degree of amplitude modulation through in order to achieve useable light levels. Figure 5.1 shows a comparison between the image of the SLM on the camera in the orientation which provides maximum intensity modulation between the 'on' and 'off' pixels, and an image of the SLM in the configuration that should give pure phase modulation only. As can be readily seen in the case of the latter there is residual amplitude modulation being transmitted so getting a series of interferograms which can be successfully unwrapped necessitates a trial and error approach to polarization optimization.

5.2.2 Fringe Stability

One of the major sources of error found commonly in interferometry is that induced by high frequency variations in the path lengths between the two arms. In the particular case of this interferometer the problem is of special concern due to the long separate path length for each arm. The two main sources for



(a) (b)

Figure 5.1: Images of a test pattern on the SLM for different orientations of the output analyser. A) shows the observed test pattern for when maximum contrast is seen corresponding to maximum intensity modulation and b) shows the modulation of the fringes when the analyser is rotated 33.5° away from this point. What is seen is not pure phase modulation as expected theoretically but rather contains a degree of intensity modulation.

these high frequency variations are mechanical vibrations and air turbulence.

There are numerous methods of reducing the effect of mechanical vibrations on the experimental accuracy, starting with mounting the whole set-up on a floating air bench to significantly attenuate the vibrations coming through from the floor from sources such as traffic. As the experiment was performed in a university building with high undergraduate activity this is especially important as the near constant vibrations from footsteps would make obtaining any results impossible. All electronics were removed from the bench to eliminate the effect of 'mains buzz' from adding vibrational sources. The two components which are the most sensitive to any vibrations, the SLM and the reference mirror, were mounted on heavy magnetic clamps with thick vibrational resistant posts to increase stability. Experimental results were obtained from two different interferometer builds, one in Durham and one built at the National Physical Laboratory(NPL) however results presented in this chapter were all obtained from the Durham interferometer. In the NPL build, Linos cage mounting was

used which further increases component stability as all optics are tightly bound together in a cage system which prevents independent oscillations from being induced in individual optics.

Air turbulence effects in the NPL interferometer were highly dominant in the absence of any compensating arrangements, so the experiment was enclosed in an Optoblok casing to minimise the entrance of air currents into the separate paths, with only the openings required for cabling and laser entrance to the system. As the laser heats up through use it was not encased in the Optoblok as it would act as a turbulence source by warming the air and inducing new currents inside the box.

The degree to which residual vibrational effects and air turbulence affect the repeatability of the experiment were determined through measuring the standard deviation of recovered values for multiple aberration sizes over repeated measurements. The increase of applied aberration on the hologram gave no consistent increase or decrease in the magnitude of the errors detected. Error bars on all graphs displayed below are taken from this analysis as are errors stated with given results.

5.3 Experimental results

The first experimentally obtained results were taken whilst displaying holograms with radius of 400 pixels. These holograms, which did not fully fill the 1024 x 1280 pixel display, were used to ensure the full extent of the pattern was illuminated by the source . Table 5.1 shows the growth of the recovered desired astigmatism term with increasingly aberrated holograms as well as the RMS error of the wavefront after subtraction of the desired aberration from the final phase map. We note that the residual errors are not merely the applied aberration minus the measured aberration as Zernike polynomials add to the RMS

Applied aberration size Waves RMS	0.5	1	1.5	2	2.5	3
Measured astigmatism Waves RMS	0.4764	0.9964	1.4934	1.9989	2.4945	2.9905
Residual error Waves RMS	0.2696	0.1611	0.4531	0.2732	0.4437	0.3523

Table 5.1: First experimental results showing the growth of recovered Zernike astigmatism with increasing aberration applied onto the SLM, with hologram radius r=400 pixels.

Applied aberration size Waves RMS	1	2	3	4	5
Measured astigmatism	0.9994	1.9976	2.9946	3.9923	4.9913
Residual error	0.1170	0.1209	0.1391	0.1582	0.1680
Applied aberration size	6	7	8	9	
Measured astigmatism	5.9912	6.9927	7.9940	8.9964	$\pm$0.00027955
Residual error	0.1931	0.2301	0.2497	0.2855	

Table 5.2: The production of Zernike astigmatism using a standard binarization technique up to an applied aberration of 9 waves RMS. Residual error is the error after subtraction of the ideal applied aberration size.

wavefront error in quadrature. The residual errors after removal via the subtraction of the desired astigmatism term seen for these holograms are a significant fraction of the recovered astigmatism term which may be directly related to the size of the hologram. Fringes on the binary hologram for this size are often only a single pixel in width which suggests the wavefront may be insufficiently sampled for accurate recreation which could be improved upon using a larger diameter hologram. Furthermore a larger hologram will diffract more light to the desired spatially filtered order which is important due to the low efficiency of the technique used.

For a larger hologram size of 1000 pixels across, using a standard binarization algorithm, the recovered astigmatism and error measurements are given in table 5.2. These are obtained using a modal deconstruction on the unwrapped phase map produced through three step interferometry, and are compared against

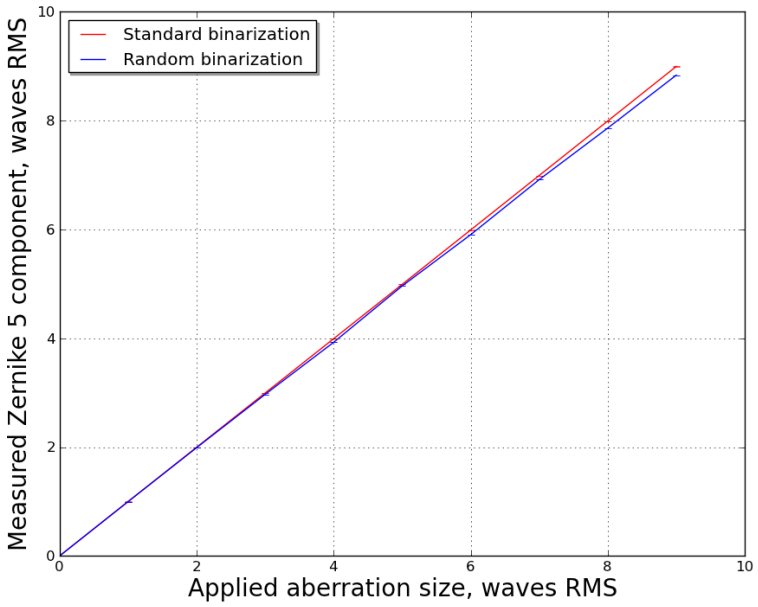


Figure 5.2: The growth of measured Zernike(2,2) astigmatism with increasing applied astigmatism on the hologram. The red and blue lines correspond to a standard and random binarization algorithm respectively.

similar results obtained using a hologram binarized via the random algorithm described in the previous chapter and shown in figure 5.2. This figure simply shows the results obtained from the hologram interfering with a flat mirror after systematic errors have been removed via direct subtraction of a phase distribution obtained in the absence of an astigmatic component to the hologram. This figure also experimentally confirms the simulated results of the previous chapter in which the randomized binarization algorithm was observed to be less suitable for the accurate production of Zernike modes than the standard one. It is also seen that the error induced through vibration and turbulence, as determined in chapter 5 section 2.2 and shown as error bars, is an almost negligible fraction of the recovered values. This suggests that the technique overall may be quite

insensitive to vibration so long as the basic precautions against it as outlined previously are taken.

As the dimensions of the SLM are 1080x1240 pixels this the close to the limiting size of the hologram capable of being produced, however it is desirable to use a hologram which does not fully fill the dimensions of the SLM as illumination of the edges can lead to spurious reflections contaminating the beam.

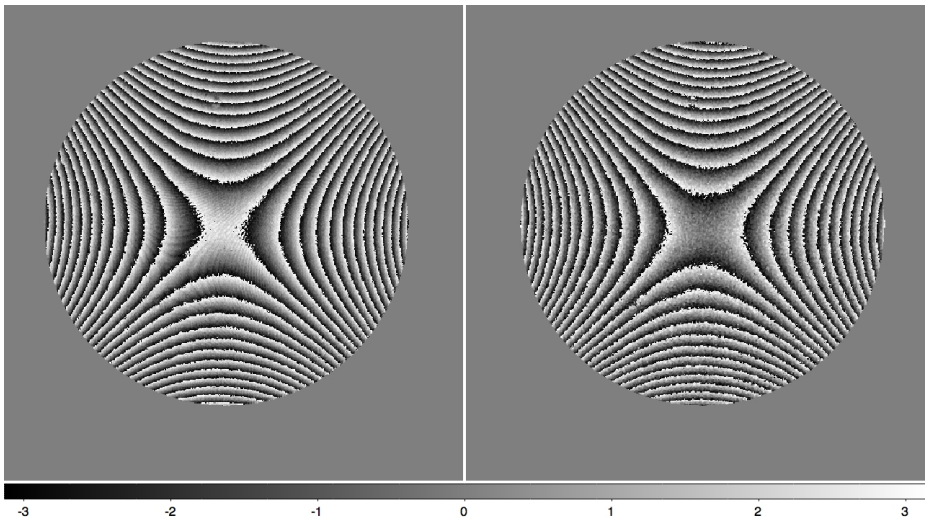


Figure 5.3: Images of the recovered phase map obtained through use of a three step phase stepping algorithm for 5 waves RMS astigmatism using a hologram with a standard binarization on left, and the random binarization on right. Z scale used is in radians.

Figure 5.3 compares the recovered phase map after using a three step phase shifting algorithm for 5 waves RMS astigmatism encoded using a standard and random binarization. The main reason the random binarization technique is considered is that it theoretically reduces the presence of higher order harmonics in the recovered phase profile and we can see in figure 5.3 that this is indeed the case as evidenced by the removal of the subtle astigmatic phase variations in the centre of the recovered wrapped phase pattern for the standard hologram.

Whilst this is true however, looking at the centre of the phase map for the random hologram shows a high degree of noise, and it needs to be quantitatively determined whether this new source of noise is more detrimental to overall wavefront accuracy than the astigmatic artifacts are. A side effect of this however is that the fringe edges become noisier, and also at the edges of the image where the fringe spacing is small there exist many islands of completely discontinuous phase values which may pose a problem when unwrapping the image.

As mentioned in the previous chapter, for the purposes of metrology we do not just require that the SLM is capable of producing exactly the desired quantity of aberration. One of the main applications of this device would be in nulling interferometry so it is important to see how large the residual aberrations produced by the hologram are in the absence of the primary astigmatism term. Due to the absence of requisite optics as well as to eliminate the time it would require to insert and align exact reference surfaces, the residual errors were evaluated by removing the intended aberration magnitude from the recovered phase profile in software before being measured.

Figure 5.4 shows the growth of this residual RMS term with increasing aberration size for both binarization algorithms and shows a discouragingly large residual error for both techniques. As expected, the random binarization algorithm gives a significantly larger error than the standard however this can be explained as mainly resulting from errors in the phase unwrapping procedure. The randomised interferograms show a worse modulation and less clear fringe boundaries than the standard algorithm provides giving large unwrapping errors which completely dominate the null interferogram RMS term. Nevertheless even for the standard holograms the residual RMS error of the null interferogram is in the range of $\lambda/10$ to $\lambda/3$ which is an order of magnitude higher than expected from simulated results and could potentially eliminate the suitability

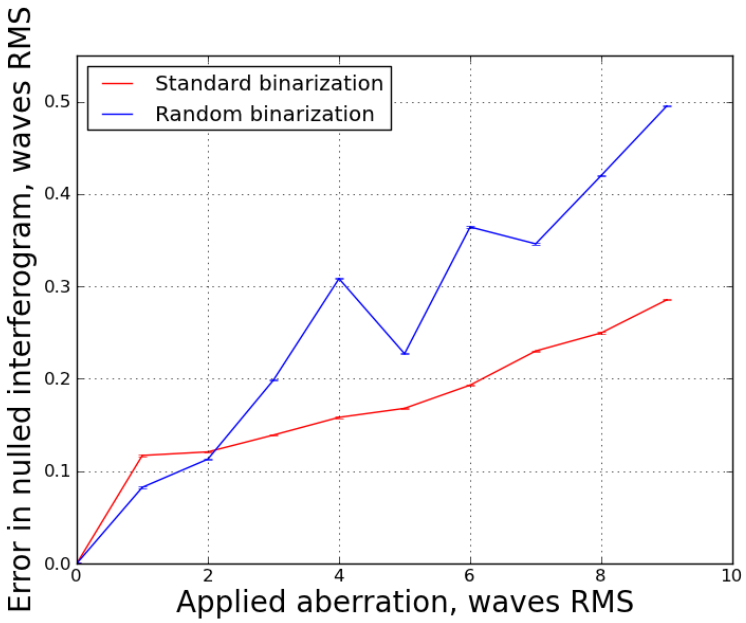


Figure 5.4: A comparison of the residual wavefront RMS measurements for the recovered phase profiles after the desired quantity of Z5 astigmatism has been removed.

of the device for use in nulling interferometry.

5.4 Residual quantisation

As seen in figure 5.4 the RMS residual error for the null interferograms is unacceptably large, however an analysis of these would be beneficial to potentially provide a means of reducing them. To this end a Zernike modal deconstruction of the recovered phase profile was used to determine if part of the residual error was down to the production of other, unwanted Zernike terms. Figure 5.5 shows the size of these for a hologram of 5 waves RMS Zernike astigmatism after the removal of the desired aberration form. For both binarization techniques there is a large peak in the Z3 term corresponding to the alternate Zernike astigmatism polynomial, which is a 45° rotation of the Z5 term. This could be a result

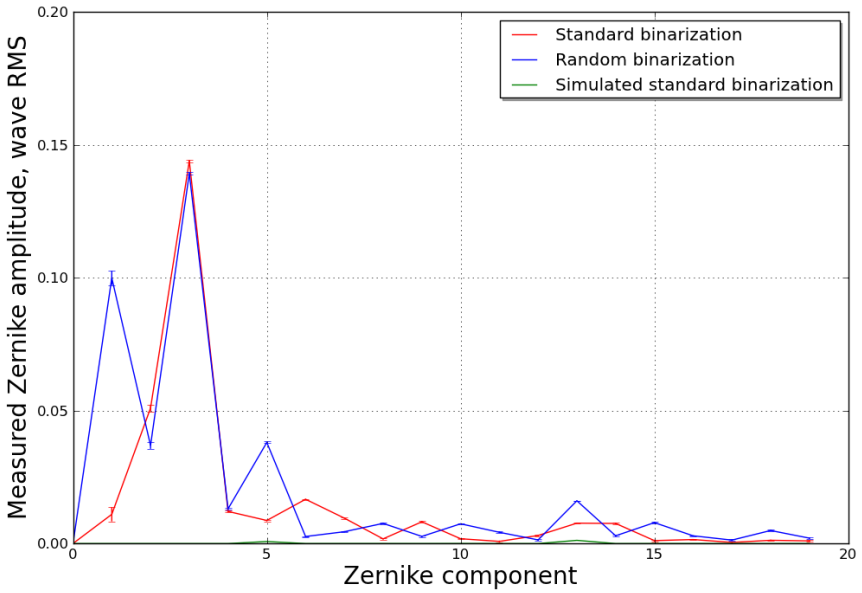


Figure 5.5: The measured magnitudes of the first 20 Zernike terms seen in the experimentally produced wavefronts for standard and random binarization holograms of 5 waves astigmatism. A comparison with the simulated Zernike terms shows that these are significantly larger than expected.

of a slight rotation in the SLM with respect to the camera which will show an astigmatic phase pattern subtly rotated in the image plane, and upon analysis will result in a small decrease in the magnitude of measured Z5 with a small increase in the amount of the Z3 term being detected. The random binarization algorithm once again shows an increase in errors over the standard one, with the detection of an additional large tilt term in the modal deconstruction as well as the large deviation in magnitude of the desired Z5 term.

