Phase Optimised General Error Diffusion for Diffractive Optical Component Design

by

Jamie L. Ramsey

A Thesis

presented to the University of Strathclyde

in fulfilment of the requirements for the degree Doctor of Philosophy

in the

Department of Electronic and Electrical Engineering

September 2013

Declaration

This Thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Acknowledgements

I would like to thank Professor Trevor Hall and Professor Ivan Andonovic for all their help and guidance throughout the research. It had been a long road and I would not have made it without them.

I would also like to thank Nic Olivier and my colleagues at ELCAN Optical Technologies who have supported me during my studies. A special thank you is extended to Jason Taylor for his assistance in writing a program to convert the data generated in Matlab into the GDSII format and Maxine Poirier and Jessica Zhang at the CPFC for their assistance in overseeing the fabrication of the DOEs for testing purposes.

Lastly I would like to thank my parents, Sharon and James Ramsey and my brother Bill for their love and encouragement.

Abstract

Algorithms for developing Diffractive Optical Elements (DOEs) are improved to achieve real time holograms capable of switching at rates of 25frames/second or greater. A Phase Optimised General Error Diffusion (POGED) algorithm optimised for quality and speed of generation of diffractive elements is the main contribution of the research. Compared to Simulated Annealing algorithms, a fourfold improvement in the speed of generation is achieved. The algorithm is further enhanced to operate in the Fresnel region with high diffraction efficiency and Signal-to-Noise Ratio (SNR).

A number of different target reconstructions are simulated to determine validity and performance of the algorithm. Diffractive optical elements are fabricated to verify performance and a free space optical beam steering application is defined to further validate a DOE generated by POGED. The performance of the diffractive optical elements is proven through the design and characterisation of a free space optical interconnect amenable to harnessing the fast switching speeds of liquid crystal spatial light modulators.

Table of Content

Table of Figure

Table of Tables

Abbreviations

DOE	Diffractive Optical Element
CGH	Computer Generated Hologram
POGED	Phase Optimised General Error Diffusion
SA	Simulated Annealing
DBS	Direct Binary Search
GS	Gerchberg–Saxton
E-beam	Electron Beam
RIE	Reactive Ion Etching
SEM	Scanning Electron Microscope
TEM	Transmission Electron Microscope
FSOI	Free Space Optical Interconnect
MEMS	Microelectromechanical systems
PMMA	Poly Methyl Methacrylate
IFTA	Iterative Fourier Transform Algorithm
SAG	Sagittal
TAN	Tangential

Chapter 1

Introduction

1.1 Motivation

Communication networks are continually evolving to meet the growing demand from more sophisticated user communities consuming an ever increasing range of rich services e.g. high definition movies, multi-player gaming [1]. Consequently, there is a pressing, near term need for the deployment of higher capacity access solutions, illustrated in Figure 1.1, the segment of the network which represents the major restriction to extending high bandwidth services to end users.

Free space reconfigurable optical interconnects [3-12] are a class of transparent optical switches used within an optical network to improve the scale of delivery of services. Transparent optical switches eliminate the optical-electronic-optical (OEO) conversion overhead executing on the routing function directly in the optical domain, in so doing extending the ultimate data transmission rate [2]. There are a number of implementation options within the hierarchy of the network for transparent optical switches [1] and indeed a number of switch types offering different set of performance parameters have been developed.

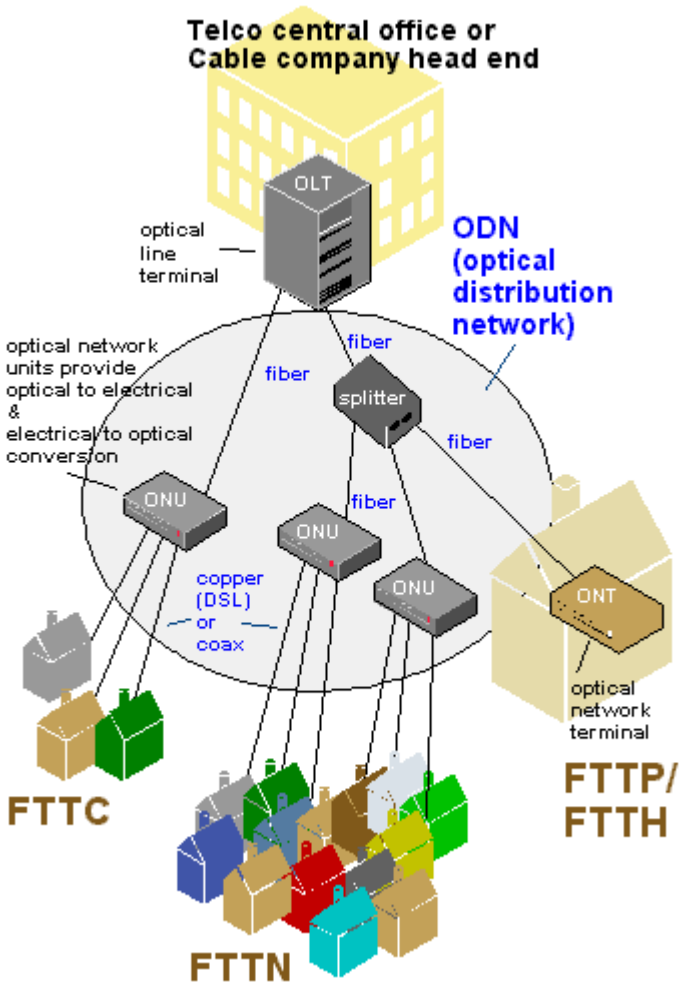


Figure 1.1 A schematic of a representative optical network hierarchy [2].

One possible implementation utilises the fast switching speeds of liquid crystal Spatial Light Modulators (SLMs) [9,12]. Currently the fastest SLMs employ ferro-electric liquid crystal on a silicon backplane, capable of switching within ~450 μs [13]. The Thesis confines the research to consider network architectures in which a SLM is used as a means to implement routing of optical signals. More specifically, the generation of diffractive optical elements is core to this evolution option and one of the limiting characteristics of SLM-inspired optical diffraction is the speed of establishing a suitable diffraction pattern. For example, algorithmic frameworks that

improve the speed of generation to rates in the micro second time scales or less allows the implementation of video frame-rate beam steering applications.

A free space optical interconnect was proposed by Marchand et al [85] to increase the scale of delivery of services. This system utilises two sets of meso-lens in the input and output planes, illustrated in Figure 1.2. To improve upon this design, a central Fourier Transform Lens is proposed to minimise off-axis aberrations, as described in Chapter 9. Further to assist in beam steering and shaping of the optical signal, two sets of reconfigurable DOEs are proposed, located in the focal planes of the meso-lenses at the input and output. With the use of reconfigurable DOEs, an algorithm must be developed that will achieve a high quality output signal and be able to generate a new DOE in the order of micro seconds or less.

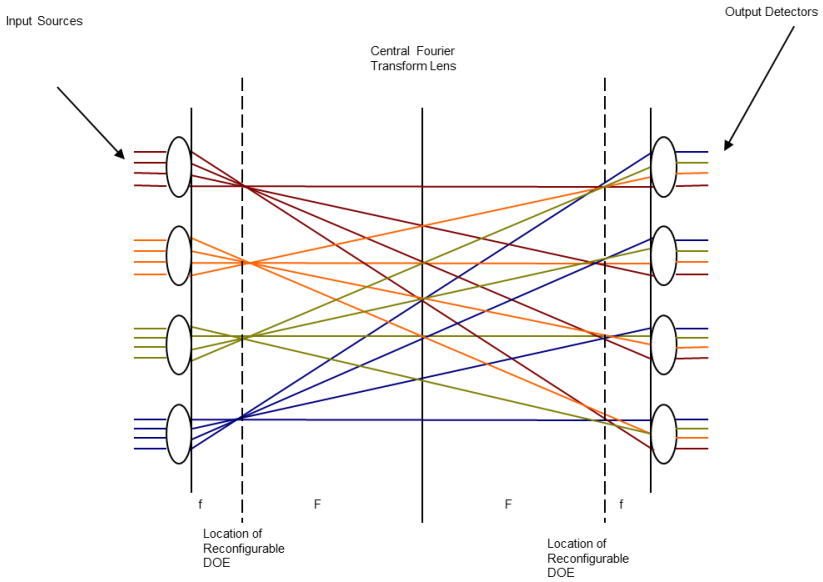


Figure 1.2 Example of a free space optical interconnect with the potential location for a central Fourier transform lens and the reconfigurable DOE's at the input and output.

Since the advent of computers, algorithms that underpin the modulation of the phase of the spatial spectrum of an optical image have been in development and applied to a variety of applications e.g. creation of a corrector lens for the Hubble telescope [14-17], beam steering [6], beam shaping [18-22] and image projection [23]. A large body of research has also focused on the development of algorithms that enable reconfigurable and accurate diffraction patterns to be implemented. Example algorithms include the Gerchberg-Saxton (GS) [24], Genetic Algorithms (GA) [25], the Iterative Fourier Transform Algorithm (IFTA) [26-27], Direct Binary Search (DBS) [28-29], Simulated Annealing (SA) methods [30-31] Minimum average error diffusion [32-33]. Each algorithm offers a mix of strengths and weaknesses, most often a trade-off weighing computational speed and accuracy versus stagnation - finding the local minimum instead of an objective function.

The Phase Optimised General Error Diffusion (POGED) algorithm is proposed as a novel formulation targeted at improving the speed of execution whilst maintaining the quality of the diffractive optical element. It has, as its basis, the minimisation of a cost function and provides a performance comparable to the SA algorithm but within a shorter time period; it yields a high quality design at computational times approaching real time where real time refers to calculations occurring in the order of milliseconds.

POGED is a modified version of the error diffusion algorithm developed originally for printing [34]. The error diffusion method is a form of half-toning where

quantisation errors are spread to yet to be processed neighbouring pixels. The diffusion of errors is accomplished through the use of a filter defined by its spectral transfer function (the ‘filter function’) or equivalently its convolutional kernel (the ‘mask function’) that, whenever a quantised pixel state is updated, diffuses the quantisation error to the neighbouring pixel.

1.2 Overview

The Thesis is divided into two main parts. The first part reviews the theoretical foundations that underpin the research, including the fundamentals of scalar diffraction theory. The core of the research - the Phase Optimised General Error Diffusion algorithm - targeted at the design of phase only diffractive optical elements of specified far-field or near-field field magnitude – the definition of far-field and near field are described in Section 2.4 - is then introduced and its structure defined. The second half presents an overview of the fabrication of the diffractive optical elements including their optical and structural characterisation and a free space optical interconnect is selected as a representative application example of an optical architecture making use of reconfigurable diffractive optical elements. The application is used in a performance comparison of POGED with the well-established Simulated Annealing approach.

Chapter 2 reviews of some applications employing diffractive optical elements. A historical background is then presented along with a review of scalar diffraction theory, the different grating types and the artefacts introduced by the naïve

application of the discrete Fourier Transform to numerically evaluate the continuous Fourier Transform, embedded within the diffraction integral.

Chapter 3 presents, for the purpose of positioning the research, relevant reported algorithms used to generate diffractive optical elements; specifically two of the more established algorithms, Simulated Annealing (SA) and the Direct Binary Search (DBS) form the focus of a comparison of performance.

The theory and performance evaluated through simulation for the Phase Optimised General Error Diffusion (POGED) algorithm applied to problems with a target reconstruction specified in the far-field region (Fraunhofer), are presented in Chapter 4. The Chapter begins with an introduction to the method and proceeds with its mathematical description. The filter function and its effects are evaluated. The Chapter concludes with an analysis of performance results for binary, quaternary and multi-level grating structures.

Chapter 5 begins with a discussion of the necessity to modify POGED for target reconstruction applications in the near-field region (Fresnel). Arguments are developed for the selection of the angular spectrum method in preference to alternative integral representations. Results for multi-level gratings and a discussion on the effects of different filters are presented.

Chapter 6, the beginning of the second part of the Thesis, reviews the photolithographic fabrication techniques used routinely in the manufacture of static

gratings. Two techniques are discussed in particular; electron beam lithography, a maskless technique; and step and repeat lithography, a technique where the pattern defined by a smaller mask, is repeated across the wafer.

Chapter 7 presents a review of techniques used to evaluate the quality of fabricated grating structures along with a discussion of results obtained.

The design of a Free Space Optical Interconnect (FSOI) architecture harnessing diffractive optical elements is detailed in Chapter 8, the optimum implementation being through ferroelectric liquid crystal spatial light modulators. However, due to limitations of the software used to define the optics for the switch, beam deflection mirrors emulate the diffractive element.

Chapter 9 provides conclusions arising out of the research and presents future work.

1.3 Contributions

The principal contribution of the research is the design and development of the Phase Optimised General Diffusion (POGED) algorithm for the generation of diffractive optical elements for beam steering applications. The core contribution can be further segmented into;

- The development of POGED for binary to multiphase diffractive optical elements
- Improved filtering through the use of Dykstra's Alternating Projection algorithm

- Fresnel POGED using the angular spectrum method to eliminate high frequency oscillations
- Fabrication and test of DOE's generated using POGED
- The emulation of a free space optical interconnect application with beam steering diffractive optical elements generated by POGED

1.4 Publications

1 J. L. Ramsey, T. J. Hall, "Improved methods for generating dynamic computer generated holograms for realising adaptive optical cross connects", Conference on Lasers and Electro-Optics/Europe, Munich, Germany, June 12-17, 2005.

2 J. L. Ramsey, S. Bidnyk, and T. J. Hall "POGED: A fast alternative to Computer Generated holograms design by simulated annealing," Diffractive Optics, Warsaw, Poland Sept. 3-7, 2005.

3 J. L. Ramsey, J. LaPointe, T. J. Hall, "Comparison of Simulated and Fabricated Computer Generated Holograms Generated using Phase Optimised General Error Diffusion", Conference on Digital Holography, Vancouver, BC, Canada, June 18-20, 2007.

4 J. L. Ramsey, W. Pijitrojana and T. J. Hall, "Novel Lens Design for Free-Space Optical Interconnects", Proc. SPIE **6342**, 634218 (2007).

5 R. Millet, J. Ramsey, P. Brock, J. Nkanta, H. Schriemer, T. J. Hall, S. Bidnyk, " Synthesis of Planer Reflective Gratings for Silicon Interconnects", Advances in Optical Technologies, Special Issue on Silicon Phonics, 2008, Article ID 315724, pp1-8 June 2008.

6 J. L Ramsey, V.P. Sivokon, T.J. Hall, I. Andonovic and C. Michie "Fresnel Based Phase Optimised General Error Diffusion Algorithm for Optical Beam Shaping", Proc. SPIE **7652**, (2010).

Chapter 2

Fundamentals of Diffractive Optical Elements

2.1 Introduction

Interference, essential to holography, has been the subject of considerable research over many years. The field has moved from generating holograms on photographic plates to complex holograms generated using computers and lithography. The diffractive optical elements (DOEs) now being generated are in use in a number of applications, stimulating new technologies to be developed, increasing production capabilities in a number of key industries and seeding other potential valuable developments. The massive impact on science research and engineering has stimulated new methods that improve the quality of, or produce real time operation of families of DOEs.

2.2 Background

The Section reviews some of the basic principles underpinning scalar diffraction theory and the beginnings of computer generated holography. A brief description of the differences between holography and photography is given.

2.2.1 Traditional Holography

In conventional photography, the process only captures intensity variations and therefore only the amplitude of the image is recorded; thus any phase information from the object is lost. Phase information allows the ability to create 3D images. This limitation stimulated the development of holography - stemming from the Greek term “holo” meaning total - a technique for recording both amplitude and phase. A hologram encodes both the amplitude and phase captured through optical interference. This process is outlined in the following references [35-40].

2.2.2 Computer Generated Holography (CGH)

As the object wavefront is required for the recording process and as such, holograms are impacted by a number of factors such as vibration [36]. For Computer Generated Holograms (CGHs), the object wavefront is now described mathematically, the interference pattern being synthesised [38-41, 47-48]. The advantages of CGHs are significant, the major capability being the easy production of complex wavefronts not possible with the traditional approach. Another key aspect centres on the recording process; CGH holograms are produced through lithographic methods not subject to problems that plague the traditional process such as coherent illumination artefacts and air turbulence.

A number of methods for fabricating and defining CGHs have been developed, including the Detour Phase Hologram [42], the Kinoform [43] and the Binary Hologram [44-47].

2.2.2.1 Detour Phase Hologram (DPH)

The Detour Phase Hologram (DPH) [42] is one of the oldest and most widely investigated techniques in the 1990's, encoding the wavefront phase using a "detour phase" for the diffracting grating. DPH encoding is performed using the shift theorem of the Fourier Transform (FT) which states that lateral shift of the FT induces a concomitant phase shift of the diffraction orders. The resultant hologram is divided into a number of equally spaced square cells forming a matrix (Figure 2.1). Based on the amplitude and phase values of the hologram, a rectangular shape of varying size is placed in each square with respect to the center of the cell.

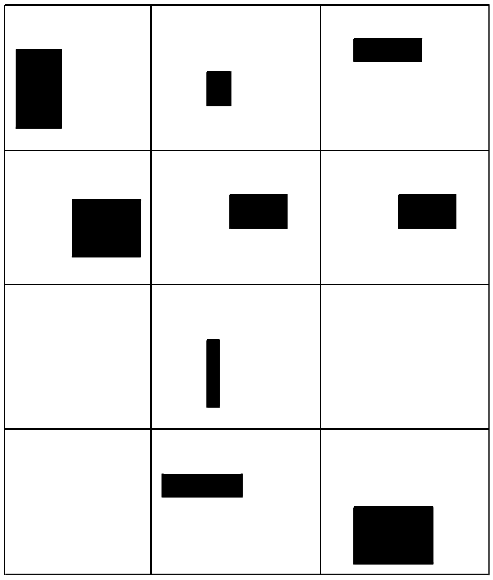


Figure 2.1 Example of a Detour Phase Hologram [40].

The amplitude and phase of the hologram are encoded but there are limitations stemming from the dislocation of adjacent apertures that store the complex wavefront, resulting in approximations to the amplitude and phase of the wavefront

owing to the quantization process. The method is also not able to treat large phase variations, since the maximum phase change that can be represented by the apertures is limited to 2π [42-43].

2.2.2.2 Kinoforms

Kinoforms are phase-shaping elements that transform a wavefront into a desired shape with constant amplitude [43] and are used in a number of devices; one example is a Fresnel lens used in colour correction of infrared (IR) lenses (Figure 2.2).

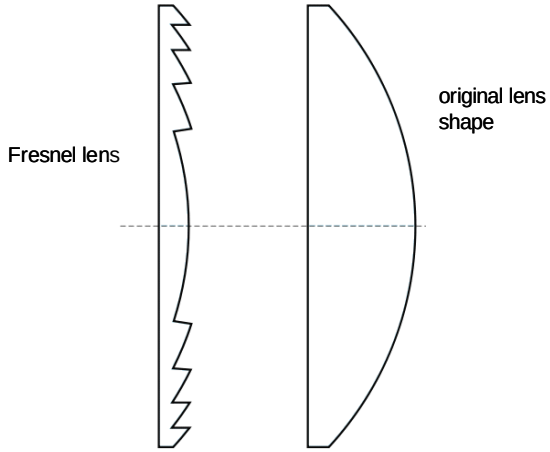


Figure 2.2 An example of a Kinoform (surface relief). (a) a Fresnel lens and b) the original lens.

2.2.2.3 Binary CGHs

Binary holograms are a special case of the Kinoform, introduced to solve the two fundamental problems of DPHs [44]. A binary hologram is not able to capture large

phase variations and only approximates the phase and amplitude of the field it encodes (Figure 2.3).

Figure 2.3 Example of a two level binary hologram.

Patterns in binary CHGs could be interpreted as interferences fringes in traditional holograms.

2.3 Technology Review

In parallel with improvements in computing power and semiconductor manufacturing, methods used to create Diffractive Optical Elements (DOEs) have also evolved [39-40, 67-68]. DOEs are now being used in a number of different sectors ranging from the commercial to the aerospace industry [14-17, 51-53]. Industries routinely use diffusers [49] to control the divergence angle of coherent or incoherent laser sources as well as beam shaping for lasers in micro machining [22] or LED sources [20] and in beam steering [6] for free space interconnects or in pattern generation [49].

2.3.1 Engineered Diffusers/Homogenisers

Diffusers create light uniform in intensity and emitted at a specific angle [49]. There exist a number of ways of making optical diffusers including ground glass, holographic, flashed opal, and greyed glass diffusers.

One mass application of a diffuser is in photography, most commonly referred to as 'flash diffusers' that spread the illumination from the flash lamp of a camera. The device makes the light appear to originate from a diffuse source rather than a point source, essentially spreading the light to eliminate any harsh shadows. Figure 2.4 shows an example of an engineered diffuser which homogenises a LED source to produce a circular pattern.

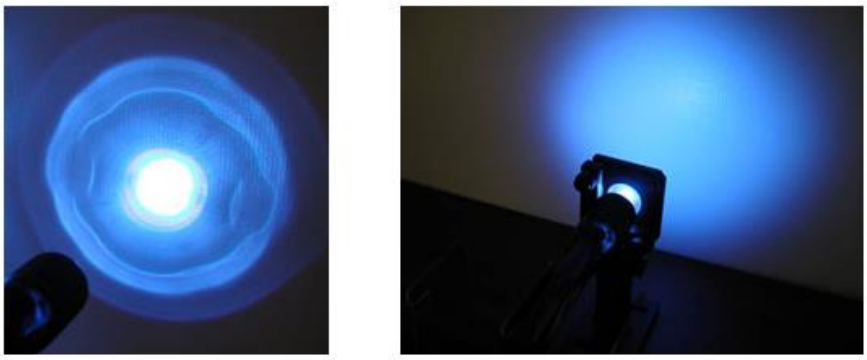


Figure 2.4 Engineered LED diffuser [49].

2.3.2 Beam Shapers

Optical beam shaping is widely used in the consumer, telecommunications and industrial markets where Gaussian beams emitted by a laser are transformed into a flat top beam [18] (Figure 2.5). The elliptical emission pattern of laser diodes requires to be reshaped into a more circular symmetry for the efficient coupling into a circularly symmetric optical fibre. New LED back lit flat screen televisions use beam shapers to guarantee uniformity of the emitted light and to remove ellipticity [19].

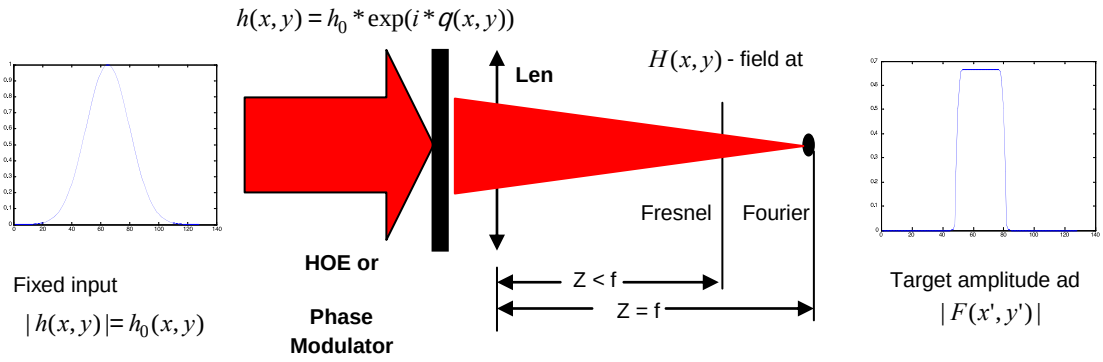


Figure 2.5 Example of how a diffractive/holographic element transforms a Gaussian into a flat-top beam.