The presence of these extraneous Zernike terms in the phase maps, whilst unexpected, is not severely debilitating to the overall accuracy of the results. As these aberrations are known to be from the specific hologram used we can remove them from the recovered phase map after data collection which will allow

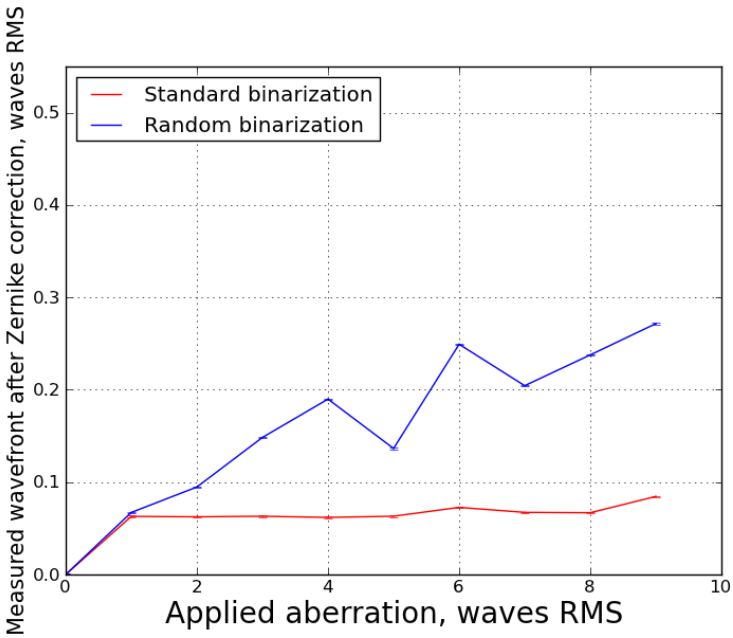


Figure 5.6: Growth of the residual RMS of the recovered wavefront after subtraction of the first 20 measured Zernike coefficients for increasingly astigmatic holograms.

us to investigate smaller scale, non Zernike errors present in the maps. Figure 5.6 shows the behaviour of the residual phase RMS with increasingly astigmatic holograms after the subtraction of the first 20 Zernike terms measured with the modal deconstruction. Once again the randomised algorithm shows significantly larger errors than the standard method, however this will be predominantly due to phase unwrapping errors as mentioned previously. The RMS for the standard algorithm however remains roughly constant as the intended aberration size increases, up until a hologram of 9 waves RMS astigmatism is displayed. This increase in residual error here is associated with phase unwrapping errors as the fringe density for 9 waves astigmatism is high towards the edges of the interferogram. Combined with noise in the image, this makes unwrapping the fringes highly problematic and requires that either a different phase unwrapping

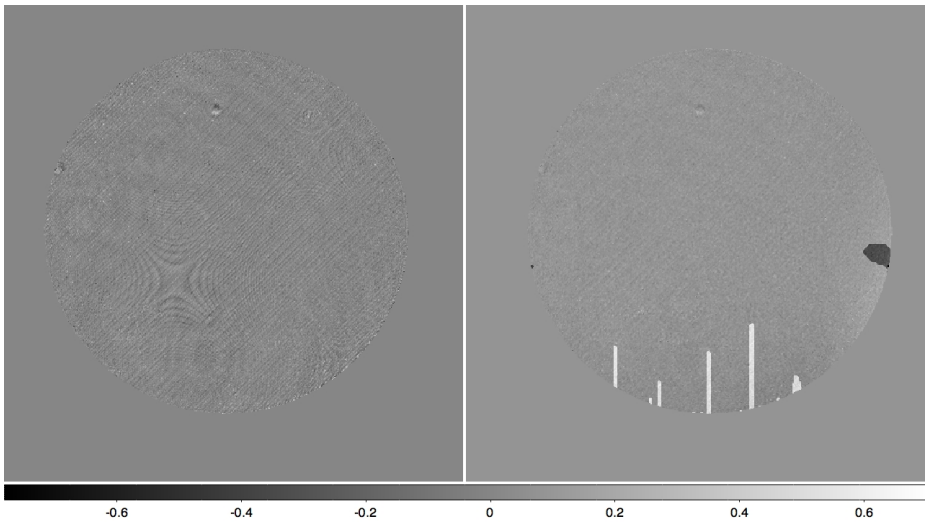


Figure 5.7: Recovered phase map for 5 wave Zernike astigmatism after the first 20 Zernike modes have been removed from the image, leaving smaller scale aberrations behind. Standard binarization on the left, and a randomly binarized hologram on the right and scale is wavefront phase given in radians. This side by side comparison serves to illustrate one of the largest difficulties encountered experimentally, the effect of phase unwrapping errors. The random binarization algorithm was found to consistently be significantly more susceptible to these errors than the standard one.

algorithm be used, or a better sampled image to be obtained ideally using a camera with a larger dynamic range.

The subtraction of these Zernike terms appears to be a crucial step in the processing of the recovered phase data to reduce the resultant error, and could potentially be applied when using the SLM to null out aberrations in test optics provided that for each hologram displayed, the byproduct Zernike terms can be separately measured. Furthermore, the general flattening of the residual error seen in the results from the standard binarization technique after Zernike correction indicates that the only major errors increasingly astigmatic aberrations produce are described fully through Zernike polynomials.

Figure 5.7 compares the recovered unwrapped phase profile for a hologram

of 5 waves RMS astigmatism after correction for the first 20 Zernike terms. Present in the unwrapped map of a standard hologram are astigmatic artefacts which are not seen in the recovered map of a randomly binarized hologram. As mentioned earlier the intention of the randomisation technique is to eliminate the production of higher order harmonics contaminating the recovered phase distribution and dominating the error. Whilst the random technique appears to successfully remove these irregularities it is unclear whether the source of the artifacts are from higher diffraction orders passing through the pinhole or due to small quantities of intensity modulation. The effect of intensity modulation on the recovered phase distribution is that lines will appear on the phase map corresponding to the lines on the diffraction grating. These will not be removed through 3 step interferometry as the phase shifting method is coded directly into the hologram rather than on a separate piezoelectric stage, and as such the faint lines recorded by the camera will have moved between each individual frame. It is more likely this artefact is a result of a contaminating higher order term as the size of the discontinuities seen are of comparable magnitude to that seen in a simulated phase map, in the order of 0.2 radians peak vally.

Also present in figure 5.7 are large errors in the corrected randomised phase map. These are introduced during the unwrapping process as the random hologram gives a higher background error term, complicating the fringe boundaries by introducing larger islands of phase discontinuities which can result in incorrect multiples of 2π being added in the local pixel phase comparison. In order to fully compare the two binarization techniques it would be necessary to use a phase unwrapping method which can unwrap both without adding any unique artifacts into only one of the phase maps, and will be considered later.

Figure 5.8 shows the residual RMS measurement for increasingly astigmatic aberrations after taking the first measured 20 Zernike polynomials away from a

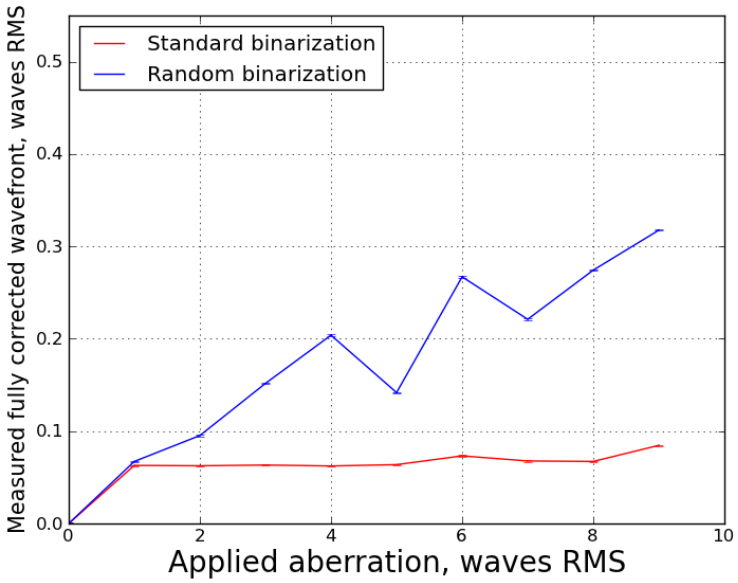


Figure 5.8: Increase of the residual RMS phase error post Zernike correction for a nulled phase map as a function of increasing applied astigmatism term.

nulled phase map. It is unsurprising to note that this behaviour is very similar to that seen in figure 5.6 as the nulled phase map is produced after the phase shifting algorithm is used, so the primary difference between the two graphs is that figure 5.8 includes the error in production of the desired Z5 astigmatism term. A proper null interferogram taken using a nulling optic in the reference arm that can accurately produce the desired aberration would allow for increased accuracy when investigating larger aberrations, as they would eliminate the unwrapping inaccuracies found when analysing very narrowly spaced fringes. Despite the absence of this, the roughly constant residual RMS error seen implies that the lack of a true null interferogram, and thus the presence of a reasonably high fringe density such as found up to 8 waves RMS Z5, is not the limiting factor in further reducing the error of these wavefronts.

Before any processing of recovered interferograms can be performed it is nec-

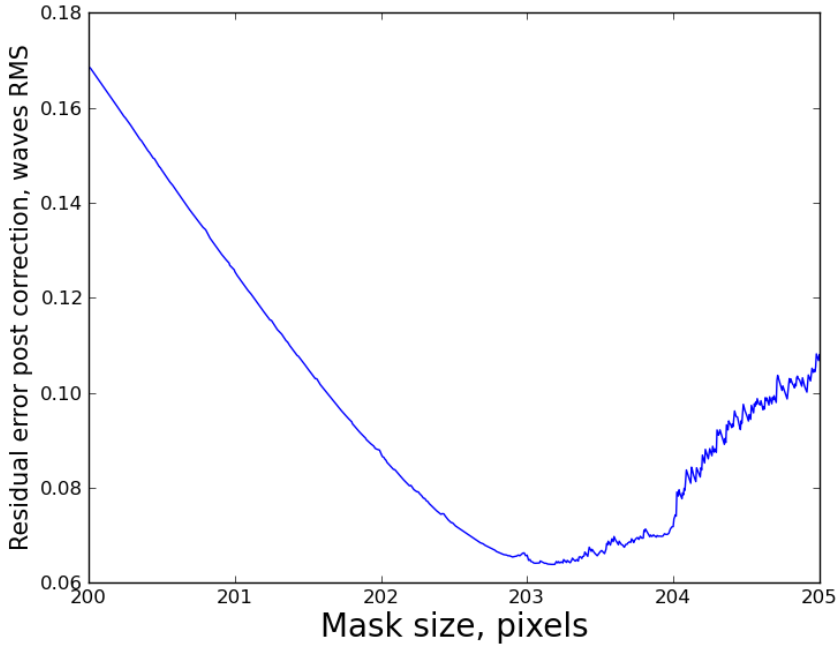


Figure 5.9: Growth of the residual RMS errror for the corrected phase maps after subtraction of desired aberration term as a function of size of the imposed software mask for an aberration of 5 waves RMS astigmatism. Initial decline in error is a result of size matching of the observed and comparison wavefront shapes, and increase in residual for large mask sizes is due to increasing inclusion of noisy pixels outside the hologram being included in the analysis.

essary to define the area of the image to be analysed, so a mask is imposed over the frames which ensures that data from outside the interferogram is not measured in the modal deconstruction analysis. The size of the mask is extremely important however. Too small and the entire hologram will not be measured and, more significantly, during the modal deconstruction the Zernike map that is compared against the recovered phase will be incorrectly scaled and thus yield an erroneous measurement of the magnitude of aberrations present. If the mask is too large however then the latter problem of incorrectly scaled Zernike map comparisons will still be an issue, but will also contain the noisy data which is

outside the interferogram and when unwrapped becomes significantly larger in magnitude as the program attempts to add or subtract multiples of 2π thinking it is a legitimate phase point. Of paramount importance here is that the wavefront measured is conjugated correctly to the plane of the SLM, otherwise the comparison of reference Zernike maps, which are defined over a circle, will give incorrect results also. As the conjugation process involves comparison against circular software masks it is also a useful opportunity to determine the radial extent of the hologram, measured in pixels on the camera, this approach however is rather crude and inaccurate so further analysis is necessary.

Figure 5.9 shows how the residual wavefront RMS after correction for applied aberration, systematic errors and aberrant Zernike terms depends on the size of this imposed software mask for a hologram of 5 waves RMS astigmatism. A gradual decline in the error is seen up to a size of 203.05 pixels as the expanding mask measures increasingly large portions of the hologram up to where the reference Zernike aberrations become of similar spatial extent to the recovered aberrations. With a pixel size of 4.65 μm this corresponds to a mask of 1.88 mm being imposed on the beam after it has been de-magnified down onto the camera. Above this however the error increases as more and more noisy irrelevant background data is introduced into the analysis. The initial gradual increase is mainly due to incorrect size matching of wavefront to comparison aberration size, and then jumps sharply when larger quantities of the noisy pixels not associated with the hologram are included. This would be an expected result as a consequence of the local approach to phase unwrapping. A pixel on the very edge of the detected hologram will be unwrapped correctly, ideally, as it is compared to a phase value inside the hologram and adjusted accordingly. Consider now a point with distance from the centre of the image larger by one pixel. This data point will not be part of the hologram however

when undergoing the unwrapping process it is assumed to be as the exact size of the ideal software mask is unknown. When this point is compared to a pixel inside the true hologram then it will be also adjusted accordingly so that it is within $\pm\pi$ of the holographic edge pixel. Moving outward radially a further pixel, we can see that whilst this will adjusted to be within $\pm\pi$ of the previous pixel phase value there is no guarantee of its value in relation to the hologram edge pixel, as the noise here is random.

We have seen that for a correctly sized software mask the FLCOS SLM is capable of wavefront control with errors in the region 0.06 waves RMS. Whilst this residual error is not as small as that seen in the simulated results contained in the previous chapter, such a result is not to be dismissed. The ability of the device to be of practical use in preserving or revealing small scale structure on an optic is examined in the next chapter, however before this is done it is worth considering if it is possible to reduce the most time consuming aspect of the process: unwrapping the recovered phase map.

5.5 Phase Unwrapping considerations

Before modal deconstruction and analysis on all simulated and experimental phase maps it is necessary to unwrap the fringes produced by the phase stepping algorithm as they come out wrapped module 2π. Phase unwrapping is a complicated and developing field[71][72][73], and there exists no algorithm which has proven to be universally reliable in the unwrapping of fringe patterns. For obtaining unwrapped phase maps of the data obtained here a simple localised approach was taken comparing the phase values of specific pixel with that of it's nearest neighbours and adding or subtracting multiples of 2π when and where necessary. This code was written in Python by the author of this Thesis. The decision to use this approach rather than adapt existing techniques was made

due to time constraints present which prohibited the in depth understanding necessary to implement previously established algorithms.

Before the main phase unwrapping algorithm can be used on the recovered wrapped phase profile however the noise in the image needs to be reduced. The raw phase profile contains fringes whose edges are highly jagged due to the effect of noise pushing phase values over the 2π limit with the result that they are wrapped from 0. This leaves isolated islands of substantial phase difference to the surrounding area and can lead to errors in the main algorithm which necessitates their elimination. To remove these and create smoother lines of discontinuity each pixel is compared to the 9 surrounding neighbours, and if more than a specified number show a $\pm\pi$ phase difference from the original, that pixel is modified by $\pm 2\pi$. The number of phase discontinuities allowed in neighbouring pixels is modified over several iterations of this procedure to eliminate as many islands of discontinuous phase as possible so as not to present possible sources of confusion in the main unwrapping procedure. Pixels with 0 phase value are also eliminated in a similar manner. Instead of just being modified by $\pm 2\pi$ however they are assigned the average value of the surrounding pixels. If the surrounding pixels overlap a fringe border then the average is taken from the side of the fringe with the highest number of bordering pixels.

For the main unwrapping algorithm, visually depicted in figure 5.10, a pixel close to the centre of the image, $p(x, y)$, was chosen as a start point from which to begin unwrapping the fringes and compared to its neighbouring pixel $p(x+1, y)$. If the difference between the new pixel and the old one is $\pm\pi$ then a phase value of $\pm 2\pi$ is added to eliminate phase discontinuities. Pixel $p(x + 1, y + 1)$ is then similarly compared to $p(x + 1, y)$. Pixel $p(x, y + 1)$ however compares phase difference with the already checked $p(x, y)$, and then $p(x - 1, y + 1)$ compares with $p(x, y + 1)$. The overall effect is for the phase checking program to spiral

out in a square around the first pixel comparing phase value with an already corrected neighbour clockwise to it in the sequence, or with the pixel directly preceding it when the former is unavailable.

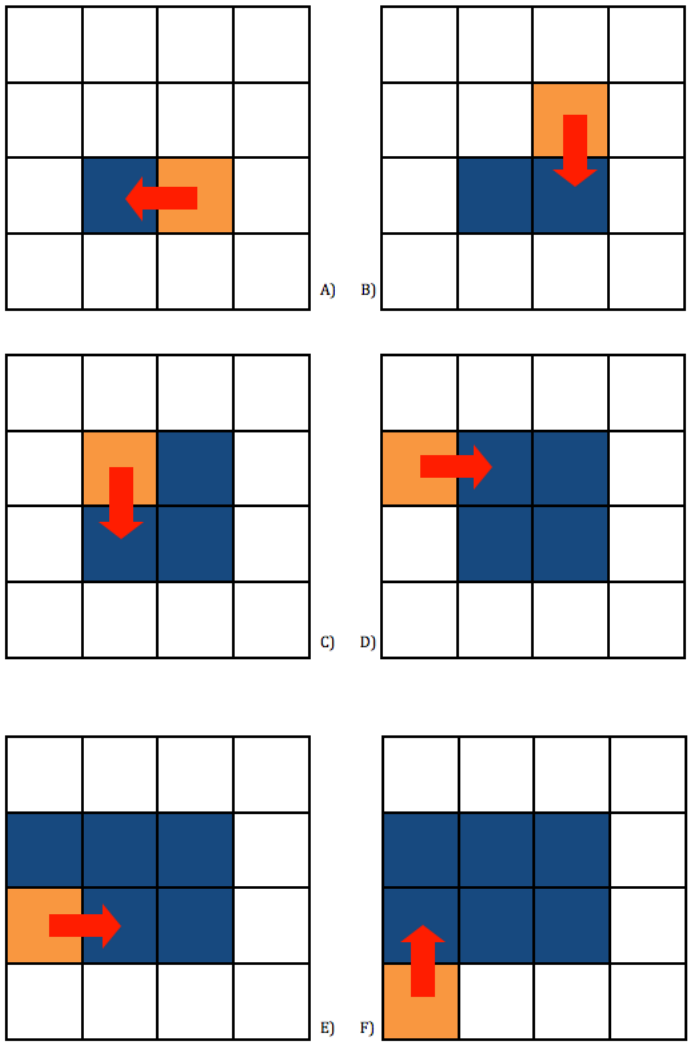


Figure 5.10: From A) to F), the first six pixel comparisons in the main unwrapping procedure. The red arrow points from the new pixel to the one it is being compared with.

One significant problem with this approach to phase unwrapping is that

an error in the unwrapping of a pixel will be propagated through the unwrap algorithm so that every pixel which is compared to an erroneous one will have a constant offset to the desired wavefront shape. In some cases this appears as a thin line of offset pixels running a partial length of the unwrapped map and therefore it is possible to reduce this effect by a further application of the initial noise elimination method. It is also time consuming when reducing the noise in the fringe edges rendering it unsuitable for use in a situation where high speed data acquisition would be desired. It is important to once again state that this is a highly simplified approach to the problem of phase unwrapping an image and a full approach to the subject would require significantly more complicated bespoke code. The unwrapping errors are especially problematic in the analysis of the interferograms corresponding to a randomly binarized hologram as seen earlier, so it would be of interest to compare this technique to another method of unwrapping the phase.

Figure 5.11 shows the growth of the recovered Zernike astigmatism RMS for the locally unwrapped standard binarization technique compared with the recovered values for the standard and randomly binarized holograms recovered using a global Fourier unwrapping method [74] with the assistance of Ali Bharmal, Durham. This method applies trigonometric functions to the Fourier transform of the wrapped phase distribution in order to estimate the number of 2π phase jumps between neighbouring discontinuous points. It can be seen that there is a drop off in the magnitude of the recovered aberration unwrapped using the Fourier approach despite the initial wrapped phase profile being identical to the one unwrapped via a local method. The Fourier approach fails to recover an accurately scaled phase distribution of the input wavefront. This scaling factor appears to not be the only source of error introduced into the unwrapped phase as determined by comparing the unwrapping of a well known arbitrary

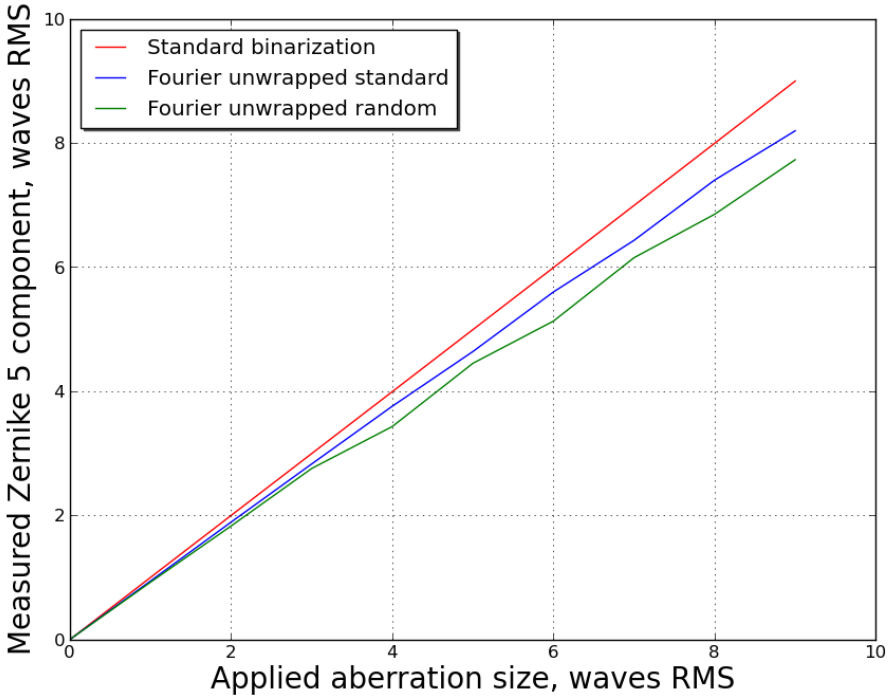


Figure 5.11: Comparison of the recovered Zernike astigmatism term for increasingly aberrated holograms between the locally unwrapped standard binarization algorithm, and both standard and randomly binzarized holograms using a global Fourier based unwrapping approach.

reference wavefront. By computing a modulo 2π map of exactly one wave RMS Zernike astigmatism and unwrapping it using both approaches it is possible to use the scaling difference as an normalisation factor for the Fourier approach and then comparing the recovered astigmatism values for the experimentally obtained map of a desired aberration of the same size. Even in this case the Fourier unwrapping procedure recovers an astigmatism value of 0.94231 waves RMS compared to the locally unwrapped 0.999411 waves RMS.

To determine other sources of error introduced by the usage of a Fourier based unwrapping approach it is worth looking at a comparison between the

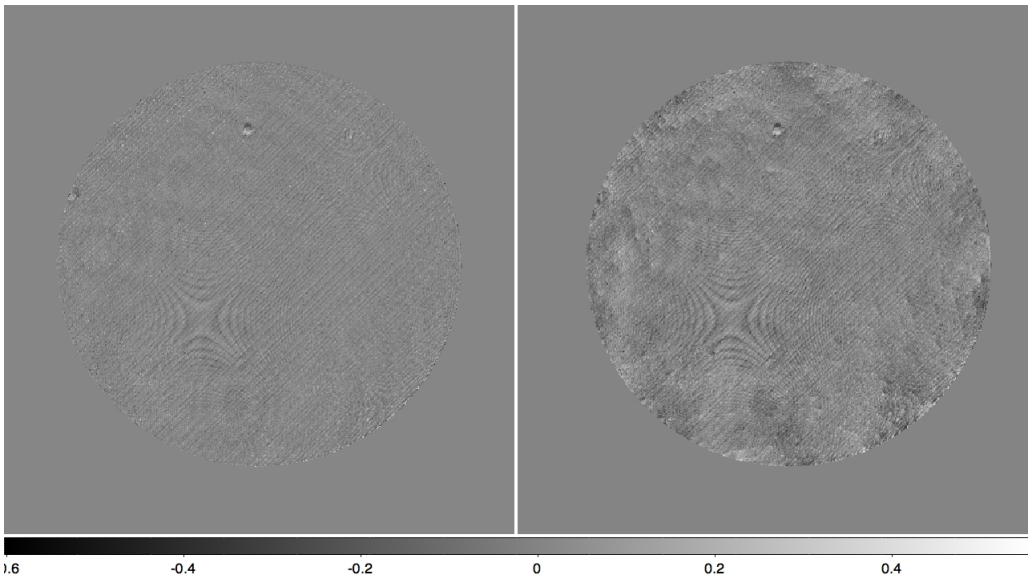


Figure 5.12: The recovered unwrapped phase map after correction of the first 30 Zernike terms for 5 waves Zernike astigmatism as produced using the locally unwrapped approach on the left, and the Fourier approach on the right. Zscale used is wavefront phase in radians.

unwrapped phase maps produced by both methods. Figure 5.12 shows this comparison of the unwrapped map of the experimentally obtained result for 5 waves astigmatism using a standard hologram and after the subtraction of the first 30 Zernike terms. The reason for the increased depth of the Zernike analysis here is that the Fourier unwrapping process induces a substantial amount of hexafoil aberration, Z27, which needs to be removed. It can also be seen that the Fourier unwrapping method produces a large degree of noise which is not seen in the locally unwrapped version, particularly at the edges of the image.

Figure 5.13 shows how the wavefront residual RMS increases with increasing hologram aberration for both binarization algorithms and unwrapping methods after correction of the first 30 Zernike terms. Significantly here we obtain a much improved unwrap of the random binarization technique allowing for a more illuminating comparison between it and the standard binarization in the

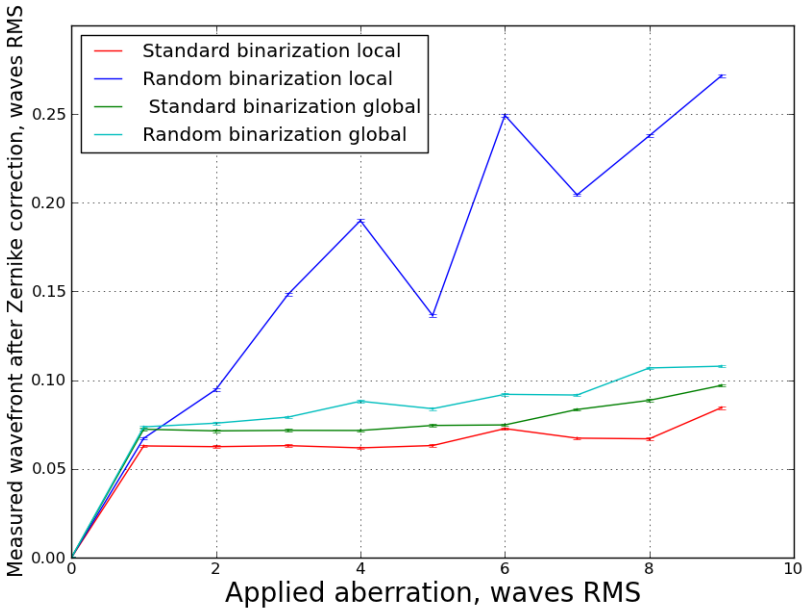


Figure 5.13: The growth of the residual RMS wavefront error after correction of the first 30 Zernike terms for the local as well as Fourier unwrapping algorithm for both binarization techniques as a function of increasingly astigmatic hologram.

level of residual error produced. Here we see that the standard holograms show a smaller residual RMS after Zernike correction than the random ones despite the elimination of intensity modulation contamination. We can also see that the local phase unwrapping technique gives less residual error than the global one, suggesting it should be preferentially used for standard holograms. It can also be seen that at small aberrations for the standard binarization the residual RMS error remains constant after Zernike subtraction for the global unwrapping technique as well as the local one, with a residual growth picking up towards 6 waves applied astigmatism. For the global unwrap however the error picks up faster than in the local case suggesting that the presence of narrowly spaced fringes is more detrimental to the accuracy of the unwrap in the former than in

the latter. Whilst the local phase unwrapping procedure is often unreliable in successfully retrieving a phase map it is still preferable to the global method as it does note alter the scaling of the wavefront shape.