Optical beam shaping for the redistribution of optical power into a specified profile is used in the laser micromachining industry to increase accuracy, speed and output power. DOEs can be used to shape the beam correctly over a refractive axicon [50] and prove to be more cost effective (Figure 2.6)[22].

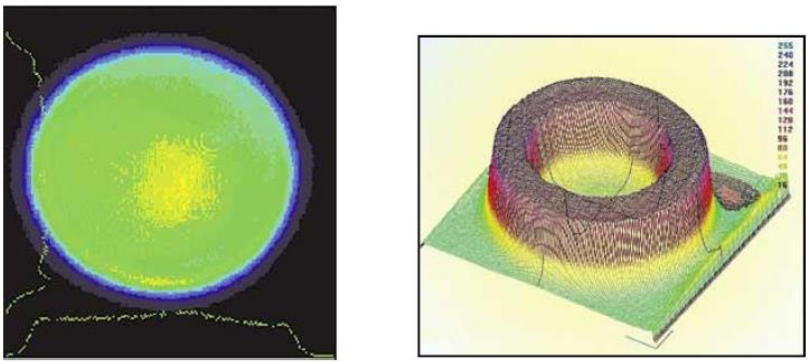


Figure 2.6 An example of a Gaussian beam shaped into a circular beam used in the micro machining industry.

2.3.3 Array Generators

An array generator consists of a single DOE that splits a beam into a number of 'beamlets' at pre-defined angles or into an array of elements of varying size [49]. In Figure 2.7 individual beamlets are imaged at infinity as a collection of bright points.

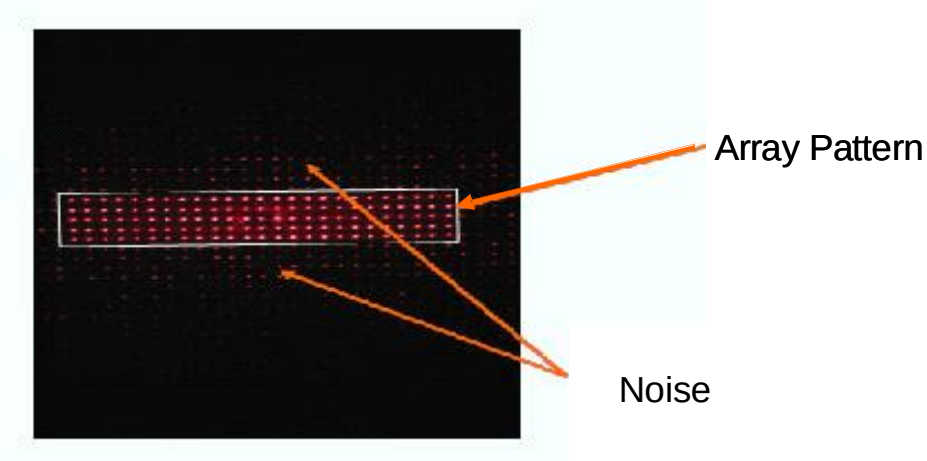


Figure 2.7 An example of an array generator splitting a beam into a rectangular pattern [49].

Array generators are useful when traditional beam splitters are inappropriate e.g. for a very large number of beams.

2.3.4 Phase Retrieval for Optical Metrology

Phase retrieval is used to correct the aberrations in the Hubble Space Telescope [14-17] and test mirrors for the James Webb Telescope [51]. Many different algorithms are used to execute this function, the most noteworthy the Gradient Search Technique [52]. Phase retrieval operates by measuring the point spread function

recovering the phase then transformed using Zernike polynomials [53] to represent wavefront deformation at optical surfaces. With this information, it is then possible to construct a correcting lens.

Phase retrieval has the potential to become very attractive for testing aspheric lenses since current testing protocols use an interferometer and are thus subjective because positioning is critical. Some techniques also require computer generated null holograms which are expensive and time consuming to manufacture.

2.4 Scalar Diffraction Theory

DOEs are usually analysed using Scalar Diffraction Theory [35, 38] which assumes light propagates substantially along the optical axis in a linear isotropic, homogenous and non-dispersive medium. Thus all electric and magnetic field components behave the same and can thus be described by scalar wave equations. For scalar diffraction theory to be valid, the size of the diffracting features must be larger than the wavelength of the incident beam and the observation plane is assumed to be located at a substantially 'far distance' [35, 38]. The far field (Fourier) region is defined at a distance significantly greater than W^2/λ, where W is largest dimension in the aperture and λ is the wavelength of light. The near field region (Fresnel), is defined as the region where scalar diffraction is only valid z is less than W^2/λ.

One condition core to the description of scalar diffraction theory is Kirchhoff Boundary conditions [35, 38]. The conditions relate to an opaque screen with a small aperture, labeled Σ (Figure 2.8) and the assumptions made on an incident

wavefront are: across Σ , the function $U_0(u,v)$ and its derivative are the same as if no screen is present and $U_0(u,v)$, and its derivative are zero in the geometric shadow of the screen.

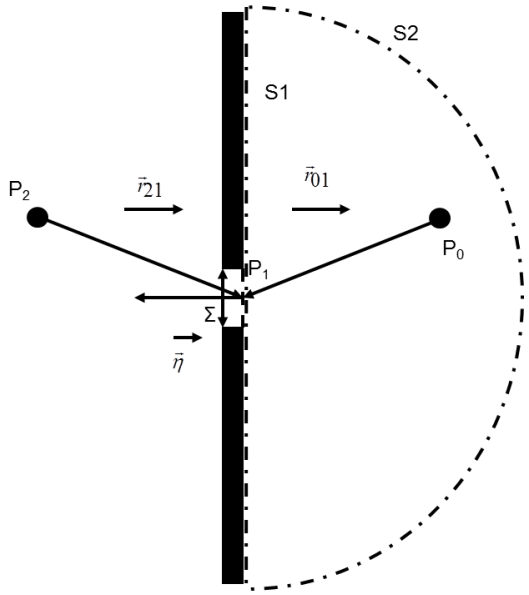


Figure 2.8 Illustration of a point-source illuminating a screen, a Kirchhoff boundary value condition.

Using the Rayleigh-Sommerfeld formulation [38] for scalar diffraction, a complex diffracted field can be described as:

$$U_1(x,y) = \frac{1}{i\lambda}\iint_{\Sigma} U_0(u,v)\frac{\exp(ikr)}{r}\cos(\vec{\eta},\vec{r})dudv \qquad \textbf{(2.1)}$$

where $U_0(u,v)$ is the incident field, $U_1(x,y)$ is the complex diffracted field at $Z = z1$, $k = 2\pi/\lambda$, λ is the wavelength of light, $cos\left(\overrightarrow{\eta,\vec{r}}\right)$ is the direction cosines between $\vec{\eta}$ and $\vec{r}$, and Σ is the aperture over which everything is described (Figure 2.9).

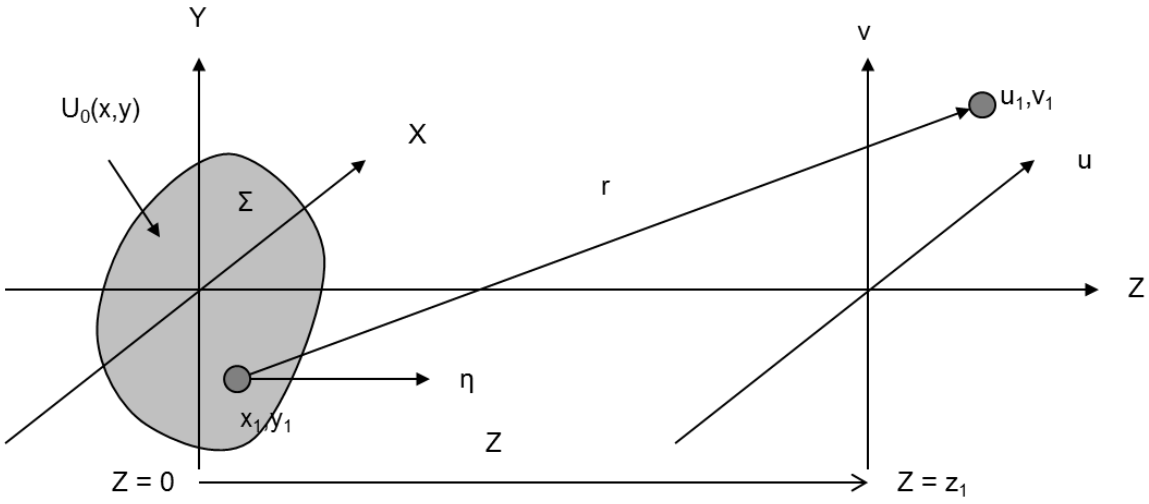


Figure 2.9 Rayleigh-Somerfield diffraction formulation [35, 38].

2.4.1 Fresnel Diffraction

To simplify the diffraction integral, the distance *r* between $Z = 0$ and $Z = z_1$ can be approximated by the binomial theorem where *r* is:

$$r = \sqrt{(x-u)^2 + (y-v)^2 + Z^2} \quad \textbf{(2.2)}$$

Applying the binomial theorem [36], Equation 2.5 can be approximated by:

$$r = \mathrm{Z}\sqrt{1 + \left(\frac{x-u}{z}\right)^2 + \left(\frac{y-v}{z}\right)^2}$$

$$\approx Z\left(1 + \frac{1}{2}\sqrt{\left(\frac{x-u}{z}\right)^2 + \left(\frac{y-v}{z}\right)^2}\right) \quad \textbf{(2.3)}$$

Substituting the approximation for *r* into Equation 2.4 results in the Fresnel/near field diffraction integral:

$$U_1(x,y) = \frac{exp(ikZ)}{i\lambda}\iint_A U_0(u,v)exp\left(i\frac{k}{2Z}[(x-u)^2 + (y-v)^2]\right)dudv \quad \textbf{(2.4)}$$

Expanding all terms of the exponential inside the integrand, the Fresnel integral becomes:

$$U_1(x,y) = \frac{\exp(ikZ)}{i\lambda Z}\exp\left(\frac{ik}{2Z}\left(x^2+y^2\right)\right)\ldots \\ \iint_A U_0(u,v)\exp\left(\frac{ik}{2Z}\left(u^2+v^2\right)\right)\exp\left(-i\frac{2\pi}{\lambda Z}[xu+yv]\right)dudv \quad \textbf{(2.5)}$$

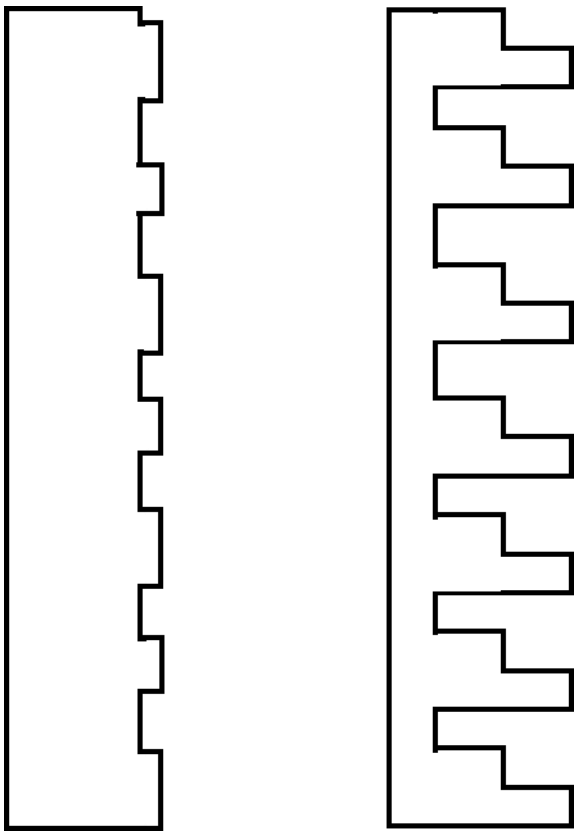
If $Z >> \lambda$, the exponential term $exp\left(\frac{ik}{2Z}(u^2+v^2)\right)$ approaches unity and Equation 2.8 becomes:

$$U_1(x,y) = \frac{\exp(ikZ)}{i\lambda Z}\exp\left(\frac{ik}{2Z}\left(x^2+y^2\right)\right)\ldots \\ \iint_A U_0(u,v)\exp\left(-i\frac{2\pi}{\lambda 2Z}(xu+yv)\right)dudv \quad \textbf{(2.6)}$$

Equation 2.9 is known as the Fraunhofer/Far-field integral.

2.5 Grating Structure Types

The range of grating structure types can be classified into two categories: amplitude and phase, either transmissive or reflective. An amplitude grating (Figure 2.10(a)) only acts on the amplitude of the incident wave e.g. dark slits on glass. A phase grating (Figure 2.10(b)) acts only on the phase of the incident wave e.g. Echellette grating which has a continuous profile [40].



(a) **(b)**

Figure 2.10: Examples of a) a binary phase grating b) a quantised phase grating.

In the Thesis all gratings are assumed to be phase i.e. either 0 or 1 (whereas for an amplitude grating phase is always 0). A phase grating can be easily quantised to generate multiple phases; a phase grating is fabricated by quantising its height, expressed as a number of quantised levels e.g. a binary grating with two levels. The levels progress in powers of two until a continuous grating structure is defined. The number of levels has differing effects on the grating output characteristics, the first and most important being the diffraction efficiency. As the number of levels increases so does the diffraction efficiency [39-40].

The mathematical framework to analyse the performance of different types of gratings is illustrated in Figure 2.11.

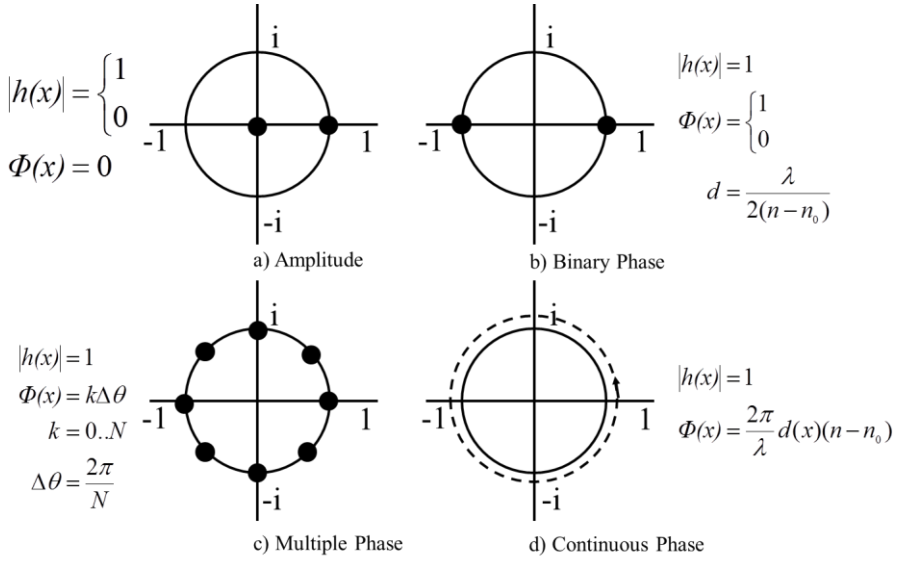


Figure 2.11: Graphical representation on the unit circle of the complex amplitude transmittance of a) a binary amplitude grating b) a binary grating c) a multiple phase grating, d) a continuous phase grating.

Mathematically a phase grating structure, labelled *h(x)*, is represented as $h(x) = \exp^{(i*\Phi(x))}$ where *Φ(x)* is its phase defined as:

$$\Phi(x) = \frac{2\pi}{\lambda} \mathrm{d}(x) * (n - n_0) \qquad \textbf{(2.7)}$$

for a continuous grating which leads to a corresponding value for *h(x)* on the unit circle in the complex plane. When *Φ =-π or 0* and *h(x) = –1 or* 1, the grating is a binary structure (Figure 2.11) [37-38]. A quaternary grating has *h(x) = -1,-i, i, or 1* with *Φ = -π,* $-\pi/2$, *0,* $\pi/2$, [39-40].

2.6 Discrete Fourier Transform (DFT)

The diffraction calculation relies on a continuous Fourier Transform [39-40]; however, in practice a numerical calculation using a representation of the problem involving a finite number of parameters must be performed. Choosing the right representation conveniently casts the calculation into a Discrete Fourier Transform (DFT), enabling a Fast Fourier Transform (FFT) to be used.

First consider a one dimensional hologram composed of *N* pixels separated at a pitch *p* and initially, that each pixel is a Dirac Delta function [38-40]. A discrete hologram function represented by *h(x)* can then be defined as:

$$h(x) = h(x)_c * comb\left(\frac{x}{p}\right) rect\left(\frac{x}{Np}\right) \quad \textbf{(2.8)}$$

where $h(x)_c$ represents the continuous hologram, *comb(x/p)* represents the sampling by the Dirac-Delta (representing each pixel) and *rect(x/Np)* represents the finite spatial extent of the hologram. The function is defined so that *rect(x) = 1* for $-½ \leq x \leq ½$, otherwise it is zero.

The diffracted amplitude produced by a beam passing through the DOE in the Fraunhofer regime is defined as *H = FT(h)* where FT represents the Fourier Transform operator given by:

$$H(u) = H(u)_c \otimes comb(pu) \otimes sinc(Npu) \quad \textbf{(2.9)}$$

Where $\otimes$ represents the convolution operator and the *sinc* function arises because aperture of the diffractive optic. The *sinc(x)* function is represented by $sinc(x) =$

$\sin(\pi x)/\pi x$ [38]. The comb term represents the replication of the reconstructed image, which is infinitely periodic owing to sampling.

Because the diffracted intensity is important rather than the phase, many periods of the hologram are illuminated, allowing the phase of the diffracted amplitude *H(u)* to be random. In this case, the *sinc* interpolation from Equation 2.12 no longer results in a continuous and smooth reconstruction, instead it produces a speckle pattern which is a result of interference of many different waves at the same frequency of different phase and amplitude.

The effects of the laser speckle may be reduced if the DOE is replicated, leading to a new definition of *h(x)* given by:

$$h(x) = h(x)_c * comb\left(\frac{x}{p}\right) * comb\left(\frac{x}{Np}\right) \quad \textbf{(2.10)}$$

with the second *comb()* representing an infinitely periodic hologram, which in turn results in a new definition of *H(u)*:

$$H(u) = H(u)_c \otimes comb(pu) \otimes comb(Npu) \quad \textbf{(2.11)}$$

The diffracted amplitude is now sampled by the second *comb()*, its intensity being zero at all locations except where it is precisely determined by the original encoding. Combining Equation 2.11 and Equation 2.13, *h(x)* now becomes (illustrated in Figure 2.12):

$$h(x) = \left\{\left[h(x)_c * comb\left(\frac{x}{p}\right)\right] \otimes rect\left(\frac{x}{p}\right)\right\} \otimes comb\left(\frac{x}{Np}\right) \quad \textbf{(2.15)}$$

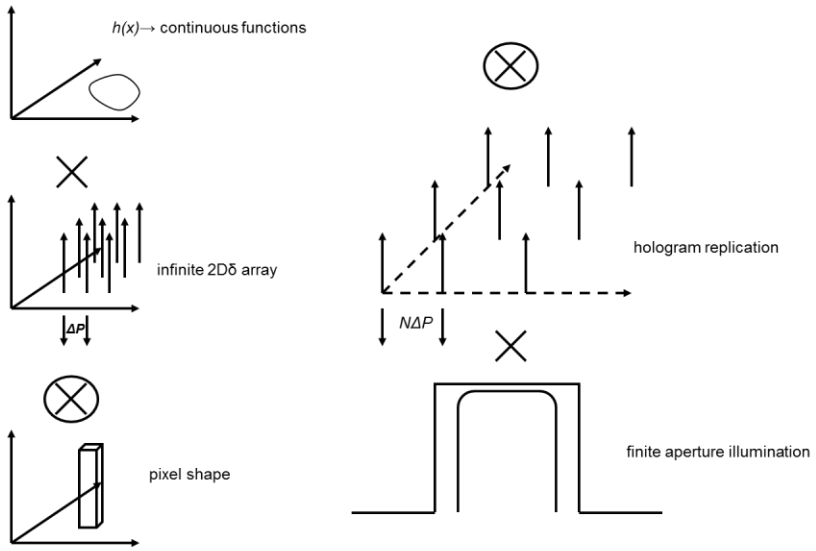


Figure 2.12 Illustration of the effects of the DFT in the hologram plane.

The first convolution term represents the sampling point convolved by the square pixel, with leads to the definition of *H(u)* as (illustrated in Figure 2.13);

$$H(u) = \{[H(u)_c \otimes comp(pu)] * sinc(pu)\} * comb(Npu) \quad \textbf{(2.16)}$$

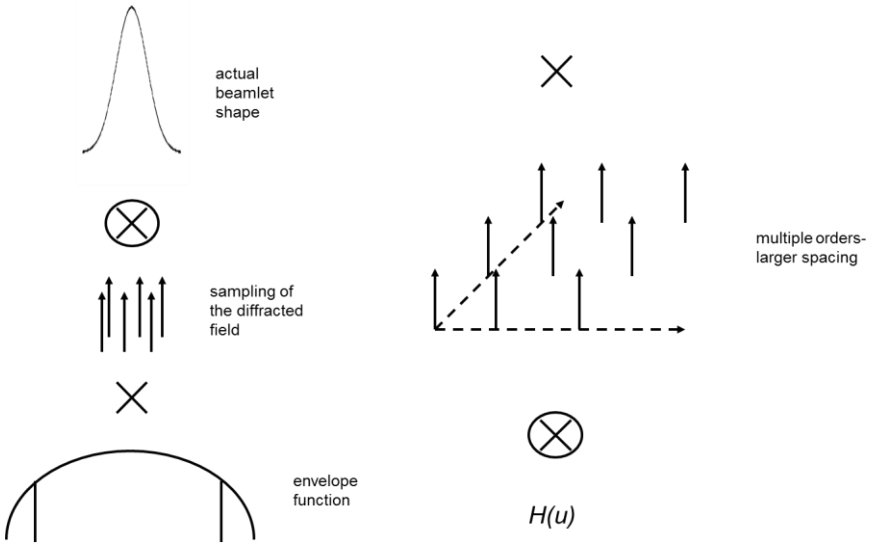


Figure 2.13 Illustration of the effects of the DFT in the reconstruction plane.

For a square pixel, a second *sinc* term is added to the diffracted amplitude, yielding a $sinc^2$ intensity envelope with the central lobe the same height as the first period of the diffracted intensity. If the amplitude being calculated uses the DFT, the *sinc* term arising from the square pixel must be explicitly included.

2.7 Conclusions

The phenomenon of interference is essential to the development of holography and has been the subject of massive development. The field has moved from generating holograms on photographic plates to holograms generated using significant computing platforms and manufactured using semiconductor mass production processes. DOEs are presently in use in a number of different applications and are stimulating new technologies to be developed, increasing production capabilities in a number of key industry sectors and pushing the boundaries of others.

With all these different opportunities, new methods that improve the quality or produce real time solutions are required to sustain and evolve progress to date. The root of any new development is an in-depth knowledge of diffraction theory to capture and formulate the problem and develop its understanding. This knowledge is also core when evaluating the operational difference between the near-field and the far-field regions. The quadratic phase factors that occurs in the treatment of the near-field region presents more specific challenges when developing an algorithm to generate a DOE and is discussed in subsequent Chapters.

Knowledge of Fourier Transforms is extremely important since computing frameworks are not capable of treating a continuous Fourier Transform and are tailored to execute on discrete quantities'. The limitation must be properly understood so that judgements on the quality and performance of any DOE generated using a computer can be robustly supported.