5.6 Conclusions

In this chapter we have seen experimental results to match those simulated in the previous chapter. The growth of a desired aberration term with increasingly large magnitudes displayed on the hologram is seen to be nearly linear and similar to that predicted from simulated data, with a robust insensitivity to residual vibrations in the interferometer. This experimental validation of the theoretical results allows us to be confident that the method used to reconstruct analogue wavefronts from binary holograms is a viable technique for wavefront manipulation. Errors in the production of desired aberrations are seen to be less than a hundredth of a wave, again as seen in our simulation, until the aberration propagated becomes too spatially extended in the Fourier plane to pass fully through the pinhole. A comparison is drawn between the standard and random binarization techniques which agrees with the results seen through simulation, the introduction of the random background noise term gives rise to a larger residual error than that corresponding to higher order terms contaminating the phase map. This result is shown in figure 5.7 and can be compared with the simulated results in figure 4.18. Despite the accurate growth of desired aberrations however the residual nulled wavefront error seen is significantly larger than that predicted, showing a factor of ten increase over values obtained through the simulation. This deviation highlights the need for an in depth analysis of the experimental errors seen. A large part of these can be eliminated through a Zernike modal analysis, as these errors are known to be a result of hologram production we can safely remove them from the recovered phase map and thus

reduce the residual RMS error to under a tenth of a wave. It is shown that after this Zernike correction the residual error is roughly independent of the size of the aberration being evaluated, and only increases when the size of the PSF becomes sufficiently spatially extended to start being cut off by the filter in the Fourier plane. Another potential reason for the difference between simulated and experimental residual error could be from the degree of stray light passed due to diffraction effects from the pixelated nature of the device, an effect not observed in our simulated model of the device.

We have also seen that one of the largest difficulties encountered in the analysis of the experimental interferograms is related to the phase unwrapping step, which is necessary to perform the modal deconstruction. In particular for unwrapping phase maps produced using a random hologram the method has shown to be unreliable, and producing a map free from unwrapping artefacts is highly difficult and further lessens the suitability of the random binarization technique. For a valid comparison between standard and random binarization however a globally based unwrapping technique was also used, which was found to successfully unwrap both binarized phase maps and it is seen from this that the standard method still gives a reduced residual RMS wavefront error. A comparison between the two unwrapping techniques is given which indicates that despite the time consuming and occasionally unreliable nature of the local unwrapping method, it is more suitable than the global unwrapping procedure which introduces it's own Zernike aberration terms as well as providing as output a phase map scaled by an arbitrary factor.

Despite these residual non-Zernike errors in the nulled phase map, the accuracy of wavefronts produced approaches $\lambda/20$ and the use of the SLM in removal of aberrations for reducing interferometric fringe density in the measurement of optics is worth exploring, as covered in the next chapter. We can also com-

pare the results here with existing values for the residual errors produced via binary holograms on FLCOS devices [12] where a hologram of 2 radians RMS spherical aberration is produced using a standard binarization technique and shows a residual wavefront error of $\lambda/10$. This residual error is larger than that seen experimentally in this thesis, with a significant reduction in higher order artifacts due to a combination of increased separation of the diffracted order terms as well as using a larger hologram to create the aberration.

Chapter 6

Implementation of FLCOS SLM in metrology

6.1 Introduction

The previous two chapters have presented simulated and experimental data on the ability of an FLCOS SLM to create arbitrary wavefront shapes as well as considerations regarding the minimising of the errors produced with them. This chapter shall cover the use of the FLCOS device in conjunction with a test optic larger than the beam width to measure the small amplitude variations in wavefront shape induced by the optic. The comparatively larger size of a test optic compared to the hologram necessitates incorporating a stitching technique to combine multiple smaller subimages, creating a composite map of the whole optic as it may not be possible to ensure sufficient sampling once the interferogram is demagnified onto the detector CCD. As the FLCOS device is reprogrammable, for every translation of the test optic it is possible to create a new hologram to null the aberration present allowing for a high degree of flexibility

in potential future applications. Firstly we will present a brief overview into how the stitching will be applied to the measured phase maps before covering the implementation of it in the measurements of two optics. Finally we discuss the limiting effect of noise on residual accuracy resulting from the camera used experimentally.

6.2 Stitching program

In order to use the interferometer described in previous chapters for measuring large optics there needs to be a way of combining smaller images to make a full map of a larger surface. This can be done by attempting to obtain information on the full aperture aberration function[75] or through the minimisation of a pre defined aberration function[76]. Image stitching, similar to phase unwrapping, is a complex field with much work involved in the optimisation of stitching algorithms [77][78]. Once again time constraints prohibited the in depth understanding required to fully implement established techniques and instead a basic stitching program was written in Python that could combine the experimentally recovered phase maps and produce a full map of the test surface. To avoid complications in translational accuracy that can be found in manual translation stages, where both errors in reading and the lack of calibration can cause inaccurate movements, two automated stages provided by the National Physical Laboratory were used with Physik Instrumente micro controllers. The stages used allowed for calibration upon start up and translational movement down to an accuracy of 0.5 μm.

In order to combine images it is important to understand the scaling between movement of the optic on the stage and movement of the image on the camera. It is inconvenient therefore to measure an optic without any form of fiducial marks or reference points which can be used to measure this scaling factor. In

the first optic measured, discussed below, the quality of the manufacturing was poor and this resulted in significant machining marks being present which could be observed in the recovered images. By fitting a circle to the machining marks of multiple translated images it was determined that a movement of 1 mm on the translation stage corresponds to a 31 pixel shift on the image, for both x and y co-ordinates.

The actual process of stitching multiple images together used is quite simple. Initially, two images are positioned relative to each other using the determined scaling between translation and image, and the mean of the overlap region for both images is obtained. Taking the difference between the two means and subtracting this DC value from one of the images should theoretically be sufficient to seamlessly stitch the two together if the wavefront of the overlap area has the same shape in each image. Further images are stitched onto this one at a time in the same manner, comparing its overlap onto the composite, until a full map is acquired. When the phase map contains areas outside of the physical optic the three step analysis recovers meaningless noise which then made more chaotic by the phase unwrapper so it is important to ensure that the overlap of these regions is not taken into account when comparing frames, achieved simply through the use of software masks. This simplistic approach does not take into account the effects of geometric distortion in the recovered wavefronts, nor does it take into account continuity of slope and curvature. As the intention of the experimental work was to model the propagation of aberrations through the system, including tilts and defocus, no modification of the wavefront form was made beyond the subtraction of the DC offset between sub frames. It is suggested at the end of this thesis that further optimisation of this approach would be beneficial, in particular an optimisation of the DC offset subtraction through dithering the offset term of one frame and measuring

the effect on residual wavefront error.

6.3 Measurement of refractive transmissives

As mentioned previously, one of the potential uses for the FLCOS device is to increase the dynamic range of interferometry. When measuring highly aberrated optics there is a reduction in the accuracy to which one can measure small scale deviations in surface form due to the high fringe density swamping out the effects on the fringe pattern. Furthermore when measuring surfaces that deviate significantly from the reference surface used, one needs to consider whether the density of fringes will become sufficiently high that sampling issues come into effect, and it is not always convenient to magnify the wavefront so it appears larger on the detector. As the FLCOS SLM is reprogrammable it gives us the possibility of measuring aberrated, highly irregular optics with a new degree of accuracy as sections of the surface containing a large arbitrary aberration not present elsewhere can be compensated for using one hologram, and then a second can be displayed for the measurement of the rest of the piece.

The optic manufacturers based at Netpark which form a part of the CfAI were asked to produce an aberrated optic for the testing of this premise, and created a refractive approximately 1 inch optic containing nominally 10 waves PV Zernike trefoil. This was then mounted on the electronic translation stages and the prominent machining marks visible on the device were used to establish the scaling relation between stage movements and image displacement on the camera as mentioned previously. It should be noted that tests of this refractive element using a Zygo interferometer within the laboratory failed, as the software could not fully detect and unwrap beyond a small central region.

The first test of the viability of using the device in such a way is to establish if the images can be successfully stitched together in the first place. Figure 6.1

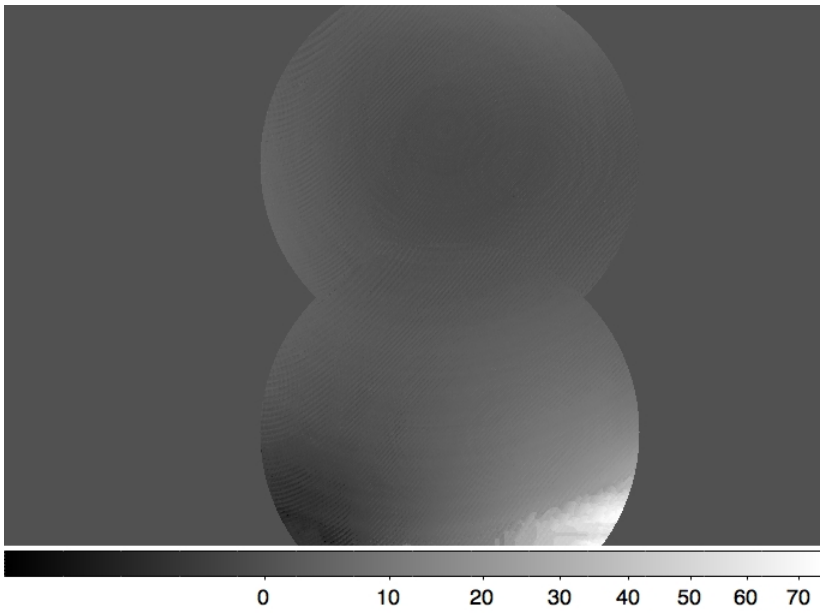


Figure 6.1: The stitched composite phase map of the first two subframes gained through interferometrically measuring the transmissive optic using a flat hologram on the FLCOS device, scale in radians. The random errors seen at the bottom of the image correspond to phase unwrapping artifacts which are produced when unwrapping the phase map at points outside of the optic. X and Y axes correspond to detector pixels.

shows the addition of two frames of data for the refractive optic measured using a flat hologram and the main point of note is that the boundary between the second and the first image is free from large discontinuities. We can establish therefore that it is likely that this simple approach to image stitching could be sufficient and once static systematic errors have been compensated for in the data analysis there is no new source of wavefront deviation present between the measurement of the first frame and the second. There are present the structured aberrations caused by intensity modulation being passed through the system as predicted in the previous chapter. Also visible are slight phase aberrations caused by the machining marks in the optic however these do not dominate as a result of using a phase stepping algorithm. Both of these artefacts are of the

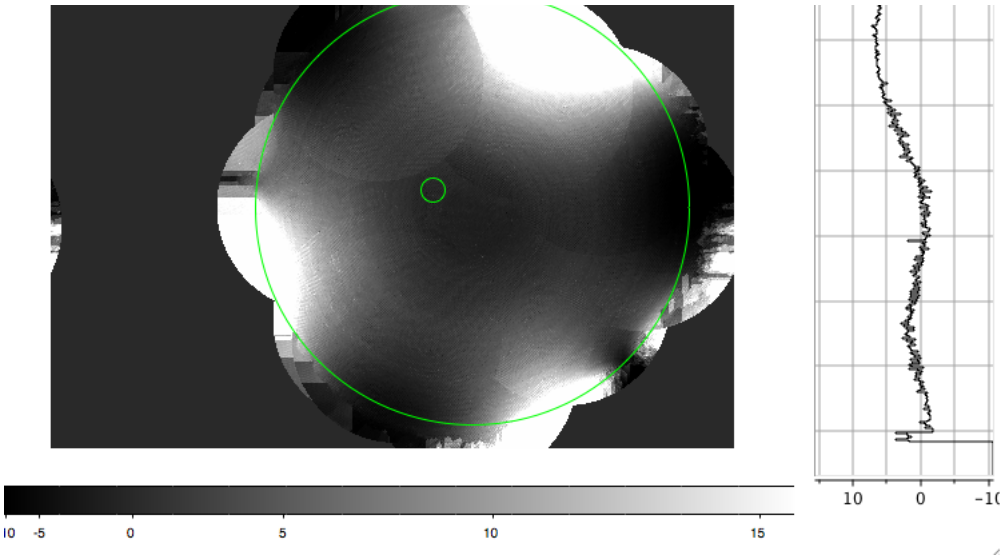


Figure 6.2: The full measurement of the refractive optic using 12 smaller images and a flat hologram. Areas of noisy unwrapping related aberrations are expected outside of the optic and help show the boundaries of the element. On the right is a vertical slice of the composite for the centre of the small green circle with scale in radians showing that the discontinuous overlap do not cause phase jumps significantly outside that of the general noise. The large green circle here shows the boundary of the usable region of the optic.

size 0.5 radians PV and not larger than the background noise levels. At the very bottom of figure 6.1 can be seen the results of the phase unwrapping acting on data outside of the optic where the results are meaningless except to show the boundary of the test piece.

Figure 6.2 shows the full stitched phase map of the refractive transmissive made using 12 smaller images all taken with a flat hologram on the SLM. Unlike figure 6.1 we can see here that it is highly noticeable where the frames overlap by where they show phase discontinuities. These errors were ruled out as stemming from an incorrect scaling factor through repeatedly stitching over a series of different values with no significant reduction in the presence of the discontinuities being observed. The areas of the optic in the outlying regions

do not show a uniform rotationally symmetric transmission pattern but contain severe errors necessitating that we define a circular area of interest centred on the machining marks, and defined in 6.2 with the larger green circle. Scale given is in radians, and a greyscale varying as the radians value squared is used to enhance the visibility of phase discontinuities. This is necessary as the optic contains an aberrated profile and small scale phase discontinuities will be lost otherwise as they comprise only a small fraction of the peak to valley range. By looking at a vertical slice through the image as seen on the right of figure 6.2, we also can see that the stitched phase map discontinuities, whilst noticeable, are not significantly large and do not stand out from the noise in the result. Nevertheless the presence of these indicates that there is some degree of stitching error, potentially from unaccounted form deviations which develop between the measurement of different frames which needs to be considered.

Despite small errors being introduced by the stitching method, the presence of phase discontinuities not significantly larger in magnitude than the background noise of the data suggests that it may be worth expanding the technique further. Figure 6.3 shows the stitched image for two frames of data taken that correspond to a nominally null interferogram. The grey scale on bottom of image is the z scale is left in radians as subframes are stitched immediately after the phase unwrapping procedure. This composite is obtained by displaying a hologram showing the same phase distribution as the originally measured optic, as taken with a flat hologram, and then correcting for systematic aberrations . This is done through the direct subtraction of a flat hologram taken without the test optic to eliminate static system aberrations as well as subtracting a phase map corresponding to errors in the new hologram as compared to the original measurement of the optic. In this case of a desired null test of the optic we see that the stitched images are not as neatly matched as previously

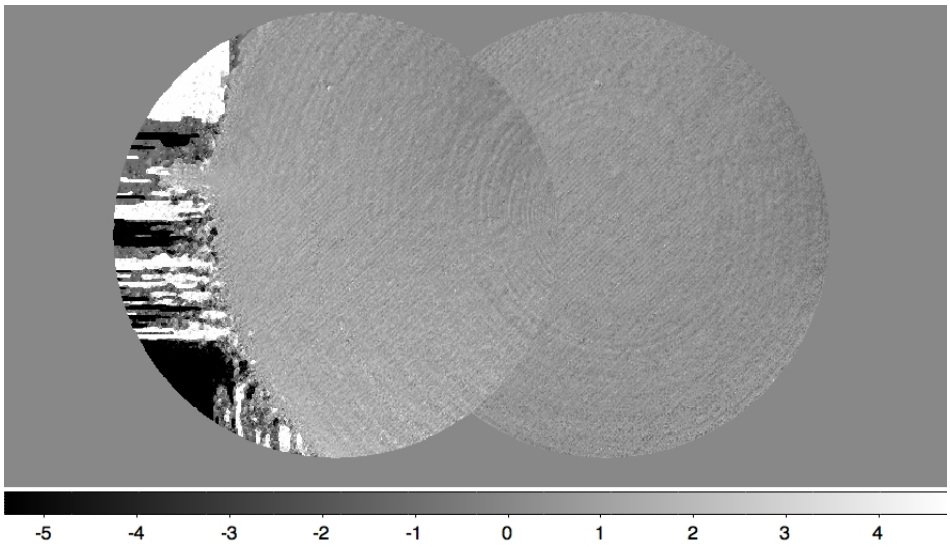


Figure 6.3: Composite phase map of the first two subframes taken using a hologram of the original aberration on the SLM, designed to completely null the recorded phase map of each section. Scale in radians.

was the case. This discrepancy is less significant than it appears however, as a cursory visual inspection of the figure leads to the conclusion that the shape mismatches are much larger than further examination suggests, with less than a $\lambda/10$ deviation in phase between the two sides of the boundary. This is not significantly different from the noisy high frequency form deviations seen in the phase map and appears to be more significant due to the fact it is a structured deviation. Nevertheless it would be beneficial to observe how this reduces the overall accuracy of results especially after the stitching of more images.

Figure 6.4 shows how the full stitched map looks after all 12 subframes have been taken and analysed. As expected there are frames which appear not to stitch neatly together as seen in previous figures however compared to figure 6.2 it appears upon first inspection to be reduced. By looking at a vertical slice through the image we can see that the effect of these mismatches is mostly lost amongst the noise, however not entirely in the case of the subframe directly

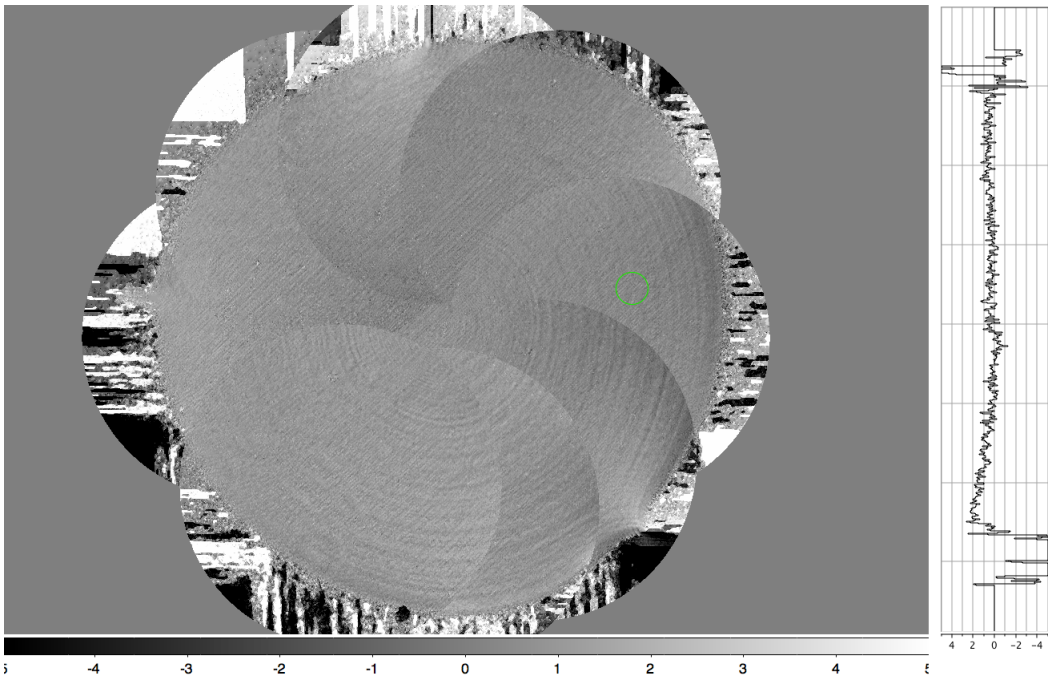


Figure 6.4: The full stitched phase map for an attempted null test of the refractive optic, with scale in radians, using holograms matching the detected phase profile of the originally measured wavefronts. Here we see stitching discontinuities that are due to form deviations between the two frames in the overlap region which are large in magnitude than the measurement noise seen in the vertical slice on the right.

below the green circle. It is also important to note here that the vertical slice does not run through the most serious of discontinuities and that a notable form deviation is present here. This is highly problematic as this is meant to be a null test and so the elimination of these inaccuracies is critical in obtaining any high accuracy measurements.

One potential cause of these stitching mismatches directly leads from the time consuming nature of the phase unwrapping software and highlights the need for setting a number of conditions on the experimental procedure. For example it is vital to fully allow the laser enough time to stabilise, as not doing so can be seen to give a time dependant phase shift to the image which will add

to the intended phase shift with the result that the three step algorithm is no longer exactly suitable. More crucially there exists the possibility for environmental variables to alter during the experiment. In order to take the 12 images required to stitch together a full nulled map of the transmissive optic, it can take up to 10 hours depending on the quality of the interferograms. This is due primarily to the time consuming nature of the phase unwrapping program used and is not an inherent limitation of the FLCOS device. Over this time it is reasonable to assume there will be variations in the air currents in the laboratory which will lead to varying degrees of air turbulence, compounded by the effects of air conditioning and motion of other members of the lab. Whilst ideally the environment would be controlled precisely and the interferometer completely enclosed this was not feasible for this experiment. It is possible to account for these effects however through re-taking new flat frame interferograms periodically to eliminate the effect of systematic aberration variations.

Figure 6.5 shows the residual phase map for the fully stitched measurement after the removal of the first 20 Zernike mode aberrations, of total magnitude 0.0504 waves RMS. This leaves a wavefront with 0.0727 waves RMS aberration which can be mostly attributed to the errors in image stitching. By looking at the vertical slice taken on the right of figure 6.5 we now see that the phase discontinuities are prominent against the background noise of the image confirming the presence of form deviations on overlap regions between frames however this does not have a profoundly detrimental effect on the recovered accuracy. Considering the results of the previous chapter which indicate for a single frame we can expect a residual error in the region of 0.065 waves RMS shows that there is only a marginal increase in the residual RMS over a region roughly four times the size which is not a discouraging result. Despite this it should be noted that the transmissive optic used did not contain an overly large aberration, only 10

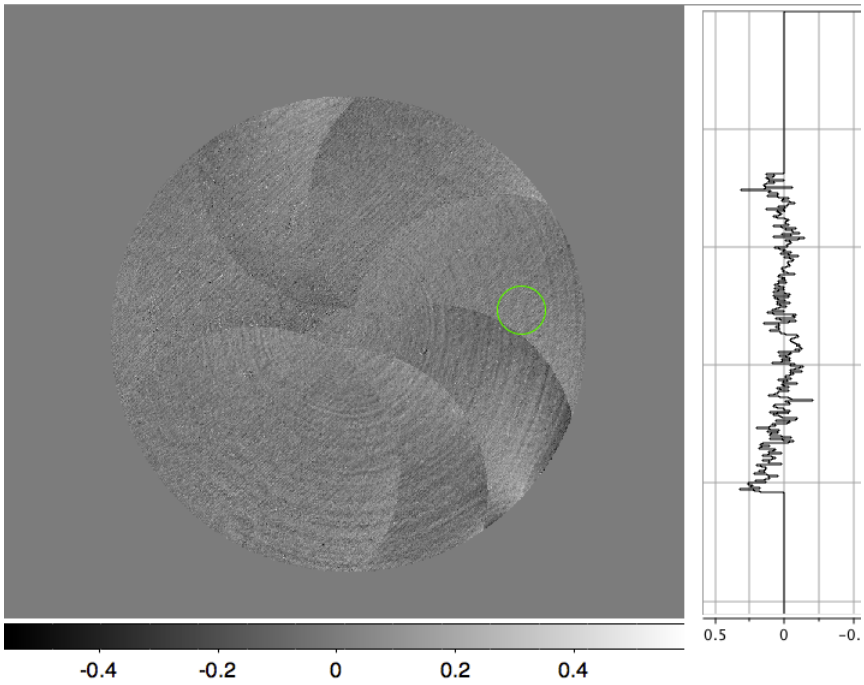


Figure 6.5: Greyscale used in waves. The stitched map after subtraction of the first 20 measured Zernike components and application of a mask diameter 2.3 cm centred using the machining marks. The RMS error = 0.0727 waves RMS after subtraction of a total Zernike based wavefront of 0.0504 waves aberration.

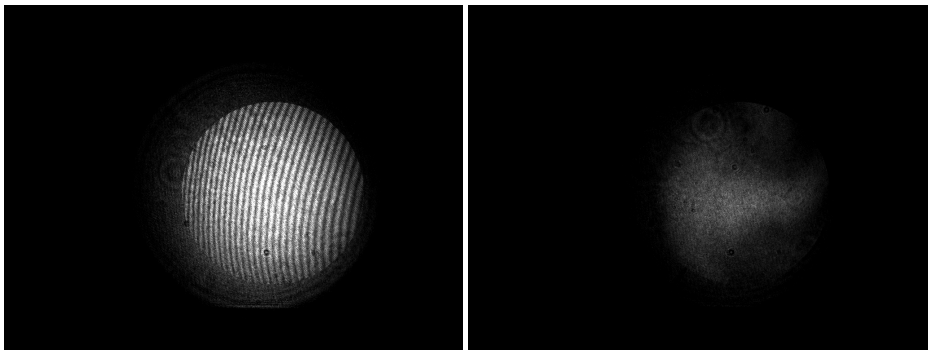
waves peak to valley, and it is unclear whether the Zernike aberrations detected were due to the errors induced through stitching or the increase in accuracy through a significant reduction in fringe density. It would be ideal to apply this method to a more aberrated optic where a high fringe density could realistically mask a smaller scale form deviation as in the following section.

A further difficulty encountered in the measurement of the transmissive was that only the central subframe provided a fully circular phase map for transformation into a hologram. The other subframes necessitated that the corresponding hologram be made from transforming the phase map directly into a hologram rather than calculating one purely to remove an orthogonal Zernike set of aberrations. This resulted in noisier interferograms than those which

would have been produced through a set of Zernike aberrations causing problems in unwrapping and vastly increasing the time needed to collect a data set. It can also be seen at the bottom of the vertical slice through figure 6.5 that the hologram did not fully reproduce an edge effect which contributes largely to the residual error seen. Furthermore the use of individual holograms for each subframe would complicate the translation of the nulled phase map into a set of parameters to determine the exact phase aberration induced, and it would be simpler to use a single hologram for each subframe. The need to take several unwrapped phase maps for each subframe: optic with a flat hologram, hologram of optic, optic with hologram of optic and in most cases a flat hologram for detection of systematic errors leads to it being a time consuming process. With the need to process multiple interferograms for the cases where phase map was difficult to unwrap, it is realistically unfeasible to use a new hologram for each subframe and unless a faster phase unwrapping process is implemented, it has little use.

6.4 Vacuum Window testing.

The Atomic and Molecular research group in Durham contacted the Centre for Advanced instrumentation for help in measuring a vacuum window they were using in their experimental research which they believed was causing significant aberrations in their beam shape. Initial attempts to measure it using a Zygo interferometer proved unsuccessful due to a combination of high fringe density and poor fringe contrast in the recovered interferogram, so it was proposed that it might be possible to measure this window using the FLCOS interferometer presented in this thesis. A successful measurement would have the immediate practical application of allowing the group to compensate for the aberrations in their experiment, allowing them to better control their beam width within



(a) (b)

Figure 6.6: Interferogram for the central subframes of the measurement of the vacuum window using a) flat hologram and b) a tilted hologram with defocus in. By reducing the fringe density using an aberrated hologram we can increase the sensitivity and accuracy of the measurements.

the vacuum which they are currently unable to do due to a lack of appropriate metrology equipment. If it is possible for them to use the results obtained here to compensate for the aberrations then it could save the significant expense of time and money which would be required to install a new set of windows into their experiment.

There are several problems in the measurement of the vacuum window not present in the above testing of a smaller refractive optic. The size of the window is larger with a diameter roughly equal to the travel range of the translation stages used, which prohibits a full measurement as the optic needs to be fully removed from the beam in order to measure system aberrations for hologram calibration. A second problem is there is no obvious window edge seen in the recovered phase maps due to the reflective mounting it is housed in. Rather than the edges giving random unwrapping errors as seen in the previous example we see an aberrated fringe pattern so we cannot easily define the boundary of the optic. Conveniently the area of the window being illuminated in the desired experiment is far from it's full extent, with only a 20 mm area needed allowing

us to provide a full measurement of the aberrations induced by the use of the window in transmission mode.

Figure 6.6 shows the interferogram for the most central subframe taken with a flat hologram being displayed on the SLM and comparing it to the interferogram given when using a tilted hologram of defocus. As can be seen there is a significant quantity of tip/tilt present giving a high fringe density and potentially swamping out smaller scale, higher order aberrations which is significantly reduced by removing the defocus and tilt terms to give a simple, easily measurable interference pattern. If the optic was illuminated fully and then demagnified onto the detector the fringe density would be sufficiently dense as to cause problems with recording a well sampled wavefront, and as such it is necessary to use a stitching method in the experimental measurement. In this example the only terms that need reducing holographically here are wavefront tilts, which could be removed through manually tilting the optic, as the defocus term is not large enough to cause problems in measurement. By programming a wavefront to remove the tilt however we know precisely the amount of slope removed from the interferogram and as such can give a precise full phase map of the optic to the user quickly and simply instead of needing to manually tamper with the interferometer. Furthermore prior to measuring the window we did not know the aberrations within it, and whilst in this particular case we did not need to null out higher order aberrations, the reprogrammable nature of the FLCOS device allows us to compare measurements taken using only a tilted hologram with those taken using the tilted hologram containing defocus as shall be seen later.

Figure 6.7 shows the full stitched map of 12 subframes corresponding to the recovered phase map of the interference of the vacuum window with a hologram of 6 waves RMS tilt in one direction and 1 wave RMS in the orthogonal one. It

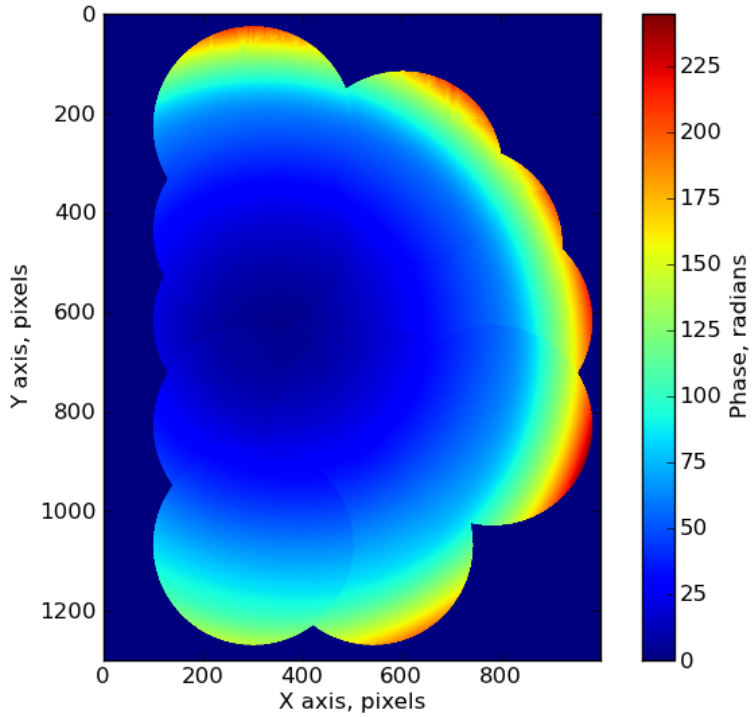


Figure 6.7: The full stitch for a measurement of the vacuum window using a hologram removing 6 waves RMS Zernike tip and 1 wave RMS tilt. As can be seen there are form deviations between numerous frames giving rise to many phase discontinuities and the appearance of stitching mismatches. Scale given in radians.

is readily seen that once more there are significant phase discontinuities between some of the overlap regions, which is more visually profound towards the centre of the image where a darker region appears to have a disjoined edge which could be the result of an incorrect scaling factor. This is not the only potential cause of the apparent form mismatch however and other sources of irregularity must be considered.