Chapter 3

Diffractive Optical Elements (DOEs)

3.1 Introduction

A number of algorithms have been developed to calculate DOEs, the most noteworthy being the Iterative Fourier Transform Method (IFTA) [26-27], Genetic Algorithms (GA) [25], the Direct Binary Search (DBS) [28-29] and the Simulated Annealing (SA) [30-31]. The aim of the Chapter is to review two of the above algorithms - DBS and SA - due to their similarity to the proposed POGED method. The root of the similarity is that these two algorithms evaluate an energy function, defined as the difference between a 'flipped' pixel and its previous state, as does POGED, the only difference being in which plane the function is evaluated and the manner it is filtered. The Chapter also provides a basis for a comparison, the metrics being the time to generate a DOE and the quality of the reconstructed image.

3.2 Direct Binary Search (DBS) and Simulated Annealing (SA)

In their simplest form, DBS and SA are downhill search algorithms that operate by finding a local or global minimum of an objective function [28-31]. For DBS this may result in arresting at a local minimum, whilst with SA the global minimum is sought with a sufficiently slow annealing schedule, executed through the use of a probability function that allows pixel states to be accepted even if their energy is not minimised.

The energy function depicted in Figure 3.1 demonstrates how the solution progresses towards optimisation. The start of this class of search technique is in the selection of the form of a random grating structure; for ease of illustration a binary grating is chosen. The random grating is then modified using any one of the highlighted techniques so that the actual image closely approximates the target image, according to calculations of the diffraction efficiency and Signal-to-Noise Ratio (SNR).

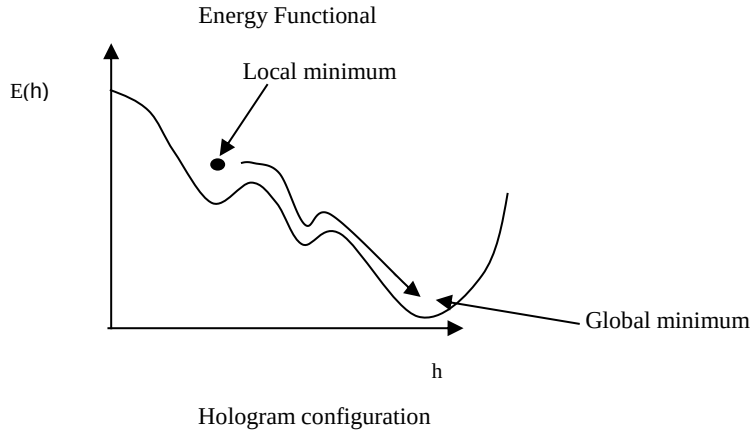


Figure 3.1 Schematic of the cost function illustrating a downhill search.

The DBS and SA methods proceed by changing the state ('flipping' in the binary case) of a pixel within a $N \times N$ array random grating structure. A calculation evaluates the difference in the energy between the new and previous states. For DBS, if the energy is reduced the pixel flip is accepted; otherwise the previous state is kept. The algorithm proceeds to the next pixel and repeats the process, continuing until all pixels have been visited. If the last pass resulted in the acceptance of one or more pixel flips, a new pass is initiated; otherwise the algorithm terminates. SA takes the above methodology one step further by additionally accepting a pixel flip

when the change in the energy is positive with a probability $p = exp(-\Delta E/kT)$ where kT is a computational temperature which decreases slowly (c.f. annealing) as the algorithm proceeds. In principle, this permits SA to locate the global optimum but at of the expense of additional computational resources.

3.2.1 Direct Binary Search (DBS)

DBS is an iterative algorithm that provides the foundation for all algorithms presented in the Thesis. The primary difference between the proposed POGED algorithm and DBS is that all calculations for the latter algorithm are executed in the reconstruction (Fourier) plane whilst for POGED, described in detail in Chapter 4, the bulk of the calculations are performed in the hologram (object) plane.

In DBS, a hologram h is defined over an array of $N \times N$ pixels, of value $+1$ or -1 in the binary case. Considering a 1D case for simplicity, the diffracted field amplitude is given by the Fourier Transform of the transmitted hologram h:

$$H(m) = \sum_{j=1}^{N} h(j) \exp\left(\frac{-2\pi ijm}{N}\right) * \operatorname{sinc}\left(\frac{m}{N}\right) \tag{3.12}$$

If the j^{th} pixel is flipped, the provisional diffracted field amplitude H' is given by:

$$\begin{aligned} H'(m) &= H(m) - 2h(j) \exp\left(\frac{-2\pi ijm}{N}\right) * \operatorname{sinc}\left(\frac{m}{N}\right) \\ &= H(m) - 2h(j)L(j,m) \end{aligned} \tag{3.13}$$

where the factor L is:

$$L = \exp\left(\frac{-2\pi ijm}{N}\right) * \operatorname{sinc}\left(\frac{m}{N}\right) \tag{3.14}$$

and is stored as a look-up table to speed up calculations [30-31]. The *sinc* envelope arises from the diffraction pattern owing to a single square pixel of the grating structure.

The energy function - a cost metric - is defined as $E = \left\| |H| - \alpha |F| \right\|_{\Omega}^{2}$, where Ω denotes that the norm is evaluated only over a signal window (Figure 3.2) where the target amplitude $|F|$ is contained within the signal window in the target plane. The reconstructed amplitude $|H|$ is contained in the reconstruction plane, α determining scale freedom [55]. The scaling factor is given by $\alpha = N_F / \eta_T N_H$, where $N_F = \sum\sum |F|^2$, $N_H = \sum\sum |H|^2$ and η_T is the diffraction efficiency defined as:

$$\eta_T = \frac{\sum_{j \in \Omega} |H(j)|^2}{\sum_{j \in \mathrm{E}} |H(j)|^2} \tag{3.15}$$

The diffraction efficiency is defined as the ratio of the diffracted energy in the reconstruction plane inside the signal window Ω divided by the total diffracted energy (the whole plane is denoted by the universal set E) [30-31]. The Signal-to-Noise Ratio (SNR) is defined as [30-31, 37-38]:

$$SNR = \frac{\iint |F(x,y)|^2 dxdy}{\iint \left(|H(x,y)| - |F(x,y)| \right)^2 dxdy} \tag{3.16}$$

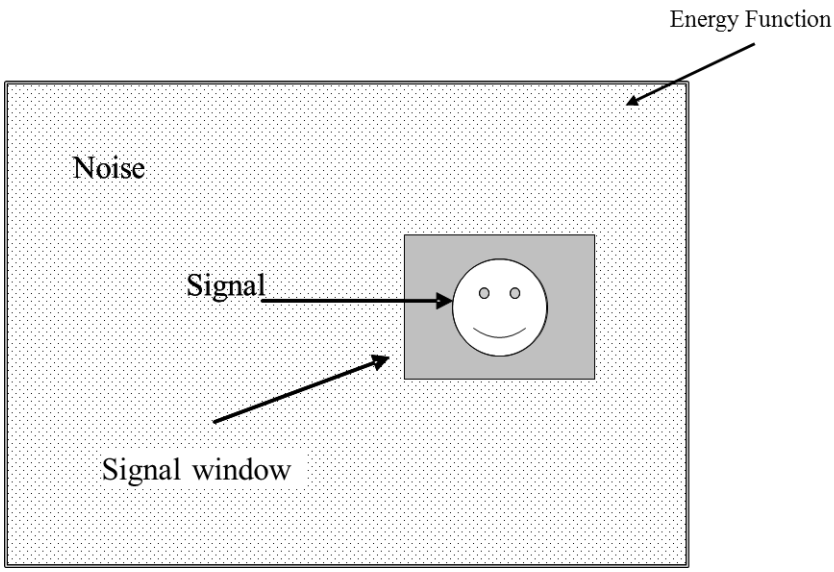


Figure 3.2 Illustration of the signal window.

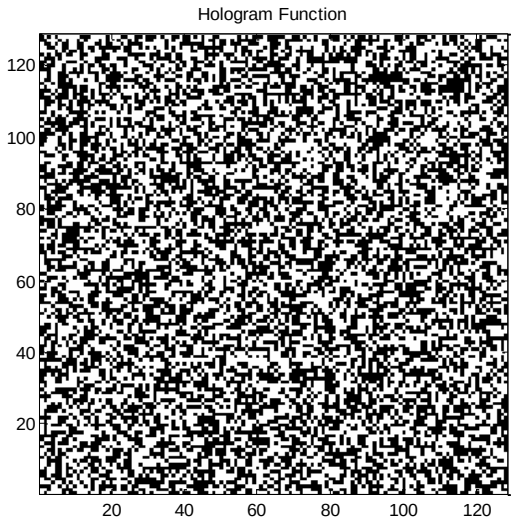
The algorithm operates by flipping the state of a pixel followed by a re-calculation of the energy; for the binary case, the pixels take the value of *+1* or *-1*. If the energy is less than the previous state, the pixel flip is accepted and the algorithm proceeds to the next pixel. If however, the energy is higher than the previous state, a pixel flip is rejected and the original pixel state is retained. This algorithm takes a short time to compute, however it is a 'greedy' algorithm which can become trapped within a local minimum [28-29]. One way to escape a local minimum is to introduce a probability function allows pixel flips to be accepted when the energy is increased, the principle embedded in SA.

3.2.1.1 DBS Results

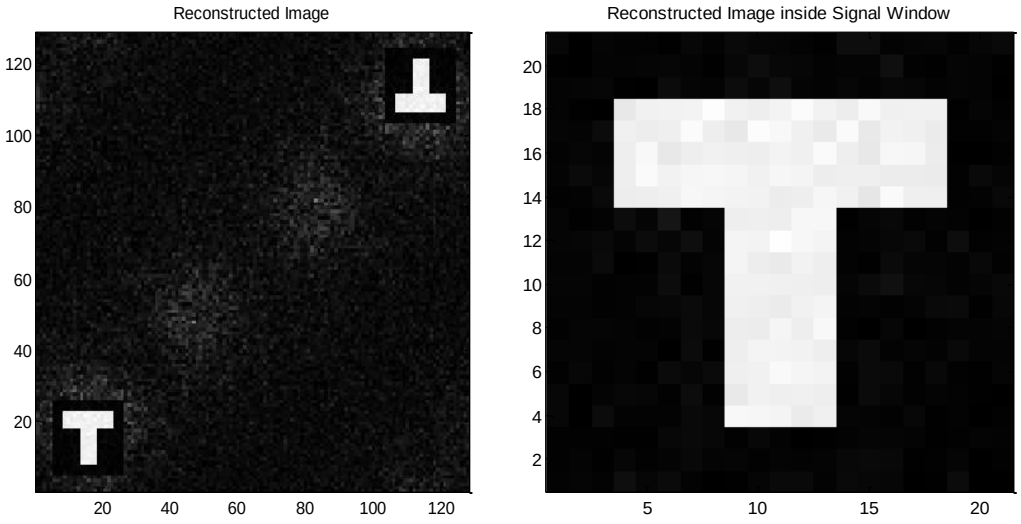
Calculations for a grating structure consisting of *128×128* pixels with a signal window of *21×21* pixels generated by DBS on a computer - a Xeon 3.2GHz

processor with 3.2 GB of RAM under Windows XP - take 1 minute and 38 seconds (Figure 3.3).

The diffraction efficiency is evaluated to be 68%, much lower than the theoretical maximum of 74%, due to the fact that there are a number of errors outside the signal window (Figure 3.3(b)). The SNR of the diffracted field is 10dB which is relatively poor, (Figure 3.3(c)) due to errors within the signal window, both surrounding and within the target.



(a)



(b) **(c)**

Figure 3.3 Results for a grating structure generated using DBS: a) hologram function b) full reconstructed Image c) reconstructed inside signal window.

Results for the DBS algorithm are poor due to the algorithm arresting in a local minimum, one of the major drawbacks of the method and the reason why the algorithm is executed in a relatively short time period.

3.2.2 Simulated Annealing (SA)

The SA method extends DBS a stage further by embedding a secondary decision step in the evaluation of the energy function, through the employment of a probability function. In SA, the same energy function is optimised, $E = \left\| |\mathbf{H}| - \alpha |\mathbf{F}| \right\|_{\Omega}^{2}$ within the signal window. The embedding of the probability function allows for cases when the energy function is higher than the previous state to be accepted, allowing the algorithm to escape any local minima, and in turn able to proceed to locate the global minimum.

The probability function is defined as:

$$p = \exp(-\Delta E / kT) \tag{3.17}$$

where T is a 'computational temperature' [30-31], k is Boltzmann's constant, and ΔE is the difference in the energy of states. A pixel flip is accepted with probability *p* and the next pixel is evaluated, progressing the probability distribution over the state space of the hologram to reach equilibrium:

$$\text{prob}(h)dh = \frac{\exp\left(-E(h)/kT\right)dh}{\int_{H} \exp\left(-E(h)/kT\right)dh} \tag{3.18}$$

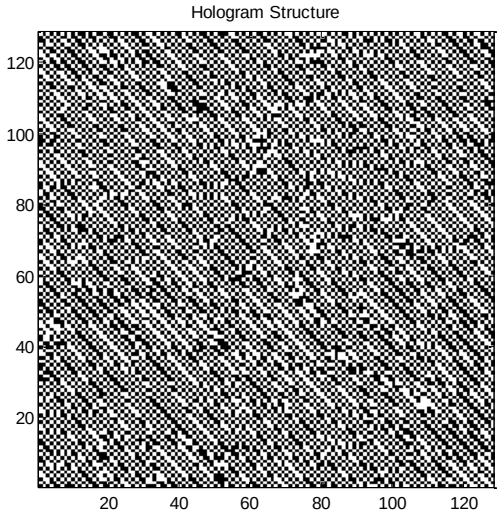
where *H* is the set of all permitted hologram functions.

If *T* reduces slowly, the probability of the energy function reaching the global minima reaches one. *T* should therefore be reduced adiabatically for the system to remain in equilibrium, the cooling defined as $T(z) = 1/\ln(z)$ where *z* is the iteration number.

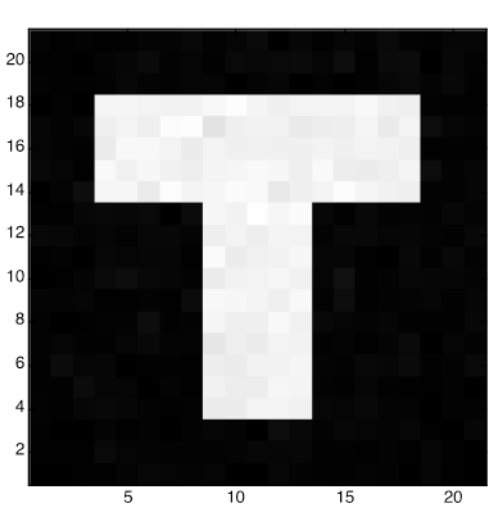
3.2.2.1 SA Results

Calculations for a grating structure consisting of *128×128* pixels with a signal window of *21×21* pixels generated by SA on an identical computing platform used for the DBS evaluation took 21 minutes and 38 seconds.

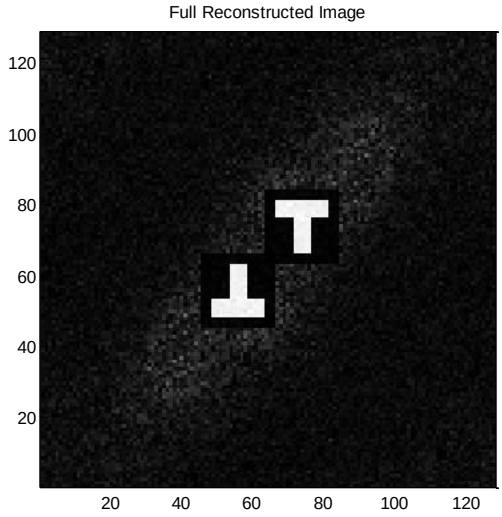
Figure 3..4 illustrates the results obtained by SA. The full reconstructed image, Figure 3.4(c) shows two off-axis reconstructed images with errors along the diagonal the images are centered on. The background in this case still contains a large amount of noise indicative of the computed diffraction efficiency of 69%, lower than the theoretical maximum of 72%, which includes both the reconstructed image and its inverted image. However, since the grating is defined as an array of *128×128* pixels, there are insufficient degrees of freedom to achieve improved results. A larger array size results in better diffraction efficiency [39-40].



(a)



(b)



(c)

Figure 3.4 a) A hologram generated using SA producing a T structure in the reconstruction plan (b) reconstructed image of the T structure in the signal window (c) full reconstruction plane showing the twin images generated by a binary grating.

The SNR for the grating structure is 26dB, consistent with little noise in the region surrounding the target in the signal window (Figure 3.4(b)).

3.3 Conclusions

The DBA and SA algorithms are successfully implemented in Matlab on a Xeon 3.2 GHz processor with 3.2Gbytes of RAM operating under Windows XP. These algorithms were chosen as both have roots in a similar methodology underpinning the proposed POGED and thus provide benchmark performance targets. The results of the evaluation are summarized in Table 3.1.

It is clear that the SA provides the best SNR but at the expense of longer calculation time. The diffraction efficiency results are governed by the principles of the algorithm; DBS can arrest in a local minimum whilst SA is able to find the optimum local minimum. Computational times are however still large in both cases which restrict the spectrum of feasible applications.

Table 3.1 Summary of results for the DBS and SA algorithms for a binary grating structure.

Algorithm	Time	Computed Diffraction Efficiency (%)	SNR (dB)
DBS	0:1:38	68%	10 dB
SA	0:21:38	69%	26dB

Table 3.2 encapsulates the advantages and disadvantages of both DBS and SA. SA provides a better SNR but its computation time is inferior to DBS.

Table 3.2 Comparison of advantages and disadvantages of the DBS and SA algorithms.

	Advantage	Disadvantage
Time	DBS	SA
SNR	SA	DBS
Diffraction Efficiency	DBA/SA	DBS/SA

POGED is designed to provide both these advantageous features, creating implementations with the quality of SA at DBS timescales.

Chapter 4

Phase Optimised General Error Diffusion (POGED)

4.1 Introduction

The Phase Optimised General Error Diffusion (POGED) algorithm is a modified version of the error diffusion algorithm first developed for the printing industry to optimise the rendering of half-tone images [34]. The use of phase optimisation improves calculation times. The quality of the reconstructed image is compared Simulated Annealing (SA) and Direct Binary Search (DBS) algorithms.

The Chapter focuses on the design of the POGED algorithm for the synthesis of Fourier plane diffractive optical elements, followed by a description of how the algorithm is implemented in the binary Fourier hologram case. An explanation is given of how it is modified to accommodate the case of multiple pixel level holograms.

4.2 POGED Definition

POGED is similar to DBS [54-56], the primary difference being that the bulk of the calculations for the latter are executed in the object plane whereas in POGED,

calculations are performed in the Fourier plane. The modification allows the speed of the DBS to be achieved with the quality of the SA algorithm, as explained later in the Chapter. Another major difference lies in the use of a convolution mask which scans across the array as the algorithm progresses, filtering out any local errors. Phase optimisation is an integral dimension within POGED based on a variant of the Gerchberg-Saxon algorithm popularised by Fienup [24, 14-16]. Phase optimisation provides further correction of the phase of the target image, yielding a well corrected reconstructed image.

Assume a plane wave of amplitude $h_0(x, y)$ passing through the DOE to create a desired target image; also assume that the DOE follows a phase function defined by $\phi(x, y)$. After the DOE, the plane wave acquires a complex amplitude, in the thin phase screen approximation [38, 54-56] given by:

$$h(x, y) = h_0(x, y)\exp(i\phi(x, y))) \quad \textbf{(4.1)}$$

After a distance *Z* from the DOE, the complex amplitude has diffracted sufficiently to form the reconstructed image, *H*, where $H = \Im(h)$ and $\Im$ represents a specific propagation operator. This operator is the Fresnel Transform if the distance *Z* is in the near field region or the Fourier Transform if *Z* falls within the far field region [38].

POGED determines the phase function, $\phi(x, y)$ that produces the desired reconstructed image, executing on this goal through minimisation of an energy

defined as $E = \langle PG, G \rangle$[1], where P is a positive function allowing the distribution of the errors described by $G = H - \alpha F$ weighted according to the significance placed upon them. H is the reconstructed image, F is the target image and α is a scaling factor.

A few key variables must be defined before progressing to the mathematical formulation that embodies POGED;

- the hologram function consists of an array of $N \times N$ elements
- the target image is defined as a separate array of $M \times M$ elements and can be either a binary or a grey scale image
- a mask forms the convolution kernel (impulse response) of a filter and is calculated using Dykstra's Alternating Projections method [60-61]. The kernel is defined as an array band limited to $S \times S$ elements implementing a real, positive transfer function (Fourier Transform). The transfer function forms the 'soft' weight P of the energy function defined above
- Dykstra's algorithm ensures that the weight function satisfies the band limited and positivity constraints for a target 'hard' weight, used as the initial value in Dykstra's procedure [60-61].

---

[1] where:

$$< A, B >= \iint A(x, y) B^{*}(x, y) dxdy$$

is the Hermitian inner product in the relevant space of two complex functions with B^{*} representing the complex conjugate.

The following variables are defined: F is the target function, H is the reconstructed image and P is the weight function defined in the reconstruction plane. The inverse Fourier Transforms of these functions are defined as f, h, p where f is the Fourier target object, h is the hologram, and p is the convolution kernel of the filter defined in the hologram plane. These Fourier Transforms pairs are:

$$\begin{aligned} F &= \Im(f) \\ H &= \Im(h) \\ P &= \Im(p) \end{aligned} \tag{4.19}$$

The energy function is defined as:

$$E = \langle P(H - \alpha F), (H - \alpha F) \rangle = \langle p \otimes (h - \alpha f), (h - \alpha f) \rangle \tag{4.20}$$

where E is a function of both the target and reconstructed image, α is a scalar variable treating scale freedom between the target and reconstructed images [57].

The constraint that P is real and positive ensures that:

$$\begin{aligned} \langle u, v \rangle &= (p \otimes u, v) \\ \langle U, V \rangle &= (PU, V) \\ U &= \Im(u) \\ V &= \Im(v) \end{aligned} \tag{4.4}$$

where u, v are any variables and are inner products in weighted Hilbert space, following from Parseval's Theorem.

The minimum of the energy function with respect to the scaling factor is found analytically by:

$$\alpha = \frac{\langle PH, F \rangle}{\langle PF, F \rangle} = \frac{\langle p \otimes h, f \rangle}{\langle p \otimes f, f \rangle} \tag{4.5}$$

POGED performs its descent on the energy function E, restricted to the optimum value of scaling factor α.

In most applications, the target function *F* has phase freedom. The POGED energy function may therefore be further refined by optimising the phase of the target image. With proper choice of the phase of *F*, the inner product $\langle PH, F \rangle$ can be maximised and hence the energy function reduces in value.

Since:

$$\left|\langle PH, F \rangle\right| \le \langle P|H|, |F| \rangle \qquad \textbf{(4.6)}$$

the appropriate choice is $\arg(F) = \arg(H)$ so that $\langle PH, F \rangle = \langle P|H|, |F| \rangle$ i.e. the phase of the target image is set to equal the phase of the reconstructed image [15]:

$$F = |F| \exp\left(i \arg(H)\right) \qquad \textbf{(4.7)}$$

The operation is most simply performed in the reconstruction plane. However, the minimisation of the POGED energy function with respect to the hologram pixel states is most efficiently performed in the hologram plane where the mask size is small since it is band limited. The phase update governed by Equation 4.7 is applied infrequently; so the bulk of the computation occurs in the hologram plane, decreasing calculation times.