Firstly is the order in which the images are stitched together. This is more relevant than in the example of the transmissive refractive as the stitching order

here does not follow any set pattern but rather follows a path exploring the outside of the optic which was used to determine the number of frames needed and their locations. A result of this is that some images were compared against overlap regions which were not continuous nor contiguous in the stitching order. Errors in the shape of the overlap region will naturally give errors, and it would be desirable for the stitching to grow outward from a point rather than loop over a region and returning to it as the combined effect of the errors over multiple frames will give large mismatches between two sides of an image. So it would be desirable to use a less arbitrary stitching order when adding frames together.

A second potential source of error which needs to be considered is the ordering of frames as they are processed in the three step algorithm. As the images were obtained at once through the rapid display of three phase stepped images it is unknown which frame corresponds to the $+0$, $+2\pi/3$ and $+4\pi/3$ frame in the algorithm. Three frames means 6 permutations of this order and for half of these the polarity of the phase map will be flipped, i.e. three tilt phase maps will have a positive gradient, and three will have a negative gradient. Furthermore it is not guaranteed that the three phase maps of one polarity state will be entirely identical to each other so it would be ideal to use the same frame ordering for all images combined to remove any possible slight shape mismatch. Rather than taking the three interferograms separately however it is sufficient to take an image of the first frame, with no phase shift, and then take interferograms as normal allowing us to continue to limit vibrational error. The frame corresponding to the zero phase shift can be matched with the first frame taken and denoted Frame1, and so long as the other two are denoted so that we get a consistent phase map without inversion each of the 12 subframes will have the same ordering of phased stepped frames.

Figure 6.8 shows the effect these two considerations have on the reduction

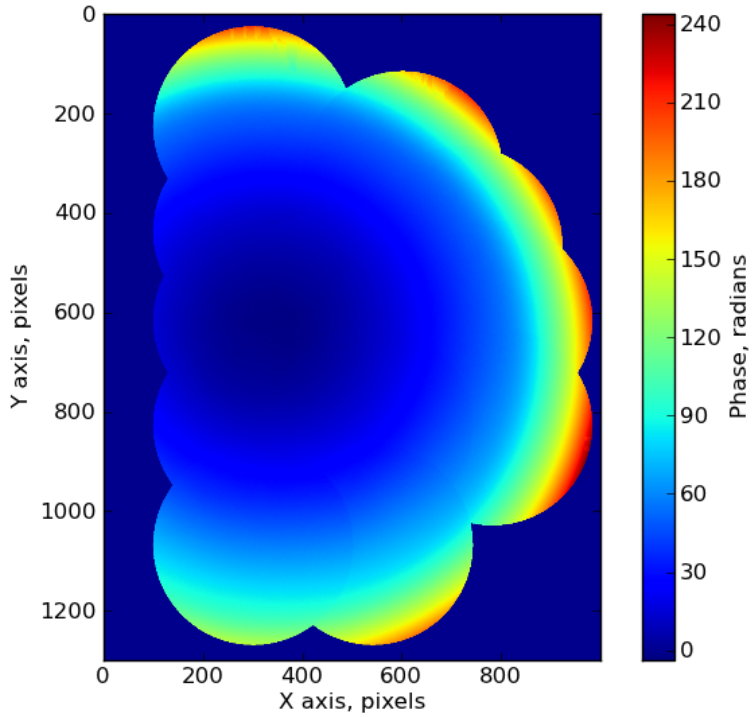


Figure 6.8: Full stitched image for the vacuum window using the same hologram as in figure 6.7 however with a refined stitching order and consistent subframe phase shifting analysis which shows a significant reduction in the quantity of phase discontinuities in the final composite. Scale in radians.

in phase discontinuities in the stitched composite. Note the central darkened region now has a significantly less disjointed boundary despite no change in scaling factor suggesting that the inaccuracies seen in figure 6.7 were unrelated to this variable. It is also seen that there are some slight unwrapping errors towards the bottom of the image however these can be ignored as they are only present in a region outside the area of interest.

As mentioned previously, the required measurement for this window need not be a complete phase profile as only a limited area is illuminated by the

experimental beam. We can therefore analyse in detail a circular region of this phase map and evaluate the remaining aberrations. Being able to consider a circular region allows us to perform another Zernike modal deconstruction which can separate out another reading of an orthogonal aberration set from non-standard residual errors. Considering a region of width 22 mm in diameter, sufficiently large for the experimental use of this element, we see an RMS aberration of 2.1722 waves RMS of which 2.1709 waves RMS is described using Zernike polynomials, predominantly tilt and defocus. This leaves a wavefront residual error of 0.068 waves RMS which consists of stitching errors, residual surface deformations and noise.

Figure 6.9 shows the residual wavefront shape for the stitched map after the removal of aberrations corresponding to Zernike modes. By looking at the vertical slice through the wavefront profile we see that the stitching discontinuities stand out from the background noise and are potentially the limiting factor in the measurement error. This complicates the detection of irregular surface features on the window which are a by-product of the manufacturing process of the window and can not be modelled using standard aberration polynomials.

Figure 6.10 shows the corrected composite map, similar to figure 6.9 however with a slight modification of the stitching process. Rather than just removing a piston term and then stitching two images together, this shows the effect on the corrected composite map when the average is taken over each detected data point. This serves to not only reduce the severity of the subframe boundary phase discontinuities but also reduces the level of noise in the measurement as can be noted by comparing the vertical slice graphs in figures 6.9 and 6.10. One effect of the phase discontinuities in the stitched phase map is to exhibit a reduction in the detected Zernike modes. For example when measuring the quantity of X-tilt in the composite we see values of 2.797 waves and 3.03 waves

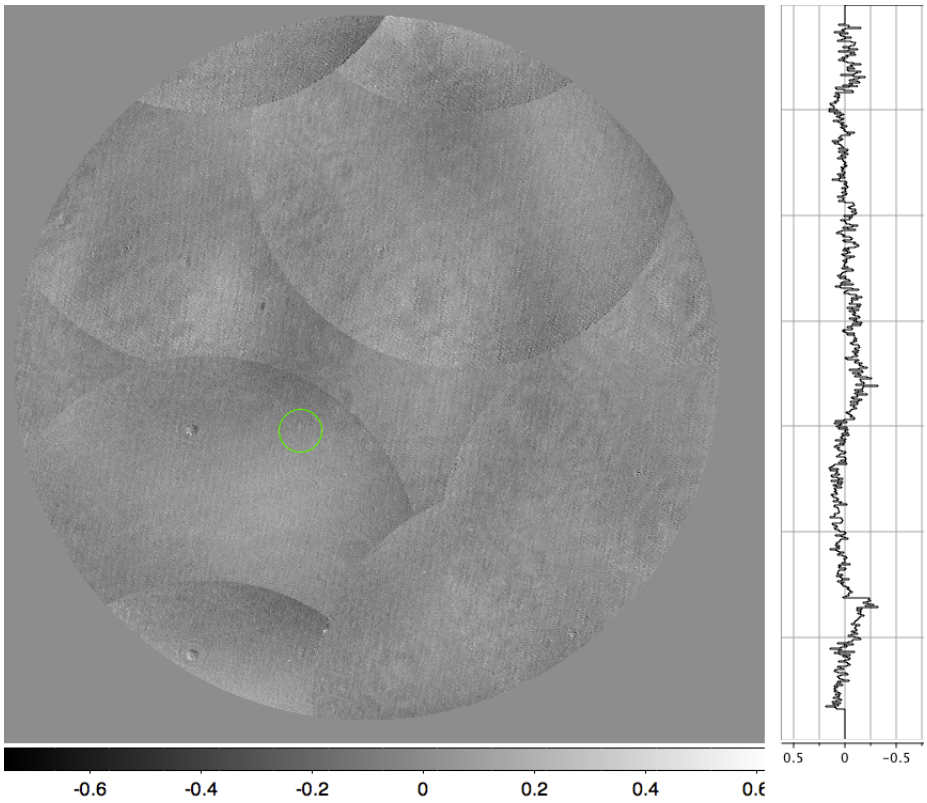


Figure 6.9: The residual wavefront of the stitched composite after removal of the first 20 Zernike terms with Z scale in waves. It is clear now that phase discontinuities from the stitching process will contribute largely to the RMS error seen however it is also possible to observe small amplitude variations from the optic itself, in the form of structure spanning multiple subapertures.

RMS respectively in the standard and averaging stitching case. Furthermore this averaging technique reduces the residual error from 0.068 waves RMS to 0.0535 waves RMS, which combined with the reduction in high frequency noise enhances the visibility of low amplitude aberrations in the optic that have structure extending over multiple subframes. The high frequency noise errors as determined through the post Zernike corrected flat frame corresponding to a hologram of purely the required tip and tilt needed to displace the first order spot is of the magnitude 0.052 waves RMS, in the same order of magnitude as

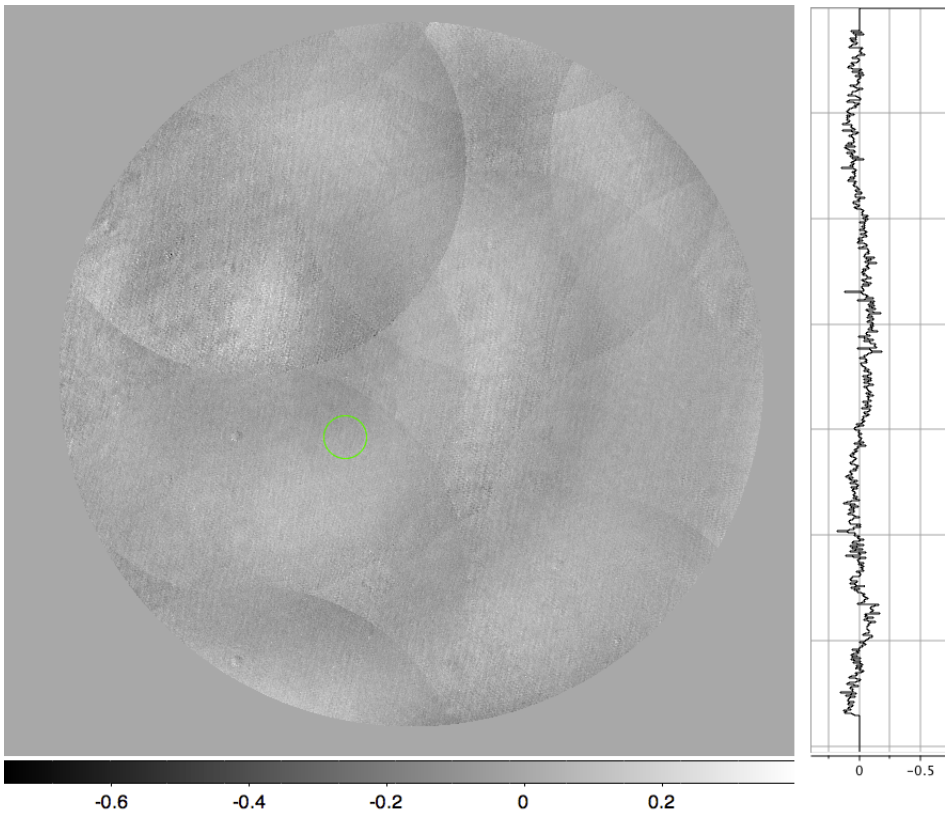


Figure 6.10: The Zernike corrected residual phase map for the composite after averaging over all subframes used with Z scale in waves. Doing so notably reduces the effect of phase discontinuities as can be seen in the vertical slice data and allows for better detection of the underlying optic irregularity.

the residual for the composite phase map indicating that there will be difficulty in further residual reductions as this noise will be present in all images due to the necessity of compensating for system aberrations.

6.5 Camera noise

As mentioned above, it is important to consider the effect on residual accuracy that noise will have on data. There will be two main sources of noise in this case; Poissonian noise from the random fluctuation of photon flux incident upon

the camera and a Gaussian noise component from read noise and thermal noise sources. For the inclusion of both of these into a simulated result the gain of the camera needs to be taken, as well as an estimation of the readout noise for the Gaussian component. By illuminating the camera with a flat field illumination we can obtain the mean signal value in counts, S_c, and the noise in counts from the standard deviation, N_c. It is also apparent that the mean and noise present in the signal count is related to the mean and noise in the number of electrons, S_e and N_e through:

$$S_e = gS_c \tag{6.1}$$

$$N_e = gN_c, \tag{6.2}$$

where g is the gain of the system. We also know that the total noise in the camera response is:

$$N_e^2 = R_e^2 + \sigma_e^2 + \sigma_f^2, \tag{6.3}$$

where R_e is the read noise of the system, σ_e the Poissonian photon noise, and σ_f is associated with deviations from uniformity of the flat field and fixed pattern noise. To minimise this effect data was taken from a central region of the flat field image. The variance of the electron noise is related to variations in photon flux, and therefore a Poissonian statistic so is equal to the electron signal, allowing us to substitute it into equation 6.3 and then by using equations 6.1 and 6.2 we get:

$$N_c^2 = \frac{S_c}{g} + R_c^2 + \sigma_f^2 \tag{6.4}$$

Plotting the mean against the variance for a range of intensity values allows for a linear fit of the form in 6.4, as in figure 6.11 and an estimation of the gain for the camera to be $22.88e^-$/adu and from looking at bias frames we can use this to calculate a value for the read noise of $10.61e^-$. These values are only

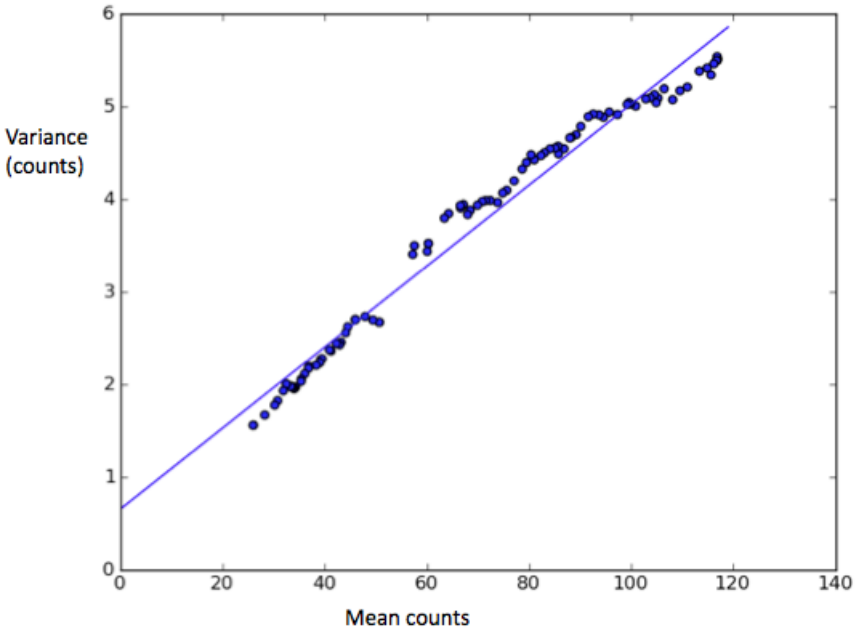


Figure 6.11: Plot of mean camera count noise against variance for a range of flat field intensities using camera settings as similar as possible to those used when obtaining interferograms. The slope of the graph is equal to 1/gain of the camera however a slight non linear response of counts with increasing intensity will result in a lowering in the estimated gain.

estimations of the gain and read noise however due to a number of complicating factors. Firstly the camera contains image analysis tools such as image sharpness and black level adjustment that were used in the acquiring of interferograms used in the above analysis. This complicates the determination of camera calibration and since it is desirable to use the same camera settings in the calibration as in the experimental data taking the effect of the sharpness parameter on gain and noise is unknown. Furthermore the stated value for gamma given in the camera documentation was not experimentally verified, and a non unity value giving a non linear intensity response could be the explanation for the slight curvature seen in figure 6.11, which would lead to an underestimated gain value.

Finally the source of flat field illumination may have a time dependant variation which could be eliminated through the normalisation of many frames of data with respect to the mean intensity measurement, before taking the mean and variance of each pixel through the resultant data cube in order to get a gain measurement. This approach was not used because a smeared out mean-variance distribution for each intensity level would have provided a less tight linear fit, and furthermore the normalisation of each frame would have given an artificial scaling to the measured gain reading, as whilst the mean intensity value would vary linearly to a scaling factor, the variance would change in proportion to the scaling squared.

Using the values given above for the camera gain and read noise, as well as the stated maximum quantum efficiency of the detector of 0.44 it was possible to simulate the effect of camera noise on the residual errors of the holograms produced. From this it was seen that the simulated residual Zernike corrected residual RMS value increases to 0.037 waves, which is dominated by the Poissonian noise and is mostly unaffected through variations in the camera read noise obtained through a series of of measurements.

It is possible to partially compensate for the high frequency noise introduced by the camera by filtering the Fourier transformed image to reduce the effect of these noisy fluctuations. Figure 6.12 shows the effect this has on further reducing the variations in surface measurements and further clarifying the presence of the surface features seen in the window giving a residual RMS of 0.044 waves. For the corresponding composite measured using without a tilted hologram a residual RMS measurement of 0.05 waves is seen which contains large phase discontinuities introduced by stitching and shows that the lower fringe density interferograms help reduce error in the overall measurement technique. By looking at the vertical slice through the composite we can see now that the

dominating source of variation is the vacuum window structural artefacts rather than from stitching errors or the effect of higher order holographic structure.

Whilst the dominant cause of the increased fringe density in this optic is tilt, we also observe that there is 1.4 waves Zernike defocus in the area of interest. We performed the same set of frame measurements using a hologram that not only contained the tilted wavefront terms but also an additional 1 wave rms of defocus to further reduce the fringe density as well as to observe the effect of producing a more complicated aberration and was seen to only marginally increase the residual wavefront error to 0.045 waves RMS.

6.6 Conclusions

Using a single hologram applied to all frames was significantly less time consuming on data processing than the method previously used in the first case, where a different hologram was created for each frame. Furthermore there was little to no difficulty encountered in unwrapping the phase profile obtained in comparison to the trouble found when using bespoke holograms per subframe. The limitation on only being able to measure half of the optic is unrelated to the FLCOS device and would have required a translation stage with a larger dynamic range, however this makes the obtaining of a second comparison measurement of the phase profile from another metrology highly problematic due to the complete lack of fiducial marks on the optic itself. Because of this we can only state that the use of a binary hologram from the SLM has been shown to reduce the residual wavefront error compared to measurements without it, and is capable of detecting phase irregularities of magnitude 0.044 waves RMS in a test piece. Smaller non-Zernike aberrations however would require the use of a more sophisticated camera however as noise from the currently implemented one is suspected to limit the accuracy to 0.037 waves RMS from simulated noisy

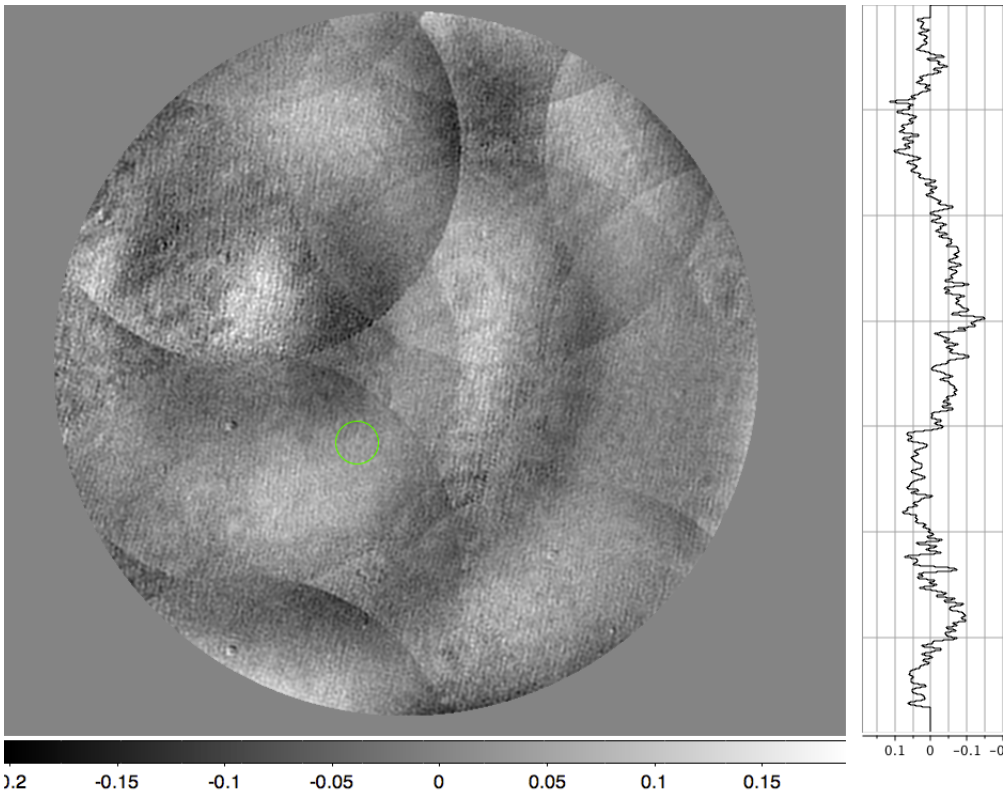


Figure 6.12: Here the residual phase map is filtered in the Fourier plane to remove very high order fluctuations associated with camera noise. The vertical slice through the map now shows a profile clearly dominated by the structure on the window itself rather than stitching discontinuities or holographic artefacts. Scale in waves

interferograms.

We compare this degree of accuracy to that exhibited in the the testing of large optics using a swing arm profilometry method [26] and see that they present a residual wavefront of 0.09 waves RMS compared to that seen thus far with the FLCOS device of 0.044 waves RMS at $\lambda = 633$ nm, suggesting that the potential exists for this technology to be used effectively in the field of metrology. In particular we note that swing arm profilometry is used as part of the initial surface testing procedure during the manufacture of EELT primary

mirror segments, and a further refinement of the work presented in this thesis could lead to the use of the FLCOS SLM technology in the manufacture of cutting edge systems. Despite this, it looks doubtful that the technology will, in its current state, manage to match the level of accuracy obtained through the use of holographic plates as through optimization of etching widths and depths they have been demonstrated to provide an accuracy in the testing of EELT mirror segment surfaces to less than one hundredth of a wave RMS [79][80][81]. Furthermore the magnitude of aberration that can be produced by this device is insufficient to null out the hundreds of waves of aberration observed in the outer E-ELT mirror segments.

Chapter 7

Conclusions and further work

Throughout this thesis we have evaluated the ability of an FLCOS SLM device to produce arbitrary wavefront profiles with the intention of using them within an interferometric surface measurement system.

Our simulated results, which were presented in chapter 4, indicate that a high degree of accuracy in the production of a desired wavefront term is possible with the main caveat of having an optimum sized spatial filter for each individual aberration produced. We predicted that the production of aberrations up to 9 waves RMS is possible with a residual wavefront error of less than 0.02 waves RMS with the dominant source of error being from the contamination of the wavefront by higher order harmonics being passed through the filter. Chapter 5 experimentally investigated this and found that it is possible to create these wavefronts with a residual error of roughly 0.06 waves RMS, which is significantly higher than that expected from simulated results however these values can partially be explained as being from camera noise. We saw in

the residual phase maps that the expected contamination of the wavefront from higher order terms is present, and also contributes to the residual error seen.

During the simulated and experimental discussion of results we compared two different binarization techniques. We proposed that, through adding randomized background noise to disrupt the highly structured PSF of the hologram in the Fourier plane, we can limit the debilitating effect of the higher order terms being passed through the filter and reduce the residual errors seen. We see both experimentally and through simulations that this is not the case. The increased background noise passed through the spatial filter gives a residual error exceeding that from the higher order structure, and therefore suggests the randomized binarization procedure is unsuitable for use. It is proposed however that since we used three step interferometry using only a single frame per step, it may be possible to increase the accuracy of the randomized holograms by collecting multiple frames, or by further increasing the number of phase steps and then averaging over them to create three subframes from any number of phase shifted interferograms.

During chapter 5 we also discussed the limitations imposed on the experimental method due to phase unwrapping considerations. We compared a local approach to a Fourier based global phase unwrapping technique and saw the introduction of both a non linear scaling factor in the Fourier unwrapped profiles, and also extra aberrations such as Zernike hexafoil therefore we decided to avoid using it, instead sticking to the local unwrapping method. Despite this there would be great benefit to spending more time either refining the unwrapper or obtaining a new technique entirely, as the local approach proved to be both slow and unreliable. The unwrapping of randomly binarized holograms in particular was highly problematic, and is one of the reasons why only limited consideration was given to the concept. For the FLCOS to be used productively in any

environment outside of laboratory research, a much faster phase unwrapping tool would be required to take advantage of the high speed offered by the SLM.

Chapter 6 presented the use of the FLCOS device in nulling aberrations measured in refractive optics and then evaluating the limitation on accuracy that can be achieved. We described how we use a stitching approach in the evaluation of these optics, as magnifying the beam to illuminate the whole element and then demagnifying it onto the camera would give a high fringe density. The degree of which would give rise to sampling issues where the observed width of the interferometric fringes was reduced to being only a couple of pixels.

We first used a transmissive element containing both trefoil and manufacturing artifacts, and for each subframe measured the aberration in the optic before creating a matching hologram with which we attempted to create a null interferogram. Once we had corrected for Zernike aberrations within the composite phase map, we saw a residual error of 0.05 waves RMS dominated by the presence of phase discontinuities introduced through the stitching process. Whilst a wavefront accuracy of 0.05 waves is not unappealing, we found that there were significant hinderances in obtaining the requisite measurements. The creation of holograms from the measured phase maps included the camera noise detected in the original interferogram which then led to noisy holograms, although in the future this could be mitigated by averaging over multiple interferograms. This caused significant problems in the unwrapping of the null interferograms, and with each subframe requiring the processing of four different phase maps, the time requirements are prohibitive for this method to be worth investigating further at this time.

The most promising result in this thesis came from the analysis of a vacuum window piece, which was provided due to it being aberrated to the extent where not only experimental use in a vacuum chamber was difficult, but also

the departmental Zygo interferometer could not produce a full phase map. By applying a single hologram to all subframes to reduce the fringe density through the removal of tilt terms and defocus, we managed to produce a stitched composite phase map with a residual error of 0.045 waves RMS, and by merely removing the tilted terms we saw a residual error of 0.044 waves RMS. Rather than being dominated through errors introduced within the stitching, we see that the residual wavefront is from low amplitude phase variations introduced by the window itself. This residual wavefront error is reduced from 0.05 waves RMS obtained through using a flat hologram, suggesting that reducing fringe density aids in detection of small amplitude aberrations.

For a comparison with real world applications, we have obtained results with an observed residual error of 0.044 waves RMS, at λ=633 nm, which is comparable to the maximum wavefront error requirement of the EELT primary mirror segments of 0.047 waves RMS and better than the observed accuracy of swing arm profilometry techniques of 0.0916 waves RMS. Despite this, there is substantial need for further characterisation of the device in order to reduce the residual errors seen.

The most vital step we feel is necessary to further evaluate the potential accuracy of the FLCOS device, is the need to improve the camera used in the experimental analysis. This is important due to the degree to which camera noise appears to limit the accuracy of all measured phase maps, from the use of a high gain in the detector. We encountered further difficulties due to the 8 bit well depth of the detector and therefore beam non-uniformities led to the saturation of central regions and difficulty in detection of signal in the outlying ones. Simulated results using the gain and read noise values obtained from the experimental camera show that we expect a residual wavefront error no less than 0.037 waves RMS, which is only slightly less than that seen from measurements

of the vacuum window where we know the dominating artefact is from the test piece.

In chapter 4 we highlighted the need for optimally sized spatial filter sizes, and showed the degree to which they can affect the residual RMS. We were unable to adequately investigate this experimentally but the addition of an electronically controllable aperture would allow for the accurate and precise measurement of the filter size, and thus could be used in experimentally reproducing the dependancy of the residual RMS error on filter size as seen in chapter 4.

Further work we suggest, includes the propagation and measurement of wavefronts that correspond to the surfaces of real optics, and the creation of holograms from theoretical surfaces before measuring their counterpart pieces. We also propose taking advantage of one of the most useful differences between the SLM and deformable mirrors by creating wavefronts with discontinuous features in. A refinement of the phase unwrapping process through the investigation of new techniques or improving the existing ones would be ideal, as the speed of unwrapping is currently a significant obstacle in the commercial viability of using the FLCOS SLM. A further investigation into optimising the stitching algorithm would also be important to consider, for example dithering the offset of one of the stitched subframes and examining the effect on residual RMS error that this has on the composite wavefront.

Binary FLCOS SLMs present a promising means by which we can alter the wavefront profile of a beam arbitrarily, with the ability to rapidly reprogram them to meet the needs of a fast changing wavefront shape. Further work needs to be done in narrowing down the residual errors associated with the production of the analogue holograms, however we have shown that it is capable of reaching an accuracy of less than one twentieth of a wave RMS and have high hopes for

future potential applications of the technology.