Thus far, the description of POGED has assumed a Fourier Transform relation between functions defined in the hologram and reconstruction planes i.e. the reconstruction occurs in the far-field of the hologram. Although the detailed relationships depend on the propagation operators, very similar expressions result in other diffraction regimes, corroborated when the Fresnel regime is considered in Chapter 5, Section 5.2

The restricted energy function is now given by:

$$E(h) = \langle p \otimes \Pi h, \Pi h \rangle \tag{4.8}$$

where the linear operator Π is defined as:

$$\Pi(\bullet) = I(\bullet) - \frac{\langle p \otimes (\bullet), f \rangle f}{\langle p \otimes f, f \rangle} \tag{4.9}$$

Moreover, easily verified by direct computation, Π is a projection:

$$\Pi^2 = \Pi \tag{4.10}$$

with a one dimensional null space spanned by f and self-adjoint with respect to the weighted inner product:

$$\langle p \otimes \Pi u, v \rangle = \langle p \otimes u, \Pi v \rangle \tag{4.11}$$

It then follows that Equation 4.11 is an orthogonal projection with respect to the weighted inner product:

$$\langle p \otimes (u - \Pi) u, \Pi u \rangle = 0 \tag{4.12}$$

Π maps any function to the closest function (with respect to the norm induced by the weighted inner product) orthogonal to f.

The above properties enable the restricted energy function to be written in a simpler form:

$$E(h) = \langle p \otimes \Pi h, h \rangle \tag{4.13}$$

Since the algorithm manipulates each pixel sequentially, it is necessary to write the energy function as a function of individual pixel values. Hence the description of the hologram function *h* requires refinement by the variable *e*, representing a pixel shape function at different array positions:

$$h(x, y) = h_{\setminus ij}(x, y) + h_{ij} e_{ij}(x, y) \tag{4.14}$$

where the subscript ij denotes array positions and $\backslash ij$ denotes the *omission* of the element indexed by (i, j) . With continuous variables, the function *e* describes the shape of individual pixels. When the continuous problem is discretised for compatibility with computer calculations, the Kronecker-Delta centered at (x_i,y_j) is followed:

$$e_{ij}(x, y) = \delta(x - x_i, y - y_j) \tag{4.15}$$

The energy function now demonstrates a quadratic dependence on h_{ij} :

$$E = a + b^* h_{ij} + b h_{ij}^* + d\left|h_{ij}\right|^2 \tag{4.16}$$

where

$$\begin{aligned} a &= \left\langle p \otimes \Pi h_{\backslash ij}, h_{\backslash ij} \right\rangle \\ b &= \left\langle p \otimes \Pi h_{\backslash ij}, e_{ij} \right\rangle \\ d &= \left\langle p \otimes \Pi e_{ij}, e_{ij} \right\rangle \end{aligned} \tag{4.17}$$

Parameters *a, b, d* are independent of h_{ij}.

Given the research is exclusively concerned with phase-only holograms for which;

$$h_{ij} = \exp(i\phi_{ij}) \tag{4.18}$$

then the energy is minimised when:

$$\phi_{ij} = \arg(-b) \tag{4.19}$$

Equation 4.19 may be used to treat holograms with continuously valued phase but needs to be modified for multi-phase gratings. In practice a suitable quantisation rule is used to select the nearest supported phase. For equally spaced phase levels with a step size $\Delta\phi = \dfrac{2\pi}{M}$, a suitable rule is:

$$\phi_{ij}^{new} = \Delta\phi \, mod\left(round\left(arg(-b)\right)/(\Delta\phi), M\right) \tag{4.20}$$

Inserting the step size into Equation 4.20, the change between the new and the previous states becomes:

$$\left|\phi_{ij}^{new} - \phi_{ij}^{old}\right| \geq \Delta\phi/2 = \pi/M \tag{4.21}$$

Alternative rules apply in the binary and quaternary cases. For the binary case, an appropriate quantisation rule is:

$$h_{ij} \leftarrow \mathrm{sgn}\left(\mathrm{real}\left(-b\right)\right) \tag{4.22}$$

For the quaternary case where $h_{ij} \in \{-1,-i,+1,+i\}$, the following rule is appropriate:

$$\begin{aligned} u &= \mathrm{real}\left(-b\right) \\ v &= \mathrm{imag}\left(-b\right) \\ &\left|u\right| > \left|v\right| \\ h_{ij} &\leftarrow \mathrm{sgn}\left(u\right) \\ &\left|v\right| > \left|u\right| \\ h_{ij} &\leftarrow i\,\mathrm{sgn}\left(v\right) \end{aligned} \tag{4.23}$$

Note that:

$$\begin{aligned} b &= \left\langle p \otimes \Pi h_{\backslash ij}, e_{ij} \right\rangle = c - d h_{ij} \\ c &= \left\langle p \otimes \Pi h, e_{ij} \right\rangle = \left(p \otimes h\right)_{ij} - \alpha\left(p \otimes f\right)_{ij} \\ d &= \left\langle p \otimes \Pi e_{ij}, e_{ij} \right\rangle = p_{00} - \frac{\left|\left(p \otimes f\right)_{ij}\right|^2}{\left\langle p \otimes f, f \right\rangle} \\ \alpha &= \frac{\left\langle p \otimes h, f \right\rangle}{\left\langle p \otimes f, f \right\rangle} \end{aligned} \tag{4.24}$$

and that α can be rewritten as:

$$\alpha = \frac{\left(p \otimes h_{\backslash ij}, f\right) + h_{ij}\left(p \otimes f\right)_{ij}}{\left(p \otimes f, f\right)} \tag{4.25}$$

Hence when a pixel state is changed, it may be updated according to the rule:

$$\alpha^{new} \leftarrow \alpha^{old} + \left(h_{ij}^{new} - h_{ij}^{old}\right)\frac{\left(p \otimes f\right)_{ij}}{\left\langle p \otimes f, f \right\rangle} \tag{4.26}$$

Because the algorithm is classified as "greedy", each pixel is visited and its state is determined according to the appropriate quantisation rule. The rule requires knowledge of each variable evaluated by the algorithm, *f*, *h*, $p \otimes f$, $p \otimes h$; all variables depend on hologram configuration and need to be updated as the algorithm progresses. The decision on when to update variables becomes critical and can either be updated after each pixel evaluation or at the completion of the quantisation pass.

The pixel by pixel update of the variable *h* requires minimal effort which in turn allows for the update of $p \otimes h$ yielding:

$$(p \otimes h)^{new} \leftarrow (p \otimes h)^{old} + \left(h_{ij}^{new} - h_{ij}^{old}\right)(p \otimes e_{ij}) \qquad \textbf{(4.27)}$$

especially useful when the filter kernel *p* is small. The variable *f*, which depends on the phase of *H*, is best calculated through the Fast Fourier Transform (FFT) only after each full pass of the whole array. The target may always be normalised so that $\langle p \otimes f, f \rangle = 1$, saving on computational resources.

4.2.1 Implementation

POGED initiates by defining a target function captured in a *NxN* array, either grey level or binary. Calculation of the mask function follows, executed using Dykstra's Alternating Projections [60-61]. Once completed, a random grating structure is defined which assume one of two forms; a binary grating of only two-phase levels, $h_{ij} \in \{-1,+1\}$ or a phase grating of sub-multiple phase levels. These phase levels are then defined as $\theta = \frac{2\pi k}{L+1}$ for $k = 0 \ldots L$ where L is the number of quantised levels

which relates to a unit circle in the complex plane. Either grating structure has the same dimensions as the target function. At this stage, a calculation of the initial energy utilising a random grating and the mask function is performed, and the process for structuring POGED is ready.

POGED begins by changing the phase of *h*, based upon a calculation of the cost function, followed by a comparison of the difference between the new and previous phase values of *h* to a specified tolerance level. If the difference is greater, the phase change of *h* is accepted; if the difference is less, the change is rejected. Once a pixel has been evaluated, the algorithm addresses the next pixel. After one pass through the array, the phase of the target structure is updated and $p \otimes f$ recalculated. In addition, $p \otimes h$ is re-calculated to prevent accumulation of rounding-off errors. The flow diagram for POGED is given in Figure 4.1 presenting each step of the algorithm. The flow diagram for the filter function is covered in Section 4.4.

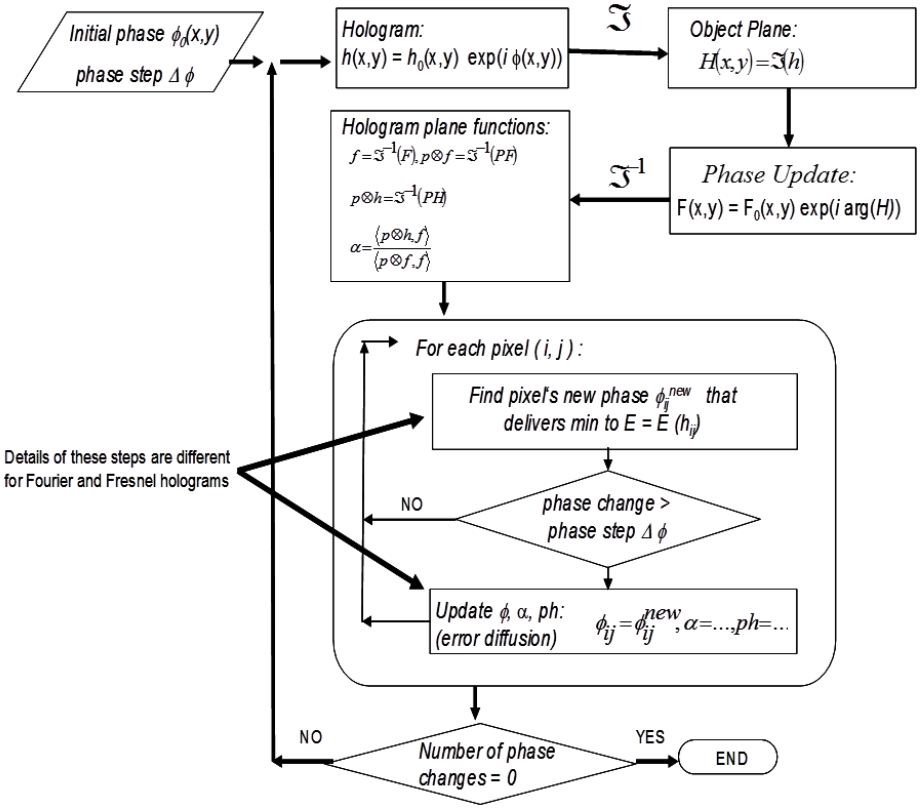


Figure 4.1 Flow diagram for POGED.

4.3 Performance Metrics

4.3.1 Diffraction Efficiency and Signal-to-Noise Ratio (SNR)

The diffraction efficiency and Signal-to-Noise Ratio (SNR) are the two performance metrics used to evaluate holograms once all pixel values are fixed i.e. when the energy has been minimised.

The diffraction efficiency evaluates how much energy is in the desired order and the SNR is a measure of background noise in relation to the intensity of the reconstructed image. These two metrics represent the overall quality of the reconstructed image and are routinely used by the community [18-20, 30-31, 38-39].

The metrics are calculated using the normalised total energy defined as:

$$
\begin{aligned}
E^2 &= \frac{(H-\alpha F, H-\alpha F)}{(H,H)} \\
&= 1-\alpha\frac{(F,H)}{(H,H)} - \alpha^*\frac{(H,F)}{(H,H)} + |\alpha|^2\frac{(F,F)}{(H,H)} \\
&\text{but} \\
\alpha &= \frac{(H,F)}{(F,F)} \\
&\Rightarrow \\
E^2 &= 1-|\alpha|^2\frac{(F,F)}{(H,H)} = 1-\eta
\end{aligned}
\tag{4.28}
$$

where the diffraction efficiency[2] η is defined as:

$$\eta = \frac{|\alpha|^2(F,F)}{(H,H)} = 1 - E^2 \tag{4.29}$$

and the SNR is defined as:

$$SNR = \frac{|\alpha|^2(F,F)}{(\chi(H-\alpha F),(H-\alpha F))} \tag{4.30}$$

where $\chi(H-\alpha F)$ is the error (noise) within the signal window, α the scaling factor, F the target function, H the reconstructed image and χ the membership function of the signal window containing $\text{supp}(F)$, η the diffraction efficiency and E the energy.

When the normalised error:

---

[2] On discretisation, the inner-product is on a space of N×N square summable sequences. The sequences represent samples of one period of the continuous field. The 'computational' diffraction efficiency η as defined here implicitly neglects the energy directed into multiple diffraction orders and their suppression by a sinc^2 envelop due to pixellation. The computed diffraction efficiency remains a useful performance measure but its value cannot be taken as the value of the physical diffraction efficiency. Indeed, the computed diffraction efficiency can exceed theoretical upper bounds on the physical diffraction efficiency of multi-level holograms [57]

$$\sigma^2 = \frac{\left(\chi\left(H-\alpha F\right),\left(H-\alpha F\right)\right)}{\left(H,H\right)} \tag{4.31}$$

is minimised, the ratio of SNR to the diffraction efficiency:

$$\frac{SNR}{\eta} = \frac{\left(H,H\right)}{\left(\chi\left(H-\alpha F\right),\left(H-\alpha F\right)\right)} = \frac{1}{\sigma^2} \tag{4.32}$$

is maximised.

A trade-off exists between SNR and diffraction efficiency. POGED can be tailored to either optimise the SNR or obtain a high diffraction efficiency through adjustment of the filter function.

4.4 Filter Function

The Section focuses on the design of the filter function which impacts on the relative contribution of errors at different locations in the reconstruction of the overall energy function.

Where errors are deemed important, they are assigned a high positive weight within the signal window where they are most detrimental to the quality of the reconstruction. When errors are deemed less important, they are assigned a low positive weight e.g. outside the signal window they have little effect on the quality of the reconstruction but reduce the diffraction efficiency and consequently should not be assigned a zero weight.

The Fourier Transform of the filter is a convolution mask function, the primary purpose of which is in diffusing the quantisation errors across the structure,

producing a filtered hologram that is compared with the continuous target hologram in the calculation of the error.

Ideally the 'hard' filter function should equal unity within the signal window i.e. everywhere the target function is non-zero and takes a small positive value (marked as ϵ in Figure 4.2) outside the signal window to promote higher diffraction efficiency. The Fourier Transform of the discontinuous 'hard' filter function yields a corresponding 'hard' mask of unlimited extent. To reduce computational complexity, it is desirable that the mask has limited extent (a 'soft' mask) and hence the corresponding filter is a smooth continuous function (a 'soft' filter) that approximates the 'hard' filter as closely as possible.

The hard filter can be derived using Equation 4.28 to Equation 4.32. It is also clear that maximising the diffraction efficiency, η, minimises the normalized error E, whereas σ is minimised when the SNR is maximised.

A composite cost function may therefore be constructed according to [56, 68]:

$$\epsilon = \frac{E^2}{E_F^2} + \frac{\sigma}{\sigma_F} = \left(\frac{1}{E_F^2} + \frac{1}{\sigma_F}\right) < \frac{\langle P(H-\alpha F),(H-\alpha F)\rangle}{\langle H,H\rangle} \quad \textbf{(4.33)}$$

where the subscript F denotes the upper limit of the target function.

The hard weight P is then given by:

$$P = \chi + \xi(1-\chi) \quad \textbf{(4.34)}$$

where

$$\xi = \frac{\sigma_F}{(\sigma_F + E_F^2)} \quad \textbf{(4.35)}$$

ξ determines the trade-off between *SNR* (reconstruction fidelity) and diffraction efficiency η; small values of ξ, defined by the condition $\sigma_F < E_F^2$, favour high *SNRs* at the expense of diffraction efficiency η. In this case filter *P* is equal to 1 inside the signal window of the target, *F*, and has a small non-zero value of ξ outside the window. However when ξ = 1, filter function *P* degenerates to a constant over the entire *(x,y)* plane, and the composite error ϵ is equal to the normalized total reconstruction error E^2 (Figure 4.2).

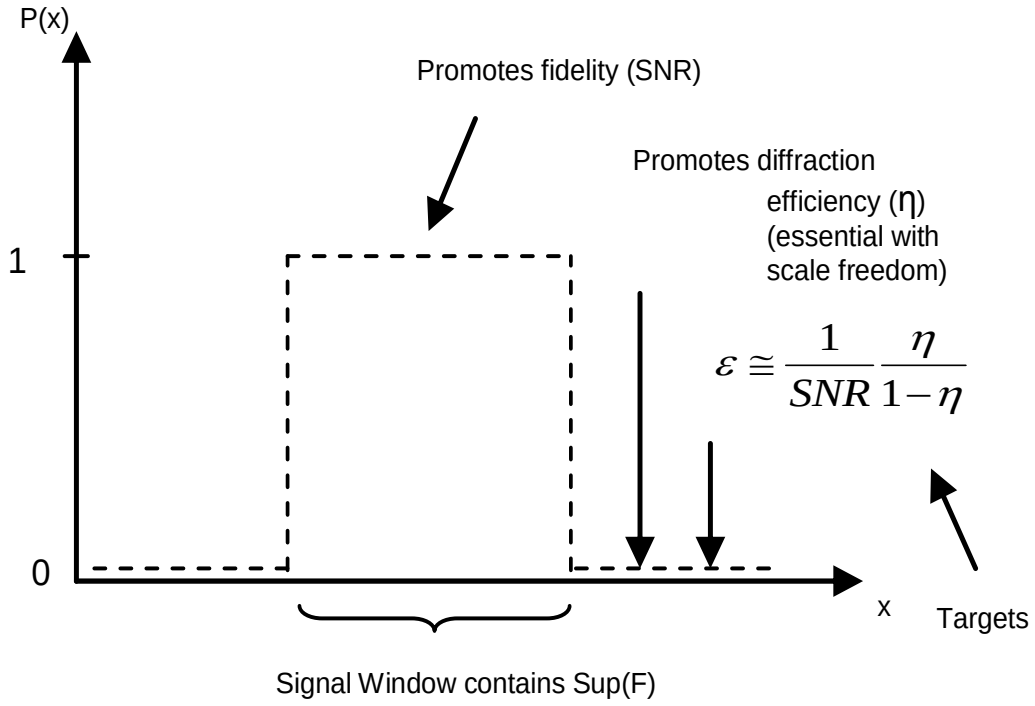


Figure 4.2 Illustration of the hard filter.

In this research, the 'hard' filter function provides a target for the Alternating Projections on Convex Sets (POCS) algorithm [58-59] that identifies a positive function, the Fourier Transform of a mask of defined extent and the closest of all such functions to the target function.

The basic POCS algorithm is performed on two constrained convex sets in order to find an element at the intersection of the two sets; the function is positive and is the Fourier Transform of a band-limited function i.e. the corresponding mask is of finite extent. The soft convolution mask is then calculated by starting with the hard filter, taking the inverse Fourier Transform of the current filter function and by setting to zero all parts of the result that lie outside the desired support of the mask. A Fourier Transform is then performed and all negative values are set to zero. This continues until the function is both positive and band-limited, (Figure 4.3). The two constraints, sequences C_+ and C_L are defined as the 'positivity' constraint C_+:

$$C_+(P_j) = \begin{cases} P_j & if \quad P_j \geq 0 \\ 0 & if \quad P_j < 0 \end{cases} \tag{4.36}$$

and the 'band limiting' constraint C_L:

$$C_L(p_m) = \begin{cases} p_m & if \quad |m| \leq M \\ 0 & if \quad |m| > M \end{cases} \tag{4.37}$$

The POCS algorithm is then theoretically defined as:

$$\begin{aligned} C_+ &= P_B(C_L) \\ C_L &= P_A(C_+) \end{aligned} \tag{4.38}$$

where P_B and P_A are projections into space *A* and space *B*.

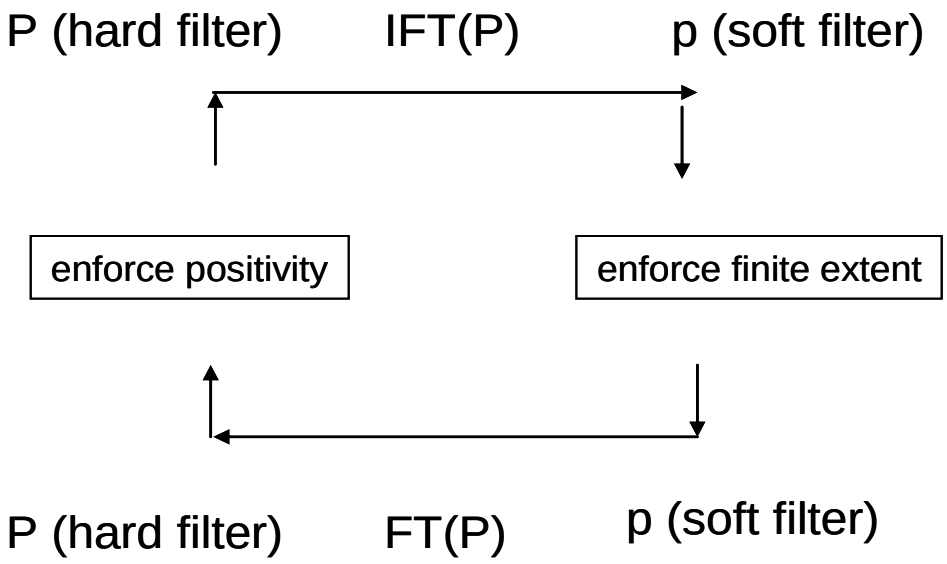


Figure 4.3 Flow diagram of the filter function implementation.

Dykstra's Alternating Projection is used to converge to a solution closer to the starting position [60-61]. Dykstra's algorithm contains the original two sequences, C_+ and C_L, used in POCS along with two additional sequences q_n and p_n theoretically defined as:

$$\begin{aligned} C_+ &= P_B(C_L + q_{n-1}) \\ q_n &= p_j + q_{n-1} - C_+ \\ C_{L+1} &= P_A(C_+ + p_n) \\ p_{n+1} &= p_m + p_n - C_{L+1} \end{aligned} \qquad \textbf{(4.39)}$$

where P_B and P_A are projections into the space A and B.

The first two stages of Dykstra's algorithm are to the same as in POCS, however convergence has not yet been reached and Dykstra's algorithm adds additional stages to reach convergence. After the above stages, the algorithm continues to reach an

even closer point since vector $a_1 + q_0$ is neglected before it can be projected onto B, owing to the outward projecting normal of q_0 .

4.5 Simulation Results

4.5.1 Effect of the Filter Function

The filter function depends on two parameters: the size of the filter and regularisation parameter, ε, the latter allowing control of diffraction efficiency and SNR. A number of trials varying ε and the size are carried out. Figure 4.4 illustrates the hard filter in the form of the support of the target function used in evaluating POGED.