References

[1] MAA Neil, MJ Booth, and T Wilson. Dynamic wave-front generation for the characterization and testing of optical systems. *Optics Letters*, 23(23):1849–1851, December 1998.

[2] http://riflebarrels.com/articles/barrel_making/rifle_barrel_accurate.htm.

[3] D. Hounshell. *From the American System to Mass Production, 1800-1932: The Development of Manufacturing Technology in the United States*. ACLS Humanities E-Book. Johns Hopkins University Press, 1985.

[4] Curtis Baffes, Terry Mast, Jerry Nelson, Eric Ponslet, Vince Stephens, Larry Stepp, and Eric C. Williams. Primary mirror segmentation studies for the thirty meter telescope. pages 70180S–70180S–15, 2008.

[5] P. J. McCarthy. The Giant Magellan Telescope project. In P. Whitelock, M. Dennefeld, and B. Leibundgut, editors, *The Scientific Requirements for Extremely Large Telescopes*, volume 232 of *IAU Symposium*, pages 420–428, 2006.

[6] ESO European Extremely Large Telescope (E-ELT) Optical Design . http://esoads.eso.org/abs/2006IAUS..232..420M.

[7] http://www.eso.org/sci/facilities/eelt/docs/e-elt_constrproposal.pdf.

[8] J.J. Ramsden, D.M. Allen, D.J. Stephenson, J.R. Alcock, G.N. Peggs, G. Fuller, and G. Goch. The design and manufacture of biomedical surfaces. *CIRP Annals - Manufacturing Technology*, 56(2):687 – 711, 2007.

[9] *BS ISO 10110-8:2010 Optics and photonics. Preparation of drawings for optical elements and systems. Surface texture; roughness and waviness*. BSI, 2010.

[10] J. Gourlay, S. Samus, P. McOwan, D. G. Vass, I. Underwood, and M. Worboys. Real-time binary phase holograms on a reflective ferroelectric liquid-crystal spatial light modulator. *Appl. Opt.*, 33(35):8251–8254, 1994.

[11] N.A. Clark and S.T. Lagerwall. Submicrosecond bistable electro-optic switching in liquid crystals. *Applied Physics Letters*, 36(11):899–901, 1980.

[12] Bosanta R. Boruah, Gordon D. Love, and Mark A. A. Neil. Interferometry using binary holograms without high order diffraction effects. *Opt. Lett.*, 36(12):2357–2359, Jun 2011.

[13] J. Y. Wang and D. E. Silva. Wave-front interpretation with zernike polynomials. *Appl. Opt.*, 19(9):1510–1518, May 1980.

[14] J. Wyant and K. Creath. Basic wavefront aberration theory for optical metrology. In *Applied Optics and Optical Engineering*, volume XI, pages 1–53. Academic Press, N.Y, 1992.

[15] Robert J Noll. Zernike polynomials and atmospheric turbulence. *J. Opt. Soc. Am*, 66(3):207–211, 1976.

[16] Virendra N. Mahajan. Optical shop testing. In Daniel Malacara, editor, *Zernike Polynomials and Wavefront Effects*, chapter 13, pages 498–546. John Wiley and Sons., 3rd edition, 2007.

[17] http://upload.wikimedia.org/wikipedia/commons/3/3d/Zernike_polynomials2.png.

[18] Guang ming Dai and Virendra N. Mahajan. Zernike annular polynomials and atmospheric turbulence. *J. Opt. Soc. Am. A*, 24(1):139–155, Jan 2007.

[19] L.M Foucault. Description des procedees employes pour reconnaitre la configuration des surfaces optiques. *C.R.Acad.Sci(Paris)*, 47(958), 1858.

[20] J. L Rayces. Exact relation between wave aberration and ray aberration. *Opt. Acta 11*, 2:85–88, 1964.

[21] V. Ronchi. Le frange di combinazioni nello studio delle superficie e dei sistemi ottici [combination fringes in the study of surfaces and optical systems]. *Riv. Ottica Mecc. Precis.[J. Opt. Prec. Mech.]*, 2:9–35, 1923.

[22] Alberto Cordero-Dávila, Jorge González-García, Carlos Ignacio Robledo-Sánchez, and Irce Leal-Cabrera. Local and global surface errors evaluation using ronchi test, without both approximation and integration. *Appl. Opt.*, 50(24):4817–4823, Aug 2011.

[23] Daniel Malacara. *Optical shop testing*, volume 59. John Wiley & Sons, 2007.

[24] R. Platzeck and E. Gaviola. On the errors of testing and a new method for surveying optical surfaces and systems. *J. Opt. Soc. Am.*, 29(11):484–500, Nov 1939.

[25] J. Schmit, K. Creath, and James C Wyant. Optical shop testing. In Daniel Malacara, editor, *Zernike Polynomials and Wavefront Effects*, chapter 15, page 672. John Wiley and Sons., 3rd edition, 2007.

[26] Hongwei Jing, Christopher King, and David Walker. Simulation and validation of a prototype swing arm profilometer for measuring extremely large telescope mirror-segments. *Opt. Express*, 18(3):2036–2048, 2010.

[27] Ho-Soon Yang and David Walker. Development of prototype stylus prototype for large optics testing. *J. Opt. Soc. Korea*, 5(2):60–66, 2001.

[28] D.S. Anderson and J.H. Burge. Swing-arm profilometry of aspherics. *Proceedings of SPIE*, 2536(1):169–179, 1995.

[29] J. Tersoff and D. R. Hamann. Theory of the scanning tunneling microscope. *Phys. Rev. B*, 31:805–813, Jan 1985.

[30] Gerd Binnig, Calvin F Quate, and Ch Gerber. Atomic force microscope. *Physical review letters*, 56(9):930–933, 1986.

[31] S Alexander, L Hellemans, O Marti, J Schneir, V Ellings, P.K. Hansma, M Longmire, and J Gurley. An atomicresolution atomicforce microscope implemented using an optical lever. *Journal of Applied Physics*, 65(1):164–167, 1989.

[32] Q. Zhong, D. Inniss, K. Kjoller, and V.B. Elings. Fractured polymer/silica fiber surface studied by tapping mode atomic force microscopy. *Surface Science Letters*, 290(1â“2):L688 – L692, 1993.

[33] N.H. Thomson, M. Fritz, M. Radmacher, J.P. Cleveland, C.F. Schmidt, and P.K. Hansma. Protein tracking and detection of protein motion using atomic force microscopy. *Biophysical Journal*, 70(5):2421 – 2431, 1996.

[34] J. Simon. New noncontact devices for measuring small microdisplacements. *Appl. Opt.*, 9:2337, October 1970.

[35] R Brodmann and W Smilga. Evaluation of a commercial microtopography sensor. In *Proc.S SPIE 802*, 1987.

[36] James Pawley. *Handbook of Biological Confocal Microscopy*. Springer, 2006.

[37] A Ichihara, R Tanaami, K Isozaki, Y Sugiyama, K Kosugi, M Mikuriya, and I Umeda. High speed confocal fluoresent microscopy using a nipkow scanner with microlenses for 30d imaging of single fluorescent molecule in real time. *Bioimages*, 4(2):57–62, 1996.

[38] M. A. A. Neil, R. Juskaitis, and T. Wilson. Method of obtaining optical sectioning by using structured light in a conventional microscope. *Opt. Lett.*, 22(24):1905–1907, Dec 1997.

[39] J Hartmann. Bemerkungen uber den bau und die justirung von spektrographen. *Zt. Instrumentenkd*, 47, 1900.

[40] R.V. Shack and BC Platt. Production and use of a lenticular hartmann screen. *J. Opt. Soc. Am*, 61(5):656, 1971.

[41] B. Platt and R.V. Shack. Lenticular hartmann screen. *Opt. Sci. Center Newsl.(University of Arizona)*, 5:15–16, 1971.

[42] F Twyman and A Green. British patent 103,832, 1916.

[43] D. Gabor et al. A new microscopic principle. *Nature*, 161(4098):777–778, 1948.

[44] N Denisyuk Yuri. On the reflection of optical properties of an object in a wave field of light scattered by it. *Doklady Akademii Nauk SSSR*, 144(6):1275–1278, 1962.

[45] Emmet N. Leith and Juris Upatnieksu. Reconstructed wavefronts and communication theory. *J. Opt. Soc. Am.*, 52(10):1123–1128, Oct 1962.

[46] H. Kogelnik. Coupled wave theory for thick hologram gratings. *The Bell System Technical Journal, Vol. 48, no. 9, November 1969, pp. 2909-2947*, 48:2909–2947, 1969.

[47] BR Brown and A.W. Lohmann. Complex spatial filtering with binary masks. *Applied Optics*, 5(6):967–969, 1966.

[48] BR Brown and AW Lohmann. Computer-generated binary holograms. *IBM Journal of research and Development*, 13(2):160–168, 1969.

[49] J.P. Waters. Three-dimensional fourier-transform method for synthesizing binary holograms. *JOSA*, 58(9):1284–1287, 1968.

[50] V P Koronkevich et al. Laser thermochemical technology for synthesizing optical diffraction elements utilizing chromium films. *Soviet Journal of Quantum Electronics*, 15(4), 1985.

[51] J H Burge. Applications of computer-generated holograms for interferometric measurement of large aspheric optics. 2576, August 1995.

[52] J. Mauthner and W. Suida. Beiträge zur kenntnis des cholesterins. *Monatshefte für Chemie und verwandte Teile anderer Wissenschaften*, 24(3):175–194, 1903.

[53] G.W Gray. *Molecular structure and the properties of liquid crystals Molecular structure and the properties of liquid crystals Molecular Structure and the Properties of Liquid Crystals*. Academic Press, 1962.

[54] Gordon D. Love. Wave-front correction and production of zernike modes with a liquid-crystal spatial light modulator. *Appl. Opt.*, 36(7):1517–1520, 1997.

[55] Thu-Lan Kelly and Jesper Munch. Phase-aberration correction with dual liquid-crystal spatial light modulators. *Appl. Opt.*, 37(22):5184–5189, 1998.

[56] Enrique J. Fernández, Pedro M. Prieto, and Pablo Artal. Wave-aberration control with a liquid crystal on silicon (lcos) spatial phase modulator. *Opt. Express*, 17(13):11013–11025, 2009.

[57] Lijun Zhu, Pang-Chen Sun, Dirk-Uwe Bartsch, William R. Freeman, and Yeshaiahu Fainman. Adaptive control of a micromachined continuous-membrane deformable mirror for aberration compensation. *Appl. Opt.*, 38(1):168–176, Jan 1999.

[58] R.N. Shagam and J.C. Wyant. Optical frequency shifter for heterodyne interferometers using multiple rotating polarization retarders. *Appl. Opt*, 17:3034–3035, 1978.

[59] J. C. Wyant. Use of an ac heterodyne lateral shear interferometer with real-time wavefront correction systems. *Appl. Opt.*, 14(11):2622–2626, Nov 1975.

[60] K Creath. Phase measurement interferometry techniques. volume 26 of *Progress in Optics*, pages 349 –393. Elsevier, 1988.

[61] P Carré. Installation et utilisation du comparateur photoélectrique et interférentiel du bureau international des poids et mesures. *Metrologia*, 2(1):13, 1966.

[62] Matt Novak, James Millerd, Neal Brock, Michael North-Morris, John Hayes, and James Wyant. Analysis of a micropolarizer array-based simultaneous phase-shifting interferometer. *Appl. Opt.*, 44(32):6861–6868, Nov 2005.

[63] Gustavo Rodriguez-Zurita, Cruz Meneses-Fabian, Noel-Ivan Toto-Arellano, José F. Vázquez-Castillo, and Carlos Robledo-Sánchez. One-shot phase-shifting phase-grating interferometry with modulation of polarization: case of four interferograms. *Opt. Express*, 16(11):7806–7817, May 2008.

[64] R.V. Shack and G.W. Hopkins. The shack interferometer. *Optical Engineering*, 18(2):182226–182226, 1979.

[65] Dong Liu, Yongying Yang, Chao Tian, Yongjie Luo, and Lin Wang. Practical methods for retrace error correction in nonnull aspheric testing. *Opt. Express*, 17(9):7025–7035, Apr 2009.

[66] Yingchao Xu, Xin Zhang, and Ying Zhang. Using computer-generated holograms to test cubic surface. *Optics Communications*, 282(12):2327–2331, June 2009.

[67] Akira Ono and James C Wyant. Aspherical mirror testing using a cgh with small errors. *Applied Optics*, 24(4):560–563, February 1985.

[68] Feenix Pan, J. Burge, and D. Anderson. Measurement of aspherical surfaces using test plate and computer generated hologram (cgh). In *Optical Fabrication and Testing*, page OWD7. Optical Society of America, 2002.

[69] J H Burge. Efficient testing of off-axis aspheres with test plates and computer generated holograms. 3782, July 1999.

[70] Jacek Kacperski and Malgorzata Kujawinska. Active, lcos based laser interferometer for microelements studies. *Optics Express*, 14(21), October 2006.

[71] Xianyu Su and Lian Xue. Phase unwrapping algorithm based on fringe frequency analysis in fourier-transform profilometry. *Optical Engineering*, 40(4):637–643, 2001.

[72] Igor Lyuboshenko and Henri Maître. Phase unwrapping for interferometric synthetic aperture radar by use of helmholtz equation eigenfunctions and the first green's identity. *JOSA A*, 16(2):378–395, 1999.

[73] Dennis C Ghiglia and Louis A Romero. Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods. *JOSA A*, 11(1):107–117, 1994.

[74] Marvin A. Schofield and Yimei Zhu. Fast phase unwrapping algorithm for interferometric applications. *Opt. Lett.*, 28(14):1194–1196, Jul 2003.

[75] Weng W. Chow and George N. Lawrence. Method for subaperture testing interferogram reduction. *Opt. Lett.*, 8(9):468–470, Sep 1983.

[76] Stephen C. Jensen, Weng W. Chow, and George N. Lawrence. Subaperture testing approaches: a comparison. *Appl. Opt.*, 23(5):740–745, Mar 1984.

[77] Mingyi Chen, Weiming Cheng, and Cun-Wu W Wang. Multiaperture overlap-scanning technique for large-aperture test. In *San Diego-DL tentative*, pages 626–635. International Society for Optics and Photonics, 1992.

[78] Shanyong Chen, Shengyi Li, and Yifan Dai. Iterative algorithm for subaperture stitching interferometry for general surfaces. *JOSA A*, 22(9):1929–1936, 2005.

[79] Ping Zhou and James H. Burge. Optimal design of computer-generated holograms to minimize sensitivity to fabrication errors. *Opt. Express*, 15(23):15410–15417, 2007.

[80] Feenix Y. Pan and Jim Burge. Efficient testing of segmented aspherical mirrors by use of reference plate and computer-generated holograms. i. theory and system optimization. *Appl. Opt.*, 43(28):5303–5312, 2004.

[81] Feenix Y. Pan, Jim Burge, Dave Anderson, and Alexander Poleshchuk. Efficient testing of segmented aspherical mirrors by use of a reference plate and computer-generated holograms. ii. case study, error analysis, and experimental validation. *Appl. Opt.*, 43(28):5313–5322, Oct 2004.