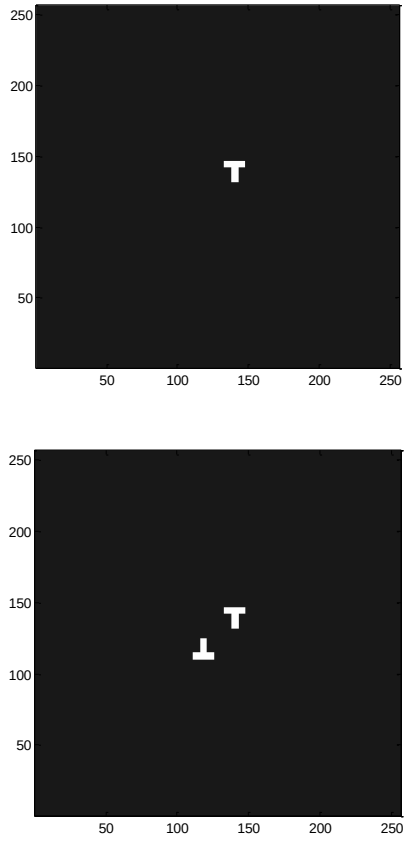
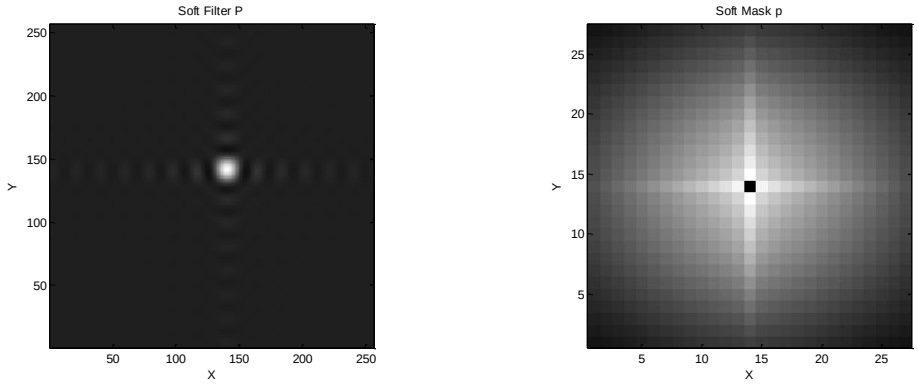


Figure 4.4 (a) Target function for a single order of a binary grating or for a multi-phase grating. (b) Secondary target function for binary gratings to accommodate for the inversion symmetry of a binary grating structure.

The target image used to evaluate the method for multi-phase grating structures and binary gratings is depicted in Figure 4.4(a). However, for a binary grating it is possible to include the second image - arising out of inversion symmetry - into the calculations of the filter function (Figure 4.4(b)) allowing for both components of the reconstructed image to be considered in the calculation of the diffraction efficiency.

Binary gratings are described by a real function; their Fourier Transforms are therefore Hermitian viz. the magnitude is inverted and symmetric. Whichever filter

is used, the same results are achieved but calculation of the diffraction efficiency is incorrect by a factor of two because it does not account for the inverted image. Figure 4.5, is an example of both the hard and soft filter functions calculated using Dykstra's Alternating Projections.



(a) (b)

Figure 4.5 (a) Soft filter function when only single signal window is used (in reconstruction plane) (b) convolution mask with suppressed central peak (done for visual purposes) corresponding to the soft-filter (a) (ε =0.1 and the width, pw = 27).

The mask function can also be used to tailor the diffraction efficiency or the SNR. When the regularisation and size parameters change, the projection algorithm produces very different mask functions which can vary in size and have varying impact on diffraction efficiency and SNR. A number of mask functions with different sizes and different regularisation parameters are generated to evaluate their effects on the quality of the grating structure (Figure 4.6).

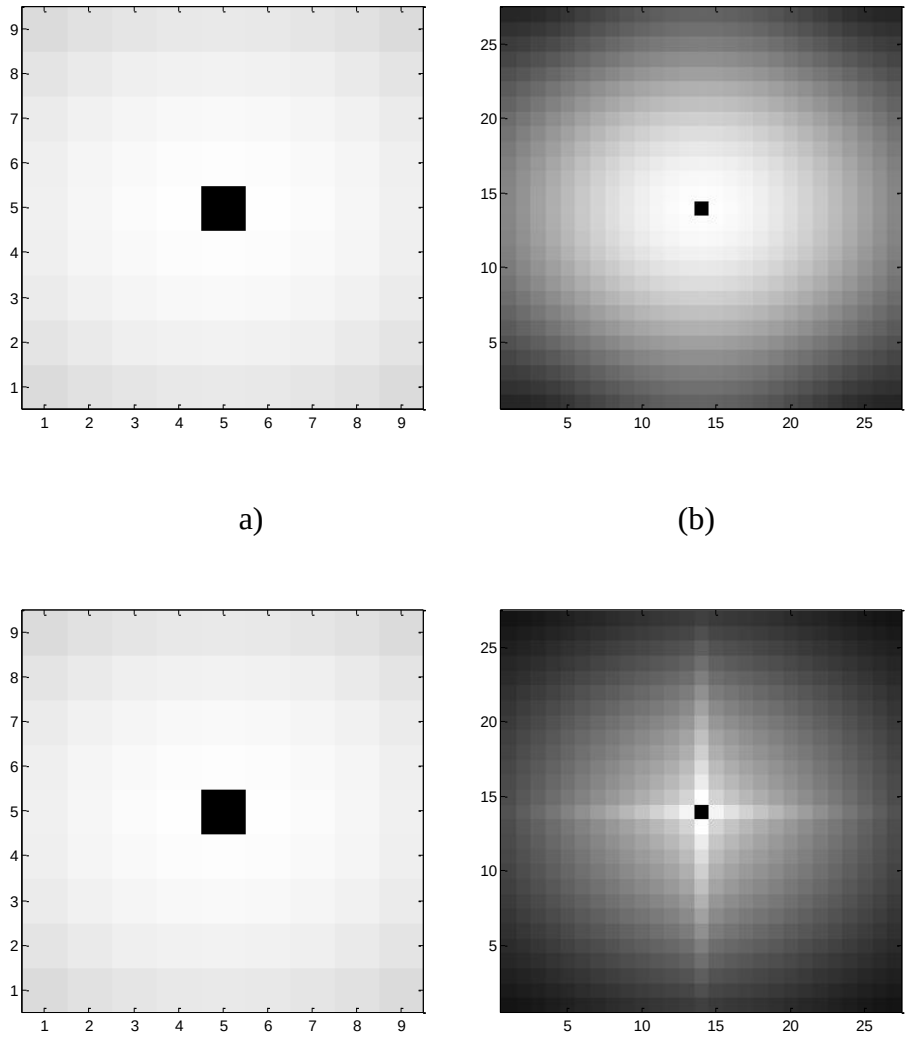


Figure 4.6 Different soft mask functions, *p,* for different regularisation and size parameters; (a) ε is equal to 0.4 and the size of the filter function is 9 (b) ε is equal to 0.4 and the size of the filter function is 27 (c) the size of the filter function is 9 and ε is 0.1 (d) the size of the filter function is 27 and ε is 0.1

4.5.2 Binary Simulations

To design a grating structure using POGED, a computer program was written in the software package Matlab [62], an interpreted software scripting language (not

complied). The simulations are then executed on a Xeon 3.2GHz processor with 3.2Gbytes of RAM under Windows XP. For the purposes of the Thesis, the letter "T" is chosen to demonstrate inversion symmetry; the letter "T" is 21 pixels across the widest section and off-axis. For the regularisation and size parameters of the mask function, one set of values is used throughout except when characterising the effect of varying parameters. The offset, $\varepsilon = 0.1$ and the size, $pw = 27$. The calculation time for an array of 256x256 pixels is one minute for an inverted and symmetric filter function; the diffraction efficiency obtained is 72% and the SNR is 32dB. When the filter function does not account for the twin image of the binary hologram, the computation time is over a minute with a diffraction efficiency of 36% and SNR of 33dB. Because the method does not yield an unique solution, all of these values can vary by a small amount when the algorithm is seeded from different starting configurations since the method devotes a significant amount of time correcting a small number of pixels. Specifying a different stop condition such as a tolerance on the cost function instead of the current condition of no more pixel flips, further improves calculation times; however this could result in a drop in the diffraction efficiency of 1%-2% and a 1dB-2 dB loss in SNR

The results of the simulations shown in Figure 4.7 are for an off-axis "T" grating structure. The structure demonstrates distinct white and dark lines representing high and low segments of the grating surface relief.

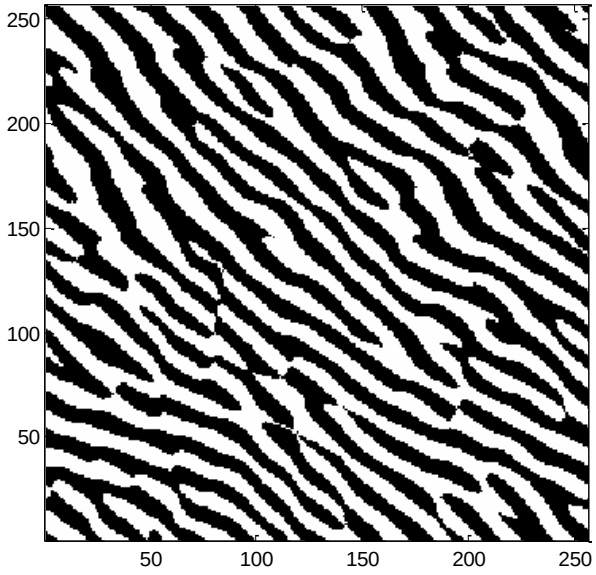


Figure 4.7 Grating structure for the letter “T” generated using POGED.

The reconstructed image of the binary grating structure of Figure 4.8 shows clearly the off-axis “T”. The images show good contrast with respect to the background, illustrating the high SNR of 32dB. The errors surrounding the reconstructed images are not dominant and appear like “clouds” of no distinct shape.

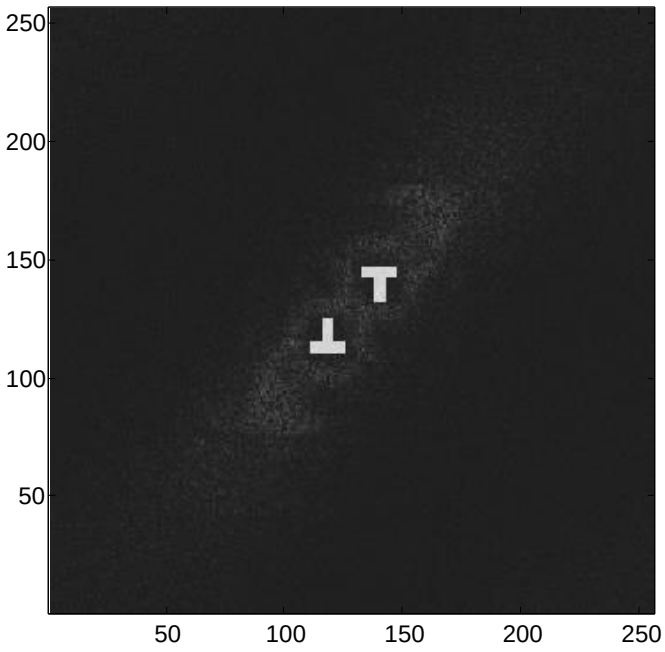


Figure 4.8 Reconstructed image of the letter “T” illustrating both the desired reconstructed image and its inversion symmetric partner.

Figure 4.9 shows the evolution of the cost function as a function of the number of iterations. The progress of the algorithm is similar to SA - Chapter 3, Section 3.2 - without the uphill portion to escape any local minimum. The other aspect to note is that most of the improvements are made in approximately the first 50 iterations and the largest fraction of the computation time is spent on correcting a small number of pixels.

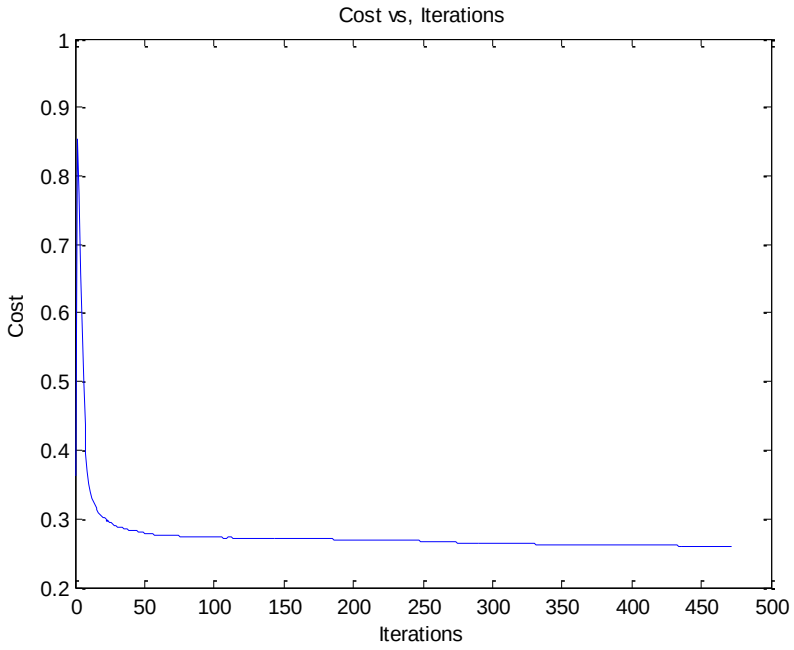
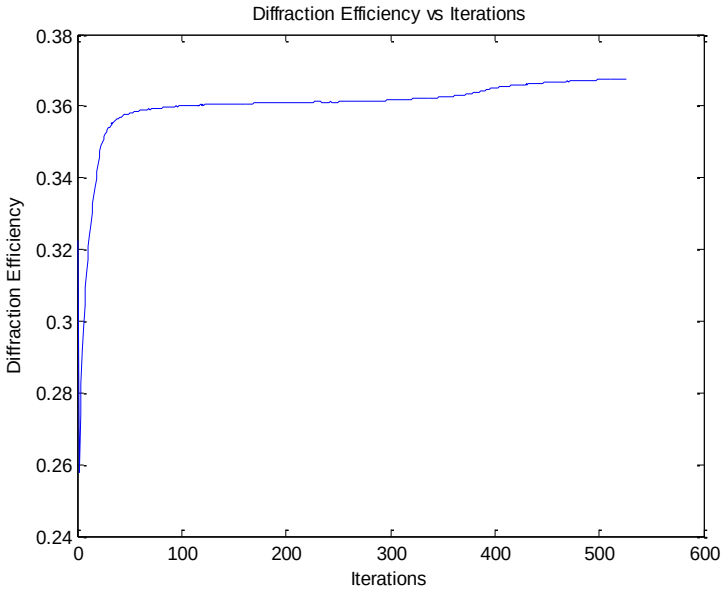
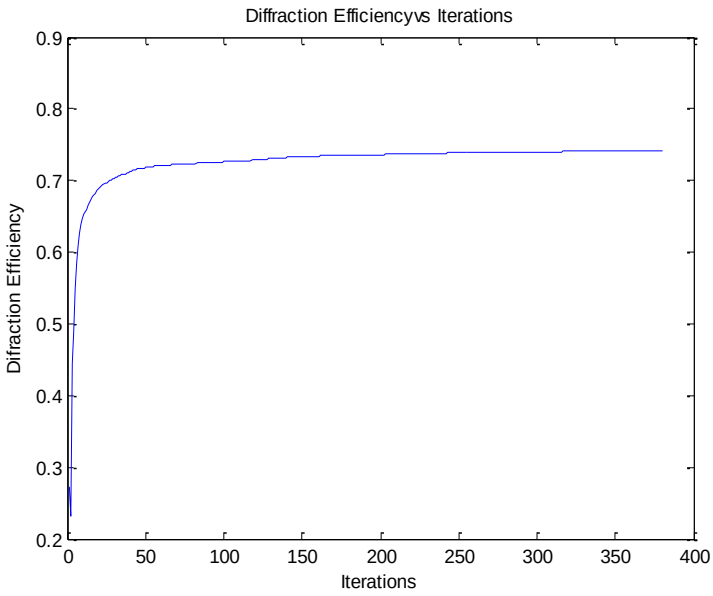


Figure 4.9 Evolution of the cost as the algorithm progresses. Most of the improvements are made in the first 50 iterations.

Figure 4.10(b) illustrates how the diffraction efficiency evolves as the algorithm progresses. In the first 50 iterations, a large increase in the computed diffraction efficiency is evident, levelling out for the remaining iterations, attributed to the algorithm trying to correct a small number of pixels to further improve diffraction efficiency. Also evident on inspection of Figure 4.10 is that the largest increase in the diffraction efficiency occurs at the same time, whether or not the inverted image is included in the calculation of the filter function.



(a)



(b)

Figure 4.10 (a) Computed diffraction efficiency when the filter function is derived from the support of the target image. (b) Diffraction efficiency when the filter function is derived from the support of the target image and its inversion symmetric parameters.

The diffraction efficiency is 36.6% (Figure 4.10(a)) and if the energy in the twin image is neglected, the diffraction efficiency is 72% (Figure 4.10(b)).

In the absence of a unique solution, it is possible to arrest the algorithm by establishing a tolerance value in the error function. Although ending the algorithm early would substantially decrease computation time, some loss of SNR and a slight reduction of in the diffraction efficiency results.

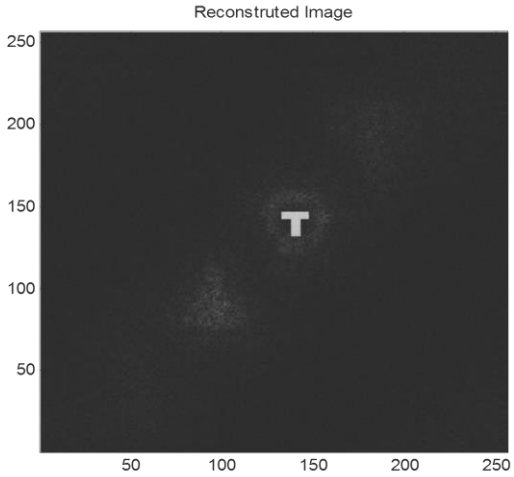
4.5.3 Multi-Phase Simulations

Quaternary and continuous phase grating structures are simulated to demonstrate multiphase POGED operation. This evaluation requires more time to implement because the algorithm no longer simply flips between two pixels but considers the phase step between the two states. "T" is again used in the simulations carried out on the same computational resources.

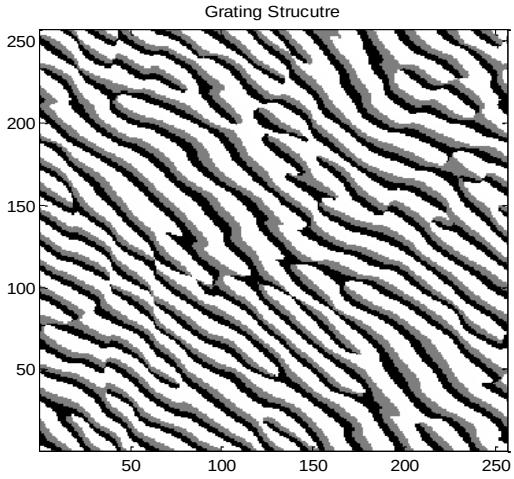
The quaternary grating structure is firstly implemented on a 256x256 pixel array. The computation time required is 5 minutes and 56 seconds, yielding a diffraction efficiency of 74% and SNR of 32dB. The performance does not improve on comparison with the binary case, except that it is effective in suppressing the impact of the twin image.

The reconstructed image for the quaternary grating shows the "T" structure visible in the upper right quadrant, close to the origin (Figure 4.11(a)). Because the grating is

not continuous and is still close to the binary case, some errors are evident where a twin image would normally be present (Figure 4.11(b)).



(a)

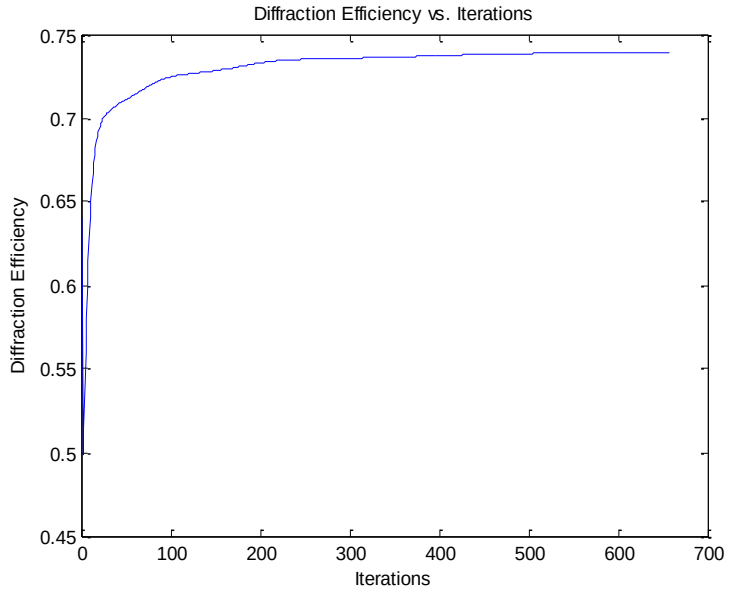


(b)

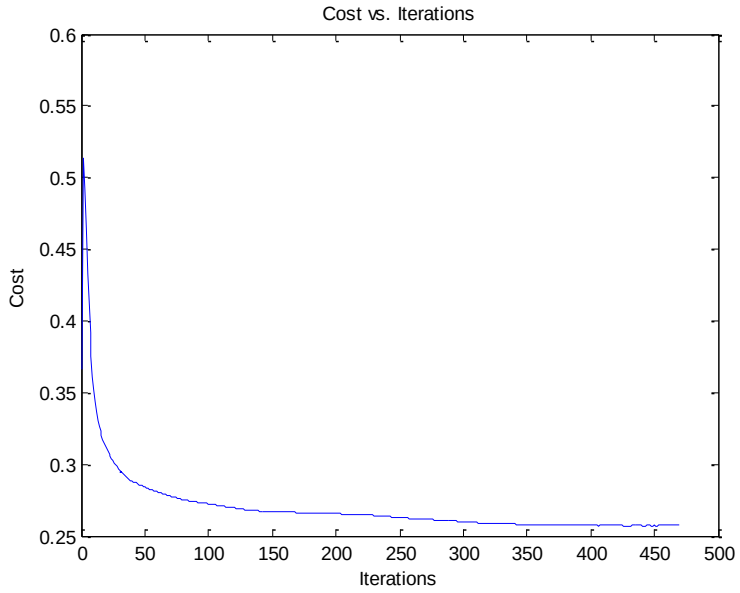
Figure 4.11 **(**a) Grating structure for a quaternary grating of an off-axis "T". (b) Reconstructed image of a quaternary grating structure of an off-axis "T".

The computed diffraction efficiency for the quaternary is marginally higher than the binary structure viz. 74% compared to 72%. The evolution of the diffraction

efficiency with iterations in the quaternary case follows the same pattern as that of the binary however, most of the rise in diffraction efficiency occurs in the first 100 iterations, again owing to the absence of an unique solution and the possibility of screw dislocations (vortices) existing in the grating structure (Figure 4.12(a)) [63]. The evolution of the cost function follows the same profile as that of the binary case where an initial sharp descent occurs as the iterations progress. The stagnation evident in the binary case is also observed after the same number of iterations (Figure 4.12(b)).



(a)

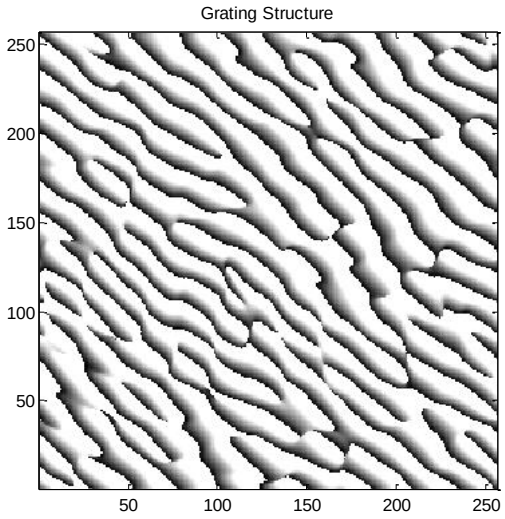


(b)

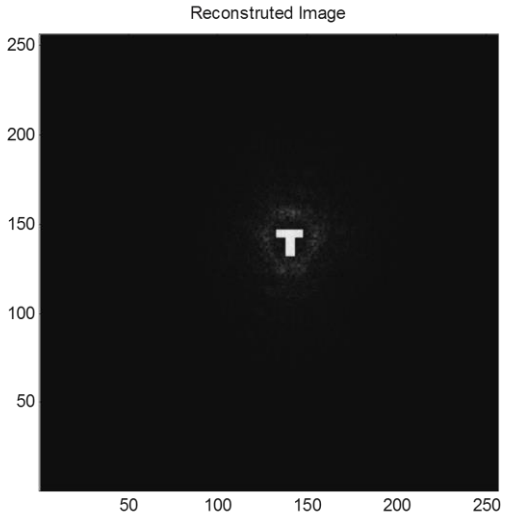
Figure 4.12 (a) Diffraction efficiency as a function of the number of iterations.(b) Cost as a function of the number of iterations.

4.5.4 Continuous Phase Simulations

For continuous phase holograms, the reconstructed image is the solid “T” structure with a few surrounding errors and no significant errors where the twin image should be present (Figure 4.13).



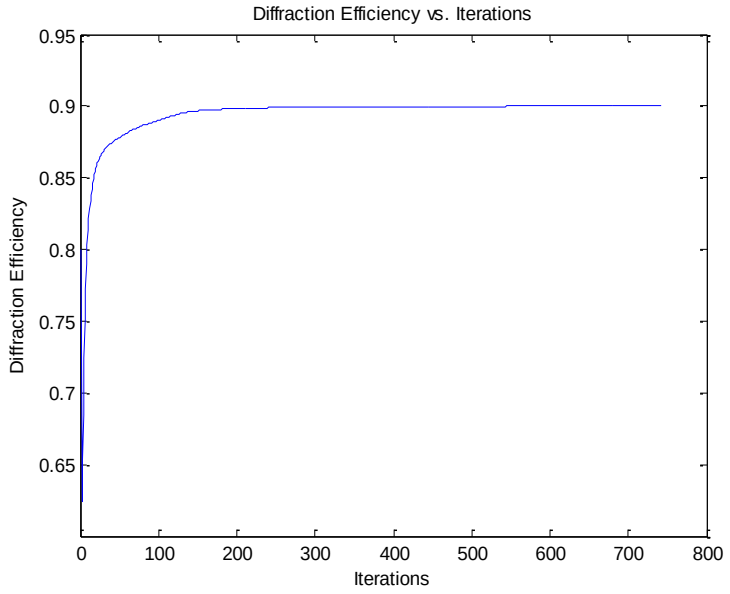
(a)



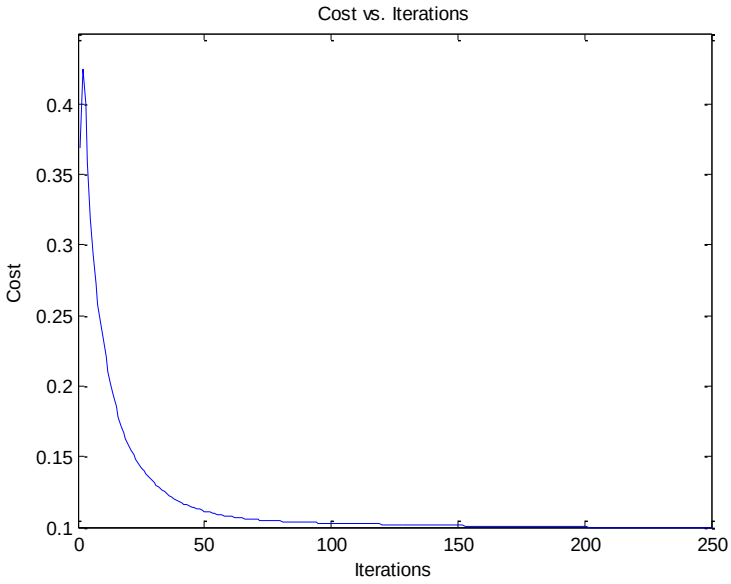
(b)

Figure 4.13 (a) A continuous phase grating. (b) The reconstructed image.

For a continuous phase structure the stagnation observed in both the binary and quaternary cases also occurs. All of the improvements are gained in the first 100 iterations, after which only a few pixels are corrected resulting in small improvements to the diffraction efficiency (Figure 4.14(a)) and cost (Figure 4.14(b)).



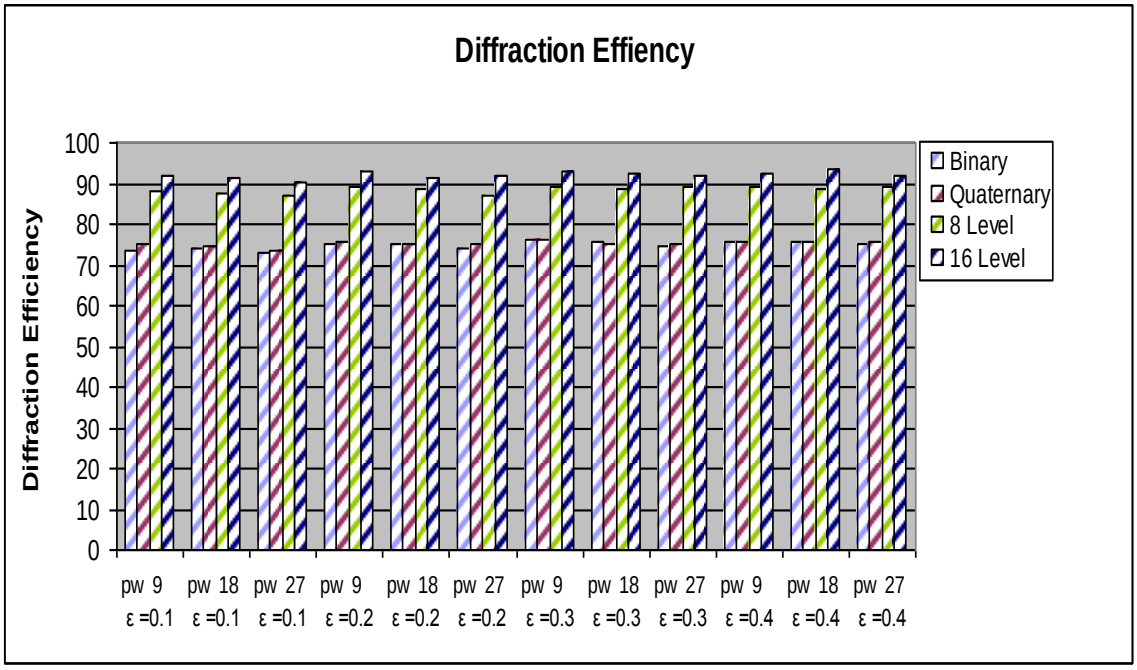
(a)



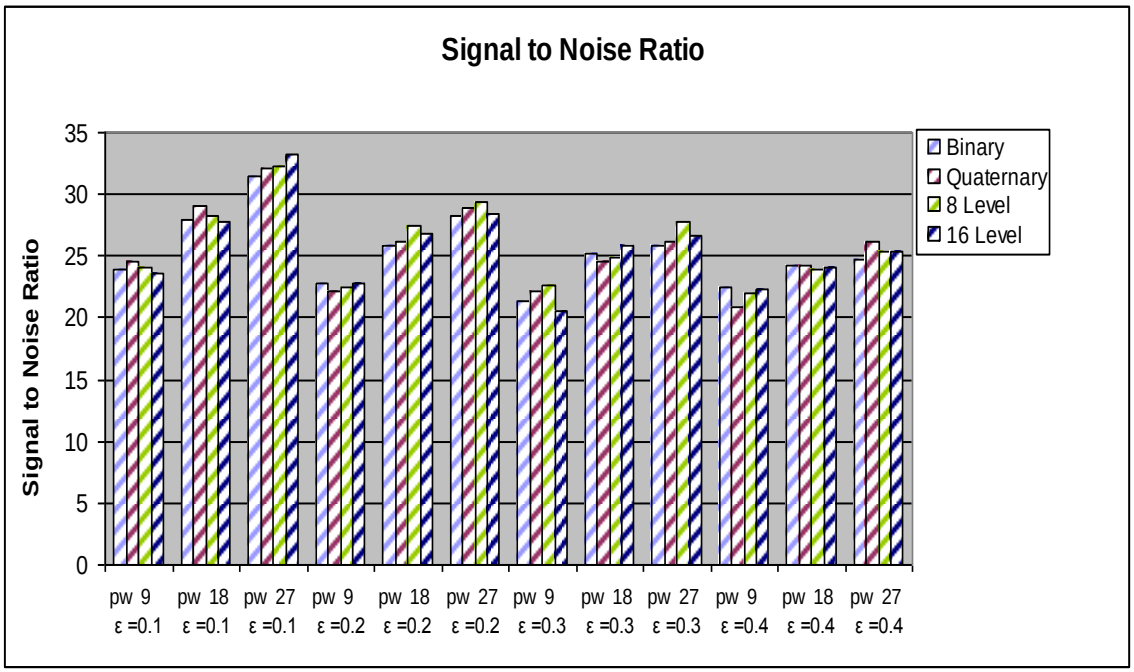
(b)

Figure 4.14 (a) Evolution of the diffraction efficiency as a function of the number of iterations. (b) Evolution of the cost as a function of the number of iterations.

As the size and the regularisation parameter of the mask are varied, some distinct trends emerge. When the size of the mask is increased, the SNR improves however the diffraction efficiency does not change significantly (Figure 4.15). When the regularisation parameter value increases, a small decrease in the SNR and a small increase in the diffraction efficiency occurs. Changes in the size of the mask do not affect overall calculation times of the algorithm leading to the decision on the size of the mask solely for optimization of the SNR or of the diffraction efficiency and not calculation times



(a)



(b)

Figure 4.15 Comparison between different sizes of the filter function and the offset value. (a) Diffraction efficiency (b) SNR.

Tailoring the algorithm for higher diffraction efficiency (or SNR) through changing the regularisation parameter value is not an effective approach. Changing the size of the mask to promote SNR however does yield good results when the filter is large and the regularisation parameter is small.

4.6 Conclusions

Figure 4.15 summarises the simulation results obtained for grating structures with different numbers of quantisation levels and the filter function. Results show that the size of the filter function and the offset have an impact on the SNR; the largest filter and the smallest offset provide the best SNR for any quantisation level. The computed diffraction efficiency does not change significantly as a function of different size and offset values. Consequently the optimum strategy is to tailor the filter function to maximise the SNR since it has negligible effect on diffraction efficiency. It is worth noting - bearing in mind that these values cannot be taken as the theoretical diffraction efficiency as indicated in Footnote 2 - that the computed diffraction efficiency is close to theoretical values (Table 4.1).

Table 4.1 Comparison of computed diffraction efficiency to theoretical diffraction efficiency.

Grating Type	Time	Computed Diffraction Efficiency (%)	Theoretical Diffraction Efficiency (%)
Binary	1:00 min	36	40.5
4 Level	5:56 min	74	81
8 Level	19:00 min	87	90
16 Level	27:00 min	90	99

The execution times vary from one minute to just under thirty minutes depending on the number of levels in the grating structure. Depending on the desired application, a solution can be found that offers the best compromises between time, SNR and diffraction efficiency

Chapter 5

Fresnel Region Phase Optimised General Error Diffusion

5.1 Introduction

The Chapter focus is on the development and modifications required for POGED to operate in the near-field region. Arguments as to why the defined theoretical approach is not appropriate as is are made and the adoption of the angular spectrum method is proposed and proven to circumvent the limitations.

5.2 Fresnel Region

A Fresnel Transform is a beam propagation method in the near field region, where the near field region is the region where z is small and thus scalar diffraction theory is valid [38]. Operation in this region is used in optical beam shaping [18-22], and in optical beam steering when propagation distances are small compared to the pixel size and less than the focal length of the system [64-68]. POGED is adapted to implement the Fresnel Transform through modifications to the algorithm.

The theoretical formulation to execute the Fresnel Transform is given by;

$$u_2(x_2, y_2) = \frac{\exp(ikz)}{i\lambda z} \times \int_{-\infty}^{\infty} u_1(x_1, y_1) \exp\left\{\frac{ik}{2z}\left[(x_1 - x_2)^2 + (y_1 - y_2)^2\right]\right\} dx_1 dy_1 \quad \textbf{(5.1)}$$

where $u_2(x_2,y_2)$ is an amplitude distribution at a distance z from the input plane with input signal, $u_1(x_1,y_1)$, illuminated by a plane wave at wavelength λ.

The amplitude distribution in the output plane can be rewritten as:

$$u_2(x_2, y_2) = \frac{\exp(ikz)}{i\lambda z} \exp\left\{\frac{ik}{2z}\left(x_2^2 + y_2^2\right)\right\} \times \Im\left(u_1(x_1, y_1) \exp\left\{\frac{ik}{2z}\left(x_1^2 + y_1^2\right)\right\}\right)\Big|_{v_x, v_y} \quad \textbf{(5.2)}$$

where $\Im$ is the Fourier Transform operator and v_x and v_y are the spatial frequencies defined as $v_x = x_2/\lambda$ and $v_y = y2/\lambda$.

Since the coordinates in the input (x_1,y_1) and output planes (x_2,y_2) are not equivalent and the output coordinates for the spatial frequencies require a scaling by $1/z$, POGED requires an interpolation and re-sampling [65-66,68]. The z factor in the quadratic phase factor must also be examined further;

$$\exp\left(\frac{ik}{2z}\left(x_1^2 + y_1^2\right)\right) \quad \textbf{(5.3)}$$

The z factor in the denominator leads to rapid oscillations as small variations in z lead to large phase changes. To correct this, significant sampling of the input field is required [65-68], inferring that the theoretical method for performing the Fresnel Transform is not applicable to digital holography.

An alternative way of implementing Fresnel Transforms is to use the convolution approach relying on the decomposition of the input field into a number of plane waves. The complex output field can then be represented as a convolution of the input field and an impulse response function *h(x,y)*;

$$u_2(x,y) = u_1(x,y) \otimes h(x,y) \quad \textbf{(5.4)}$$

where function *h(x,y)* is defined as;

$$h(x,y) = \frac{\exp(ikz)}{i\lambda z}\exp\left\{\frac{ik}{2z}\left(x^2 + y^2\right)\right\} \quad \textbf{(5.5)}$$

Using the properties of the Fourier Transform, the convolution of the two functions becomes a multiplication in the object plane:

$$U_2(v_x,v_y) = U_1(v_x,v_y)H(v_x,v_y) \quad \textbf{(5.6)}$$

where U_2,U_1 and H are the Fourier Transforms of u_2, u_1 and h.

The Fourier Transform of the function h is now defined as:

$$H(v_x,v_y) = \exp(ikz)\exp\left\{-i\lambda z\left(v_x^2 + v_y^2\right)\right\} \quad \textbf{(5.7)}$$

H now describes the retardation of the phase for each plane wave depending on the propagation angle and distance z.

As both the input and output planes are now equivalent, no interpolation or re-sampling of the fields is required and the phase transfer function is now linearly dependent on z; consequently small changes in z no longer represent rapid but dampened oscillations in frequency. This is a compromise which allows for the correct sampling of the quadratic phase factor. A critical distance, Z_c, which depends on the pixel, DOE, image size and wavelength applies. The integral method needs to be used when $Z<Z_c$ and the convolution approach when $Z > Z_c$.

Fresnel POGED

Using the above formulation allows the necessary modification to POGED to be made:

$$U(x,y) = Fr\{u(x,y)\} = \Im^{-1}\{q(u,v;z)\Im\{u(x,y)\}\} \quad \textbf{(5.8)}$$

$$u(x,y) = Fr^{-1}\{U(x,y)\} = \Im^{-1}\{q^{*}(u,v;z)\Im\{U(x,y)\}\} \quad \textbf{(5.9)}$$

where $\Im$ and $\Im^{-1}$ represent the forward and inverse Fourier Transforms, q^* is the complex conjugate to q, and the quadratic phase factor $q(u,v,z)$ is defined in frequency space (u,v) dependent on wavelength λ and propagation distance z as [61-62]:

$$q(u,v;z) = \exp\left(i\pi\lambda z\left(u^2 + v^2\right)\right) \quad \textbf{(5.10)}$$

The key variables are the wavelength λ, the propagation distance z and the pixel pitch at $z=0$ and $z=Z$.

Since the convolution is no longer valid in the Fresnel region, the variables $p \otimes f$ and $p \otimes h$ need to be redefined as;

$$ph = p \otimes h, \quad pf = p \otimes f \quad \textbf{(5.11)}$$

In addition to the propagation operator becoming a Fresnel Transform, a number of other changes need to be made when the scaling factor or the function $p \otimes h$ is updated. The update for the scaling function follows:

$$\alpha^{new} = \alpha^{old} + (h_{ij}^{new} - h_{ij}^{old}) \frac{pf^{*}}{\langle pf, f \rangle} \quad \textbf{(5.12)}$$

and the update for $p \otimes h$ follows:

$$ph^{new} = ph^{old} + (h_{ij}^{new} - h_{ij}^{old})\, Fr^{-1}\{P\, Fr\{e_{ij}\}\} \quad \textbf{(5.13)}$$

Updates for $p \otimes h$ now involve a Fresnel Transform performed on every pass throughout the array, resulting in the slowing down of the algorithm. This is caused by the fact that the convolution is no longer valid in the Fresnel region. In the Fourier region it was possible to simplify the update to $p \otimes h$; however in the Fresnel region $p \otimes h$ can no longer be simplified resulting in the pre-calculation of $Fr^{-1}\{P\, Fr\{e_{ij}\}\}$ to be stored in a lookup table. Computer memory can then become a limitation, especially for 32-bit computers. Other ways to speed up the calculation may be pursued but due to limits in time, they have not been investigated.

5.3 Fresnel POGED Simulation Results

The angular spectrum method is used to simulate the Fresnel Transform since it is more resilient to high frequency oscillations compared to the integral method [61-62]. Results for the Fresnel Transform are generated using a Gaussian beam transformed into a flat top output beam (Figure 5.1) using a 16-level grating structure generated on a Xeon 3.2GHz processor with 3.2Gbytes of RAM under Windows XP (a 32-bit computer with a 3-bit version on Windows XP allowing the implementation of a look up table). The amount of memory available placed a limit on the size of the grating structure; thus a 128x128 grating array structure is used. A comparison to the Gerchberg-Saxton Algorithm is made to verify the validity of the proposed method.

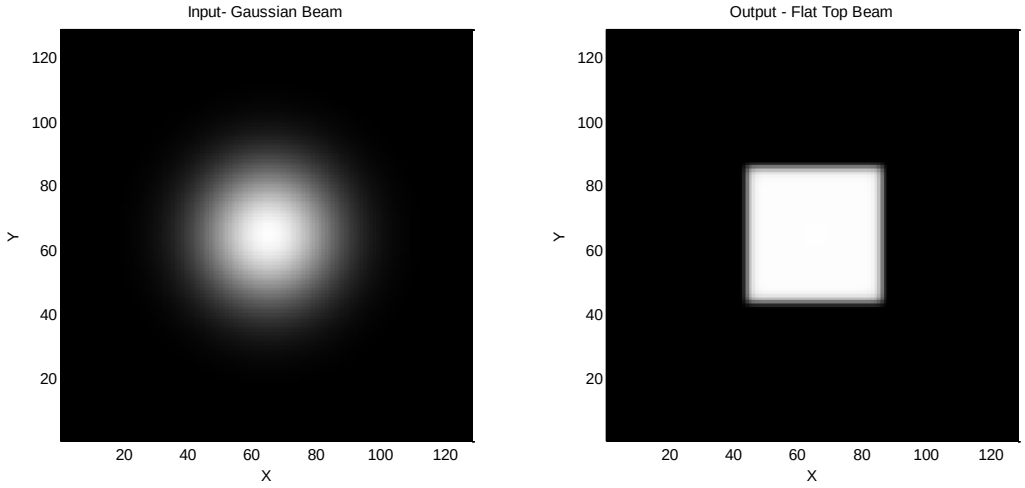


Figure 5.1 A Gaussian input beam is transformed into a flat top beam using POGED in the near field.

The design parameters to create the necessary output beam are as follows: beam wavelength = 632.8 nm, initial ($z=0$) beam *1/e* amplitude radius of 1mm, distance to target plane $z=Z$=496 mm. Hologram h has M=16 discrete phase levels and is simulated as a 128x128 square array with side width of 8 mm, giving a pixel size of 62.5 microns. The target amplitude F_0 considered is a 2x2 mm square. However, when the grating structures are coded into the algorithm, a dimensionless quantity for the z value is used.

Figure 5.2(a), illustrates the phase function that generates a flat top beam from a Gaussian beam. Because the propagation distance is in the near field, the shape of the reconstructed image is embedded in the phase function. The reconstructed image that corresponds to the phase grating, (Figure 5.2(b)) exhibits a square shape with few errors surrounding the image governed by the ε value. There are some errors

inside the image - black lines - since at this value of ε, the diffraction efficiency is being promoted to the detriment of the SNR.

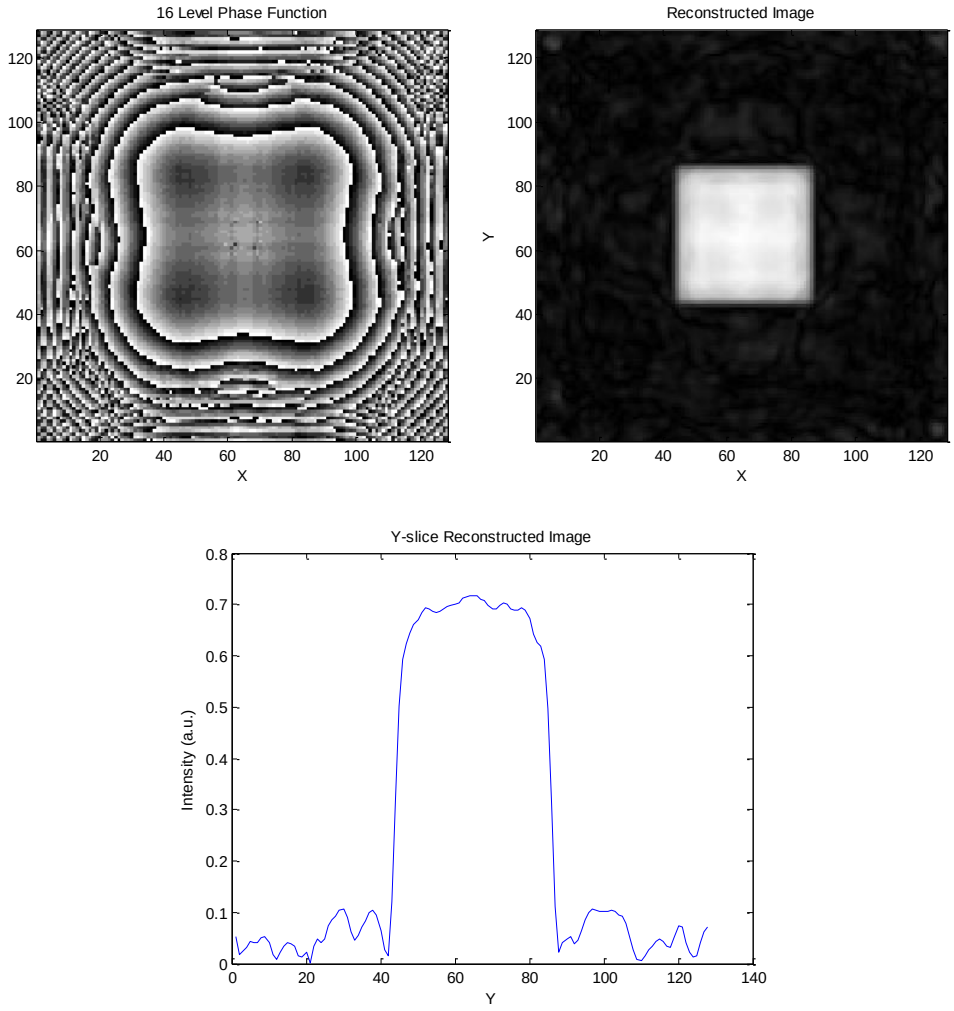


Figure 5.2 (a) Phase function that shapes a Gaussian beam to a flat top beam. (b) The flat top beam produced by the corresponding phase function with (c) as the y-slice of the reconstructed image.

The performance metrics used to evaluate the grating structure are again the diffraction efficiency and SNR. Figure 5.3(a) shows the energy function as it evolves whilst Figure 5.3(b) depicts the evolution of the diffraction efficiency. For both metrics, the bulk of the improvements are completed within the first ten

iterations. If calculation time is deemed more important, the algorithm can be modified to exit when the change in energy is less than a specified tolerance.

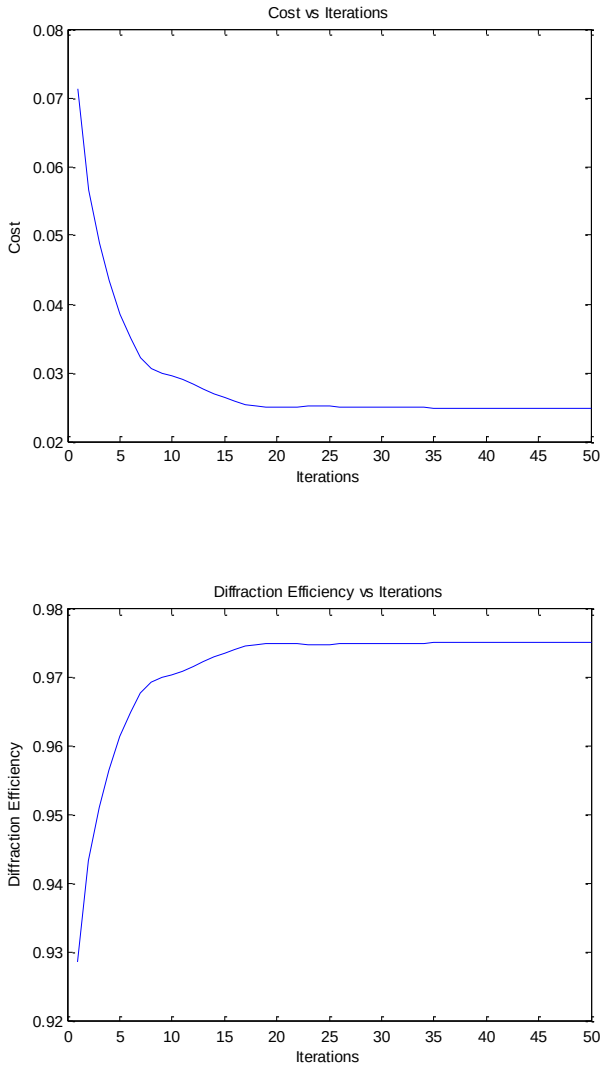
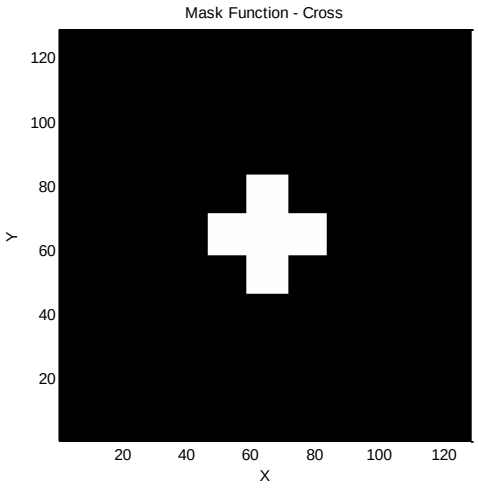


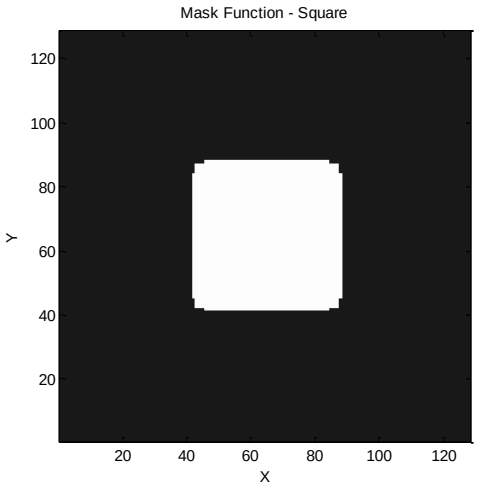
Figure 5.3 (a) Evolution of the cost function and (b) of the diffraction efficiency as a function of the number of iterations.

The mask function for the Fresnel POGED can be defined in two ways. The filter function can use the target image to define the hard filter which in turn is used to calculate the soft filter, an example is depicted in Figure 5.4(a) where a cross is used as the output beam, labeled as P0 for all the analysis. The other method for defining

the filter function is to use a square enclosing an area larger than the target image (Figure 5.4(b)), labeled as P1 in future analysis These two methods were evaluated determining which shape of filter would produce optimal results for the SNR and diffraction efficiency.



(a)



(b)

Figure 5.4 Examples of different hard masks. (a) A hard mask function in the shape of the output image. (b) Hard mask that encompasses an area larger than the target image that is in the shape of a square

When calculating the filter function two variables can be changed; the values of ε, the offset from zero, or the size, pw. Variation of ε tailors the SNR and the diffraction efficiency whilst varying the size of the filter function, changes the time owing to the necessity of using a lookup table to store $p \otimes h$ and the precision of the algorithm. Figure 5.5 shows the filter function for ε =0.2 on a 128x128 array, the total size of the grating structure. The hard filter is the function used to generate the soft filter using Dykstra's Alternating Projection, described in Chapter 4 Section 4.4. A few ripples are evident in the soft filter directly attributable to the value of ε which in turn governs SNR and diffraction efficiency.

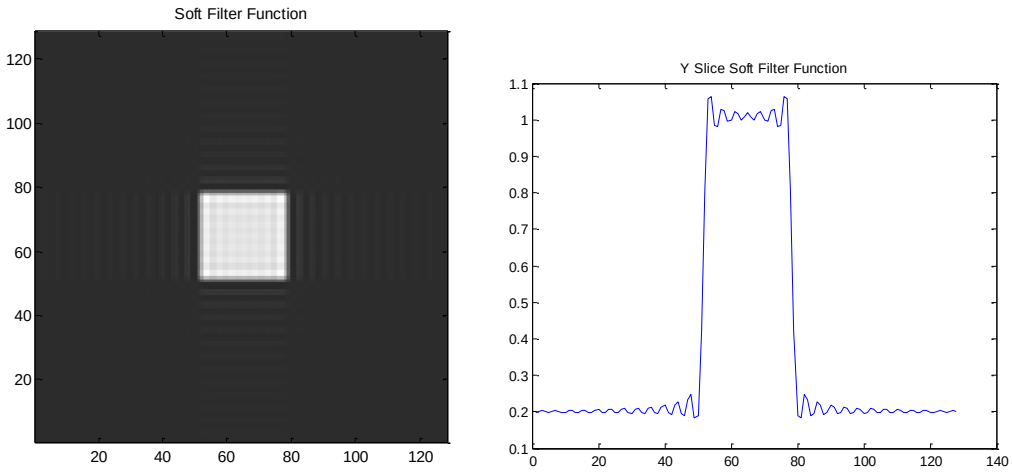


Figure 5.5 (a) Soft filter function (b) Y-Slice of soft filter function, plotted on a 128x128 array.

Significant ripples start to appear when the grating structure width of the filter function is halved and ε is set at 0.2, causing a degradation in the SNR and an increase in the diffraction efficiency (Figure 5.6). If a small degradation in SNR and diffraction efficiency can be tolerated and time is a more important parameter, then a smaller filter function is acceptable.

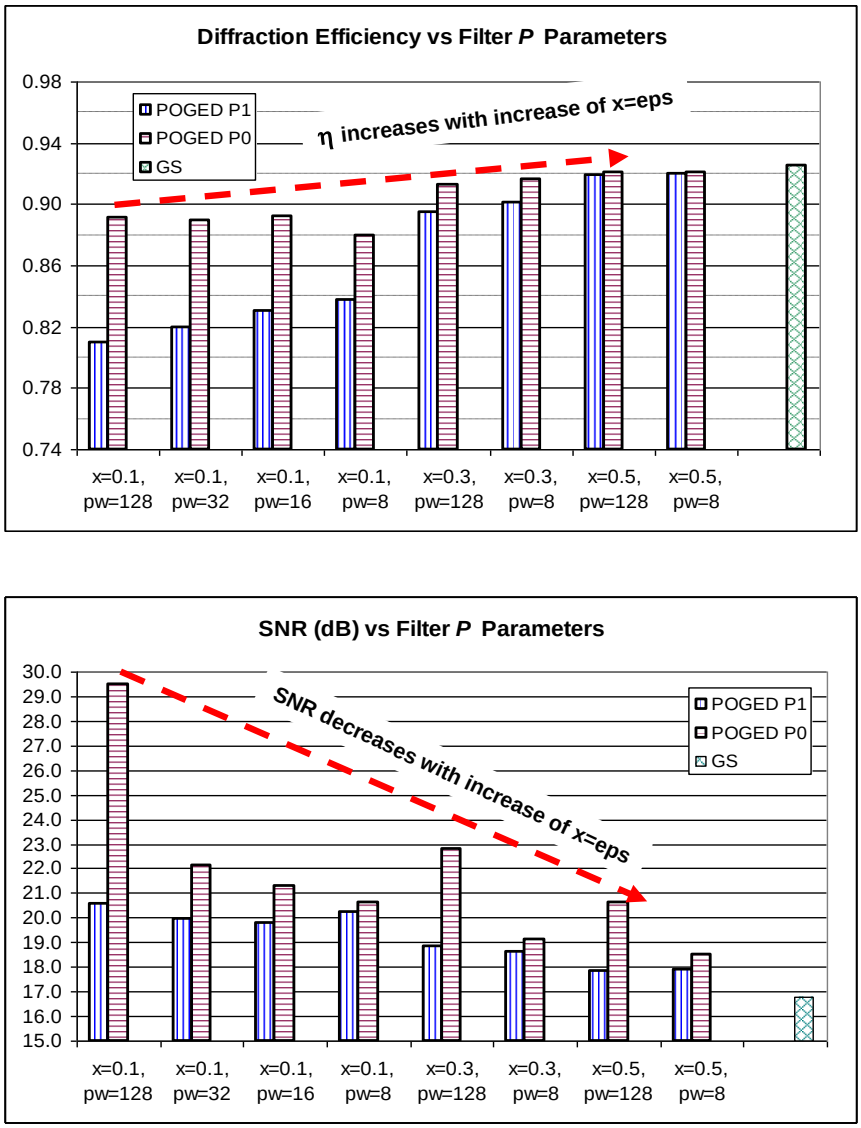


Figure 5.6 Demonstration of the effects of the size of the filter and the offset on (a) diffraction efficiency and (b) SNR

POGED is compared to the single stage Gerchberg-Saxton (GS) algorithm in Figure 5.6. The GS algorithm is used for comparison purposes since it is a relatively straightforward formulation and results can be generated in a comparatively short period of time. POGED has not to date, been effectively implemented in the near-field region and GS is a valid alternative for the purpose of performance comparison.

The GS algorithm, for the same variables and the same number of levels, produces a lower SNR than obtained by POGED whilst the diffraction efficiency is comparable.

5.4 Conclusions

The POGED algorithm has been successful adapted to operate in the near-field/Fresnel region. The modifications to the algorithm are necessary to treat the high frequency oscillations observed owing to the *1/z* term in the quadratic phase factor and in how the update of $p \otimes h$ is performed. Updates are time consuming and the use of a lookup table was successful in reducing the calculation time; the demonstration of the principle was nevertheless restricted by available computer memory. The update process will further reduce the calculation time if this hardware restriction is removed.

The best results for the SNR occur when the filter function is large and the offset is small. This trend is a distinct property of POGED, clearly observed in the near field implementation also. The characteristic leads to the ability to tailor the DOE to the performance required for the overall system. If diffraction efficiency is important, than the offset needs to be large with a small filter and if SNR is important the offset should be small and the filter large.

Chapter 6

Fabrication of Diffractive Optical Elements

6.1 Methods

The primary methods used to fabricate DOEs are focused ion beam lithography [74], electron beam (e-beam) lithography [73], laser ablation [75], and photolithography [69-70]. A sub-category of photolithography is 'step and repeat' lithography, one of the methods used to produce the grating structures in this work. An overview of the lithographic processes is outlined in the Chapter; a more detailed review of the process for photolithography can be found in [69] and for electron beam lithography in [66]. Each method has a number of disadvantages, and a comparison of the techniques is provided to illustrate why some methods are better suited for specific applications.

6.1.1 Fabrication Process

The Section describes the fabrication of the grating structures generated using POGED. For simplicity and governed by the ease of access, two lithographic methods are used.

Figure 6.1 illustrates the important steps of the fabrication process [40, 69-70]. The first step is the selection of the wafer, either quartz initially coated with chrome or a polished wafer of Si or GaAs covered either by an oxide or nitride layer. Glass is

used for transmission whereas a semiconductor substrate can be used effectively in transmission mode if the design wavelength is in the infrared, since in this spectral region, semiconductor material is transparent [76].

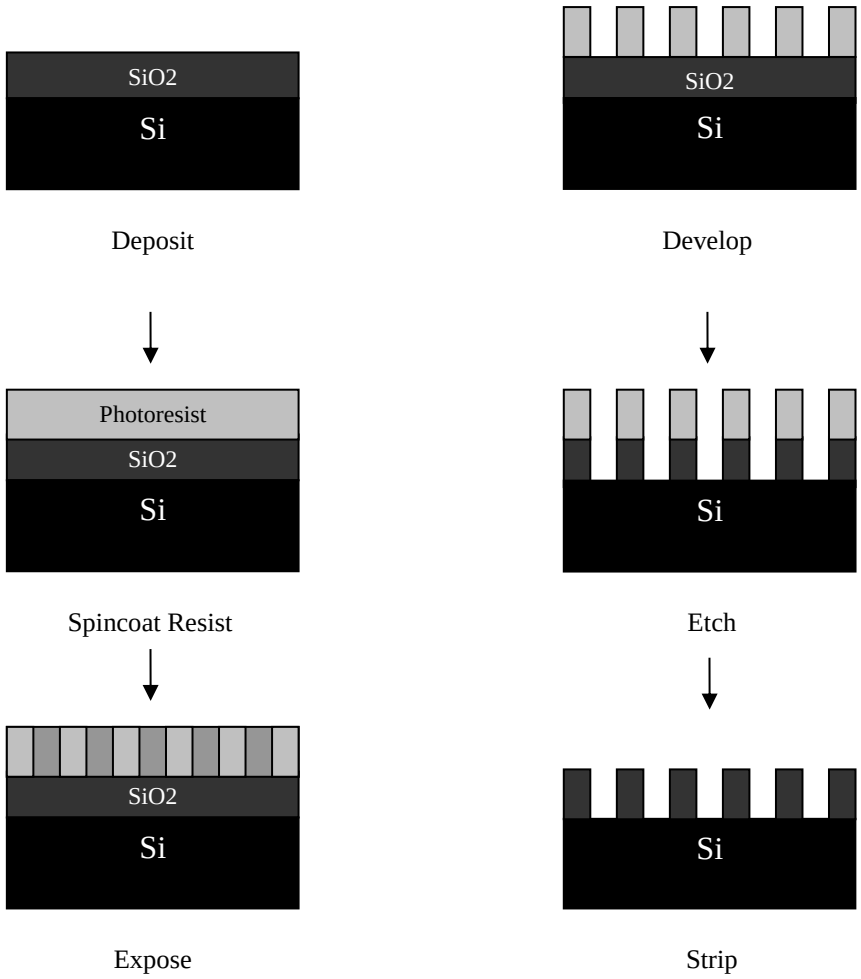


Figure 6.1 General outline of the steps involved in the fabrication of a grating structure by a lithographic method. Steps involved are deposit photoresist, expose with an energy source, develop, etch and then strip away photoresist.

Once the wafer material is selected, the surface is cleaned and stripped of all contaminants. A photoresist layer is spin coated onto the wafer after the deposition of an oxide layer.

Next in the sequence is the exposure step executed by photolithography or electron beam lithography. These lithography techniques are chosen primarily governed by the ease of access; other methods are thus not reviewed. After exposure, the photoresist is developed, removing the exposed - or unexposed - resist as appropriate to the type of resist. The wafer is then baked in an oven to remove residual solvents which also brings an added advantage of hardening the resist that remains to better withstand the etching process. Once baked, the wafer is etched by reactive ion (dry) or by acid etching (wet). The photoresist is then removed and protective coatings are applied in preparation for dicing or cleaving.

6.1.1.1 Spin Coating

The photoresist creates a relief structure on the substrate, central in subsequent processing steps. The photoresist Poly Methyl Methacrylate (PMMA) is exposed with a lithographic method and developed, forming a "template" structure directly on the sample, defining the areas to be etched. Spin coating is performed by placing a drop of liquid resist on a clean substrate and spinning the substrate at speeds in the range of 2000 rpm-4000 rpm to produce an even coating. The rate at which the substrate rotates determines the layer thickness deposited on the surface. Subsequent baking in an oven ensures that no solvent remains within the PMMA.

There are two types of photoresist viz. positive or negative; the former dissolves, whereas the latter remains after the development process. Either type of resist can be used, the choice dependent on proximity and under-cutting effects [69-70].

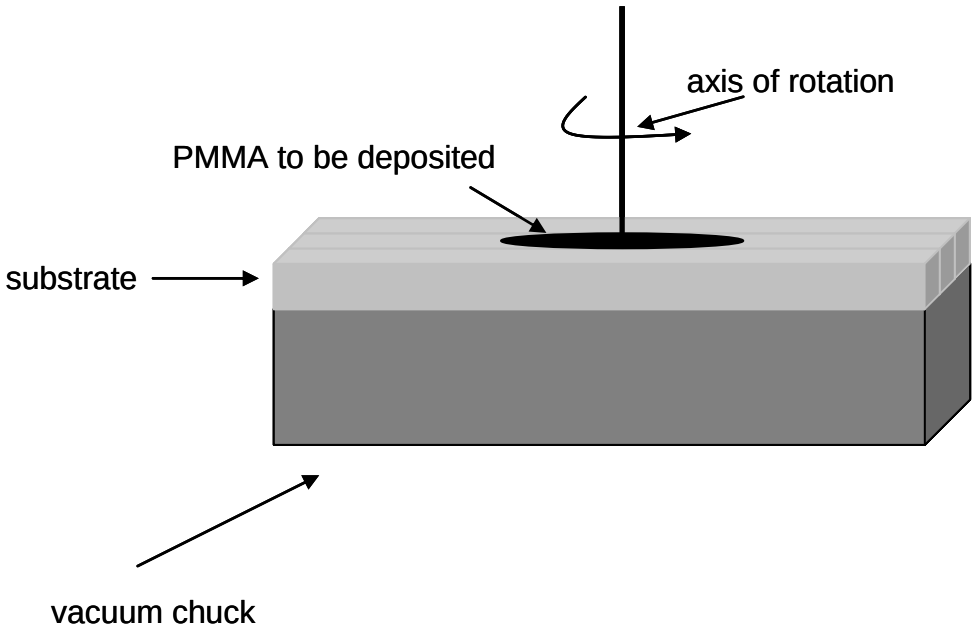


Figure 6.2 Schematic of the apparatus used in spin coating the photoresist onto the sample. Photoresist coats a sample as it spins between 2000-4000 rpm.

6.1.1.2 Lithography

The lithographic process exposes the coated wafer to energy from two sources: light in the UV or deep UV depending on the desired resolution, or electrons. Each source applies different optics since e-beam systems are direct write and UV systems are projection methods. Figure 6.3 shows a general photolithographic setup, the sample illuminated using a point source. Exposure time is dependent on resist type and thickness [69].

6.1.1.3 Photolithography

Photolithography [69] exposes the substrate with ultraviolet light, for a predetermined period of time through a mask re-imaged onto the substrate. The mask is the same size as the desired area and can be expensive depending on desired features and complexity. In some applications, many masks and exposures may be needed but for the research, since a binary DOE is fabricated, only one mask with one processing step is required, reducing the cost of manufacture and limiting errors associated with the use of multiple masks.

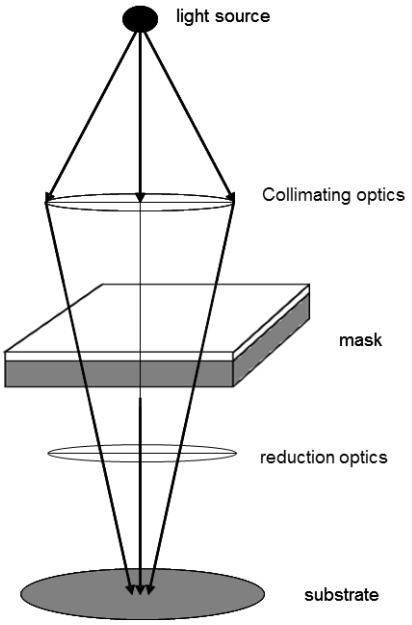


Figure 6.3 Schematic of a general lithographic method. Energy from a source illuminates a photo mask projected through optics onto a wafer coated with photoresist that is subsequently processed.

6.1.1.3.1 'Step and Repeat' Lithography

'Step and repeat' lithography is similar to photolithography except that the wafer is exposed across small areas and then stepped across to expose the entire wafer rather than all at once. These areas are referred to as dies and are tiled across the wafer (Figure 6.4). The process uses a mask or reticle stepped across the entire surface of the wafer, but only a single die is exposed at one time. Only a small pattern needs to be created instead of one that encompasses the entire surface area of the wafer.

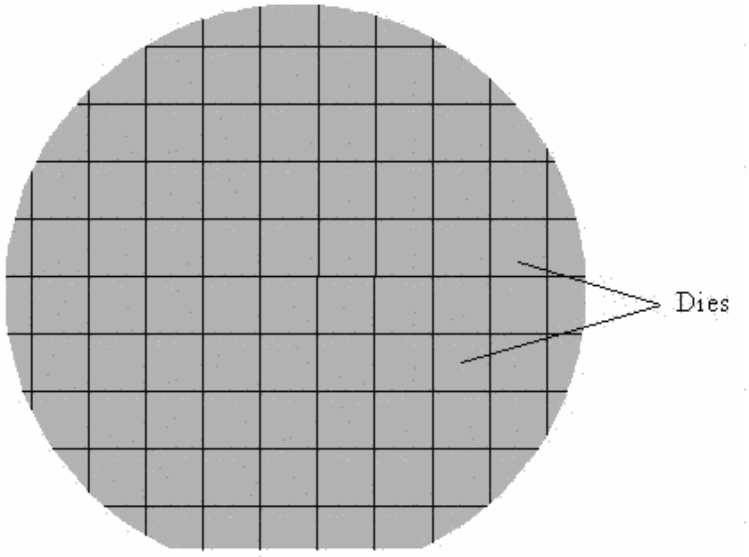


Figure 6.4 Die pattern of a wafer that is used in step and repeat lithography. The lines represent the boundaries of the "images".

To expose the wafer, light is projected through the reticle (Figure 6.5) onto the wafer at specific die locations. The most common type of stepper is a reduction stepper where images on the reticle are larger than the image to be written onto the wafer; projection optics reduce the size of the desired image before it reaches the wafer.

The advantage of this type of stepper is that the resolution for the lithography is not limited by the size of the image on the reticle. The most common types of reduction magnification for steppers are 1×, 4×, 5×, 10×; in this work, a 1× stepper is used.

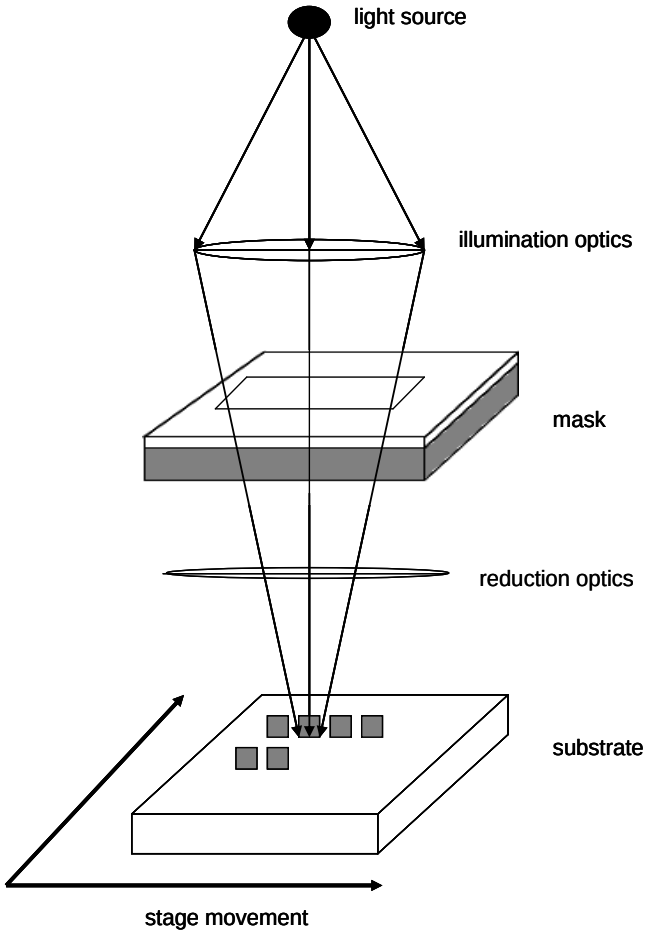


Figure 6.5 A schematic of the step and repeat lithography method. The method scans across a wafer to different die locations which are patterned lithographically. The technique allows for a smaller mask to be fabricated and is useful when a large number of features are repeated on the same wafer.

Another feature of the stepper is alignment markers, usually written on the outside of the image and core to aligning the reticle to the wafer. They are not transferred during the lithographic process. Usually two types of alignment markers are used:

one to align the reticle to the stepper column, the other to align the reticle to the wafer. Once the first "image" is transferred, the alignment process is simply a translation across the wafer. An example of these markers is shown in Figure 6.6.

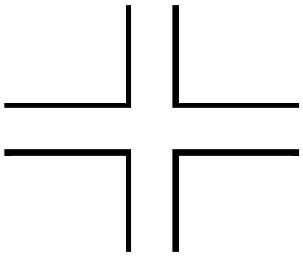


Figure 6.6 Example of an alignment marker used to aligning the mask to the wafer. These markers guarantee that the "images" are all aligned to each other and that they are properly spaced without overlap.

6.1.1.3.1.1 Electron Beam Lithography

Electron beam (e-beam) lithography is a maskless method that exposes a photoresist layer directly with an electron beam. The method scans a beam of electrons across the wafer, coated with photoresist sensitive to the energy of the electron beam e.g. PMMA, exposing it with a specific dose. The dose is the amount of electrons needed to penetrate the resist to the required depth.

One of the errors with e-beam approaches is the proximity effect which can lead to discrepancies in features. This is a result of the shape of the electron beam – a tear drop – since once it penetrates the sample, secondary electrons are generated (Figure 6.7). Irregular and small features require a different dose than larger features; if the dose is too large some features are lost. Another error is scattering from primary

electrons, which can be scattered multiple times in the photoresist layer over a large area, leading to exposure away from desired locations.

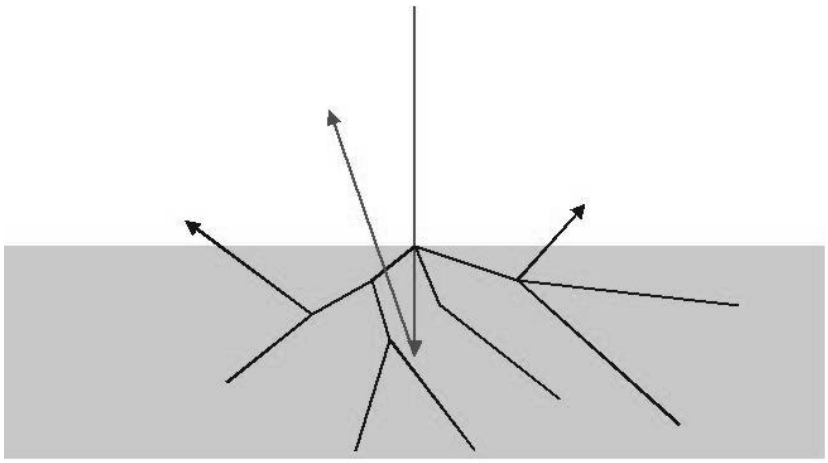


Figure 6.7: Profile of the secondary electrons as they penetrate and scatter into the wafer, highlighting why the dose of the electron beam is important and why proximity errors occur.

6.1.1.4 Etching methods

Two of the most commonly used etching methods are segmented as ‘wet’ and ‘dry’. Wet etching is most readily distinguished by the chemical used, producing either an isotropic profile that etches uniformly in all directions (but does not produce vertical sidewalls) or an anisotropic profile which etches along certain crystallographic planes (Figure 6.8).

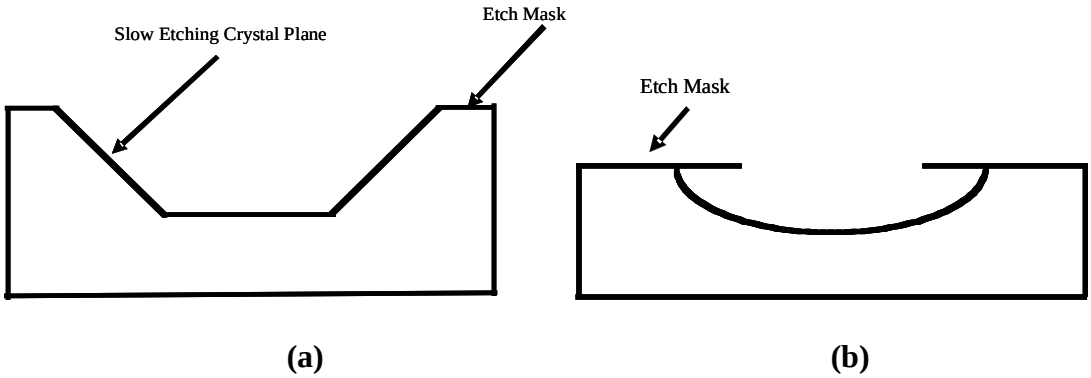


Figure 6.8 Difference between wet etch techniques and how the particular etching method progresses. (a) Anisotropic etch produces sidewalls that follow the crystallographic axis. (b) Isotropic etch, producing sidewalls rounded in the corners.

The research considered using an anisotropic etch - hydro fluoric (HF) acid - which etches at a rate that does not depend upon the exposed features, at least for the feature sizes utilised. For SiO_2, HF etching results in a sidewall angle of 53^o, which yields poor reconstructed images and was therefore not used.

Dry etching relies on a combination of chemical reaction and ion bombardment. The most widely used technique is reactive ion beam etching (RIE) which produces vertical sidewalls, but owing to surrounding features and the geometry of those features, variations in etch depth may occur (Figure 6.9).

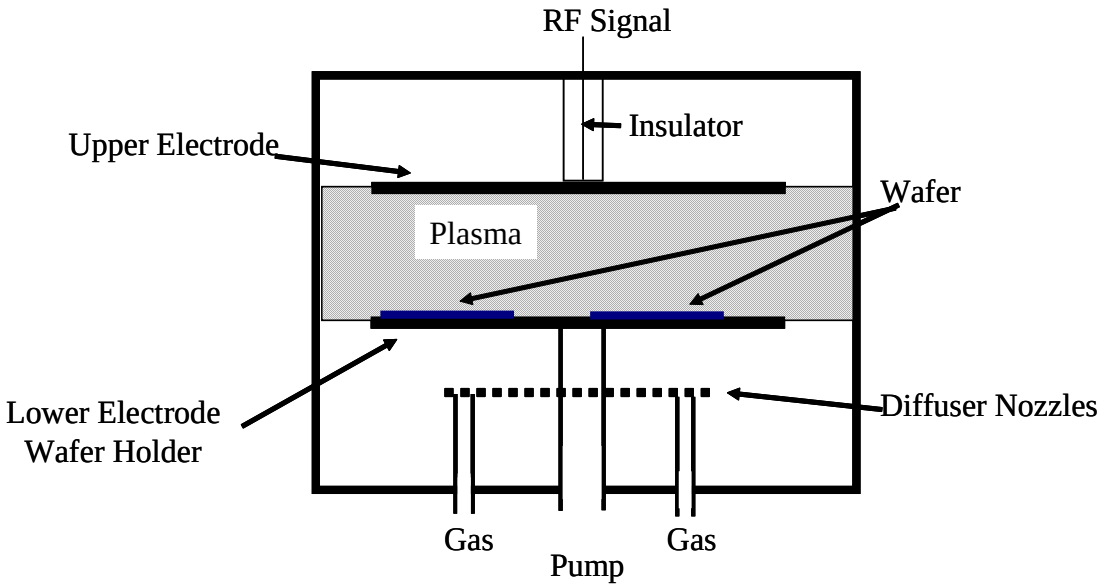


Figure 6.9 A reactive ion etch chamber.

6.1.1.5 Fabrication Process; Conclusions

Because photolithography is a replication technique, it is possible to produce a large number of elements with only one mask. In high volume production environments, Nano-imprint lithography is the best choice for fabricating diffraction gratings in an production environment however due to limited availability this method was not used in the fabrication of any DOE's for this thesis. When prototyping, electron beam lithography is a good choice because the fine resolution of the beam promotes accuracy and obviates the need to produce a quality photomask.

The required optical quality of the fabricated grating must be considered before selecting a particular etching method. If the image produced by the grating structure is viewed by eye only, a wet etch might be more preferable because it is of lower

quality and less expensive. If higher quality is required, a more expensive dry etch must be used.

6.2 Fabrication of Grating Structures

A number of samples were fabricated at the CPFC (Canadian Photonics Fabrication Center) using the step and repeat method and both wet and dry etching, to determine the quality of the grating structures. The structures are in the form of a 2×2, a 4×4, an 8×8 arrays together with a grid and circular patterns. The POGED method is used to generate the grating structures where all of the pixels with a value of one are extracted from a text file. Because the extracted files are in a text format, a conversion into a format that could be read by the Canadian Photonic Faction Center (CPFC) facility is required. The preferred software package is Design Workshop 2000 (DW2000) operating on a GDSII file formats. When importing files into DW2000, some information may need to be corrected, such as changing the layer information or the units to micrometres (Figure 6.10).

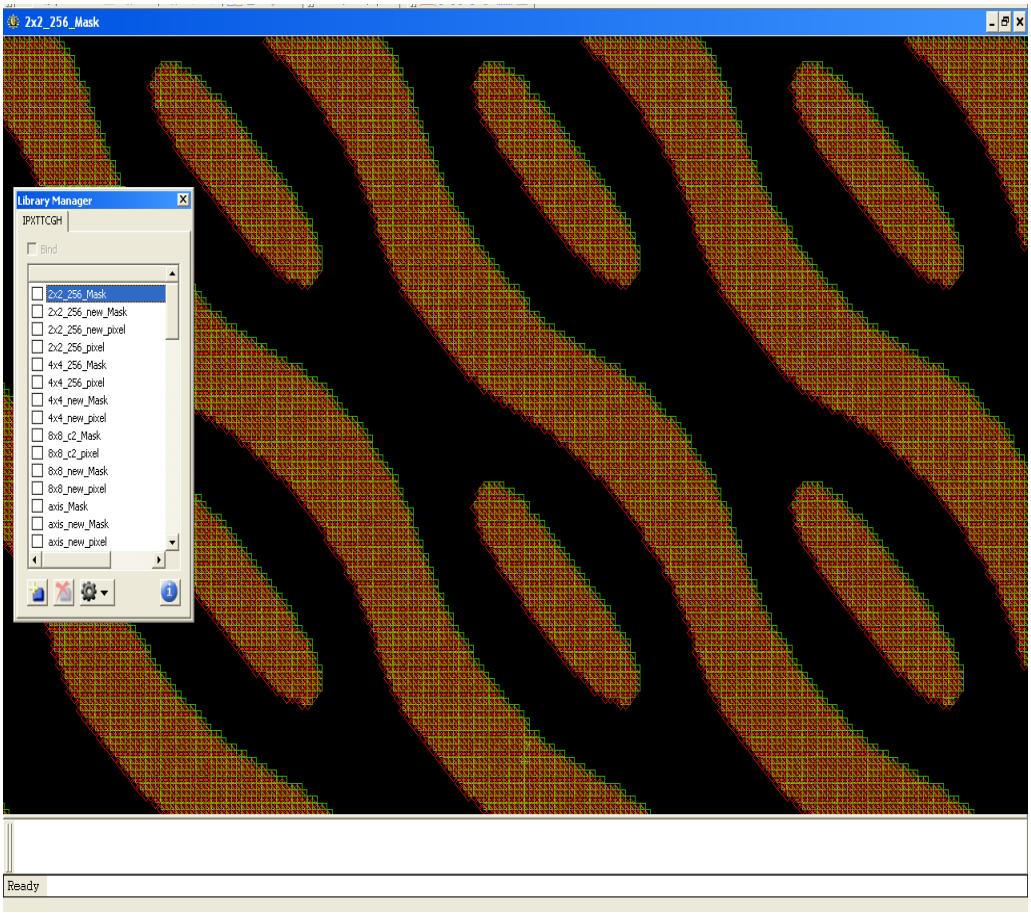


Figure 6.10 Outline of a grating structure that produces a 4×4 array generator. The image is a screen shot from the design package DW2000.

Once all of the separate GDSII files are on the same layer - layer 10 in this case - and the units are correct, the final file is assembled by merging the imported GDSII file into the template provided for the fabrication process.

The grating is then laid out. A square centred at 0,0 outlines the optimal area for the reticle, approximately 15 mm per side and comprising the six “images” placed in their required specific centre values (Table 6.1). The outline square and all images must be centred about 0,0 for the mask to be fabricated correctly.

Table 6.1 Location of each of the images as they are represented in DW2000. The correct locations of the images are crucial to mask fabrication.

Image Name	Image Coordinates				Image Center		Image Size	
	Top Left		Bottom Right					
1	-5850	6500	-1650	3500	-3750	5000	4200	3000
2	1650	6500	5850	3500	3750	5000	4200	3000
3	-5850	1500	-1650	-1500	-3750	0	4200	3000
4	1650	1500	5850	-1500	3750	0	4200	3000
5	-5850	-3500	-1650	-6500	-3750	-5000	4200	3000
6	1650	-3500	5850	-6500	3750	-5000	4200	3000

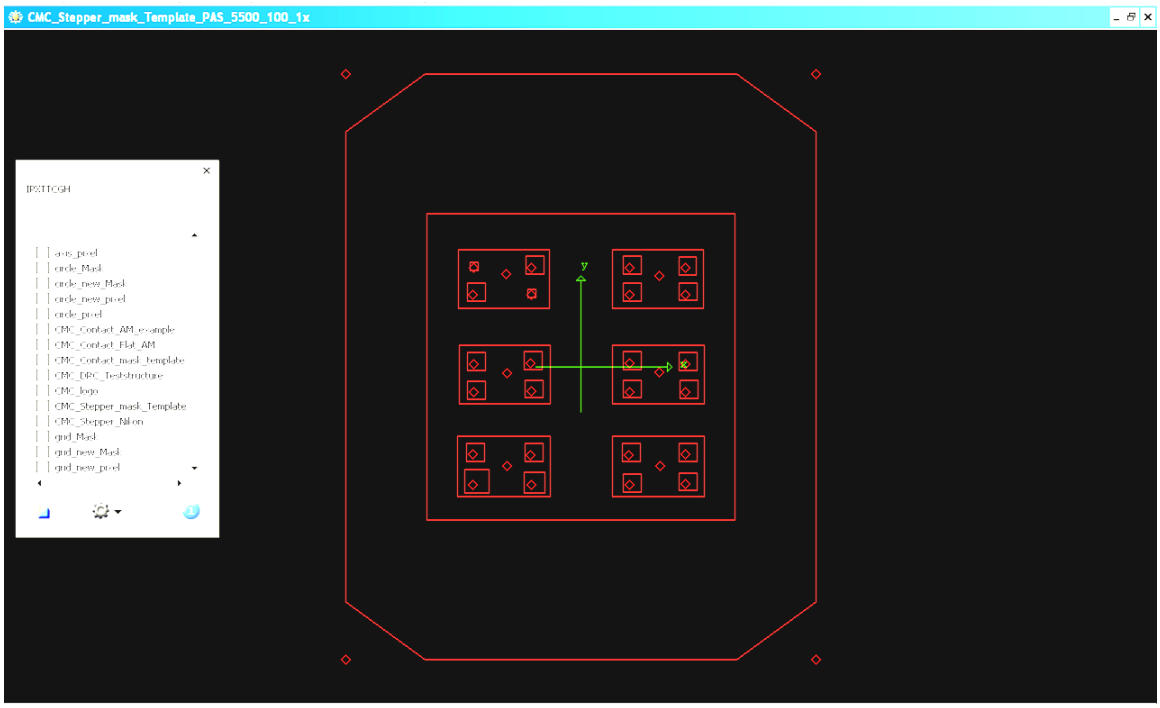


Figure 6.11 Template used to layout grating structures.

Once the entire layout process is complete, a decision on regions that are chrome and which are glass is reached. Owing to the chosen data extraction method, the chrome on the reticle is displayed in green (Figure 6.12) and individual squares represent individual pixels. These are the regions that remain after etching whereas the glass regions are the areas exposed and etched.

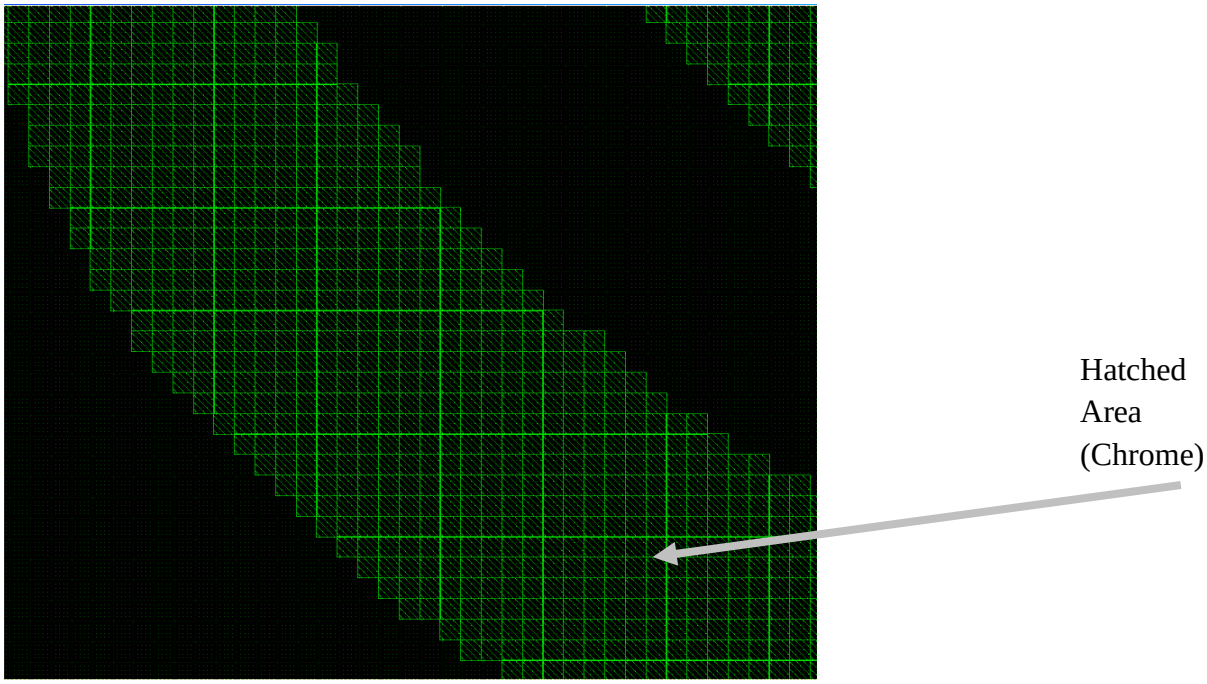


Figure 6.12 Example of what is written and what is not; the green part is chrome, the black part is glass.

6.2.1.1 Design Parameters

The GDSII files are all separated and named in a fashion that provides a description of the shape e.g. a 'twobytwob' is a file representing a 2×2 array generator. An example of file listings is given below, in addition designating the inner dimensions

of the grating as repeated. The grating comprises 9 separate structures laid out in a 3×3 pattern to help eliminate speckle. The total size of the grating is also listed.

'2x2'

A – 'twobytwob' is a GDSII file of a grating with inner section of 128×128 pixels; Size = 384 μm × 384 μm

B – 'twobytwo_new' is a GDSII file of a grating with inner section of 256×256 pixels and uses an updated version of how the pixels flips were performed; Size = 768 μm × 768 μm

C – 'twobytwo_256' is a GDSII file of a grating with inner section of 256×256 pixels uses the original method for flipping in the binary case; Size = 768 μm × 768 μm

The elements are separated by 9 mm in the vertical and 5 mm in the horizontal direction, also representative of the size of the die. Thus elements are separated by two die locations in either direction. The etch depth is determined by the material characterised by its index of refraction [36-37]. The etch depth is calculated as:

$$(n-1)d = \lambda/4 \quad \textbf{(6.1)}$$

for a reflection grating, where $\lambda = 632$ nm, $n = 1.457$ for SiO_2 and d = etch depth.

For reflection mode gratings – as opposed to transmission – a ¼ wavelength path difference is sought (instead of ½ wavelength); thus $d_{SiO2} = 345.68$nm for a reflection grating in SiO2 (Figure 6.13).

Figure 6.13 Example of fabricated binary grating structure. For a multiphase structure, the SiO2 follows a triangular staircase shape.

Figure 6.14 shows the image distributions on the reticle. The pattern follows 1, 2, 3… for the first level, repeated across the entire wafer. The next level begins where the previous level finishes e.g. if the previous level ends at 4, the next level begins at 5. A letter indicates the image placement on the reticle. Once fabricated, the wafer is diced or cleaved into individual "images". At the centre of each image (the six separate squares in the centre of the larger one) is a label displaying a numeric value which distinguishes each piece on wafer cleaving.

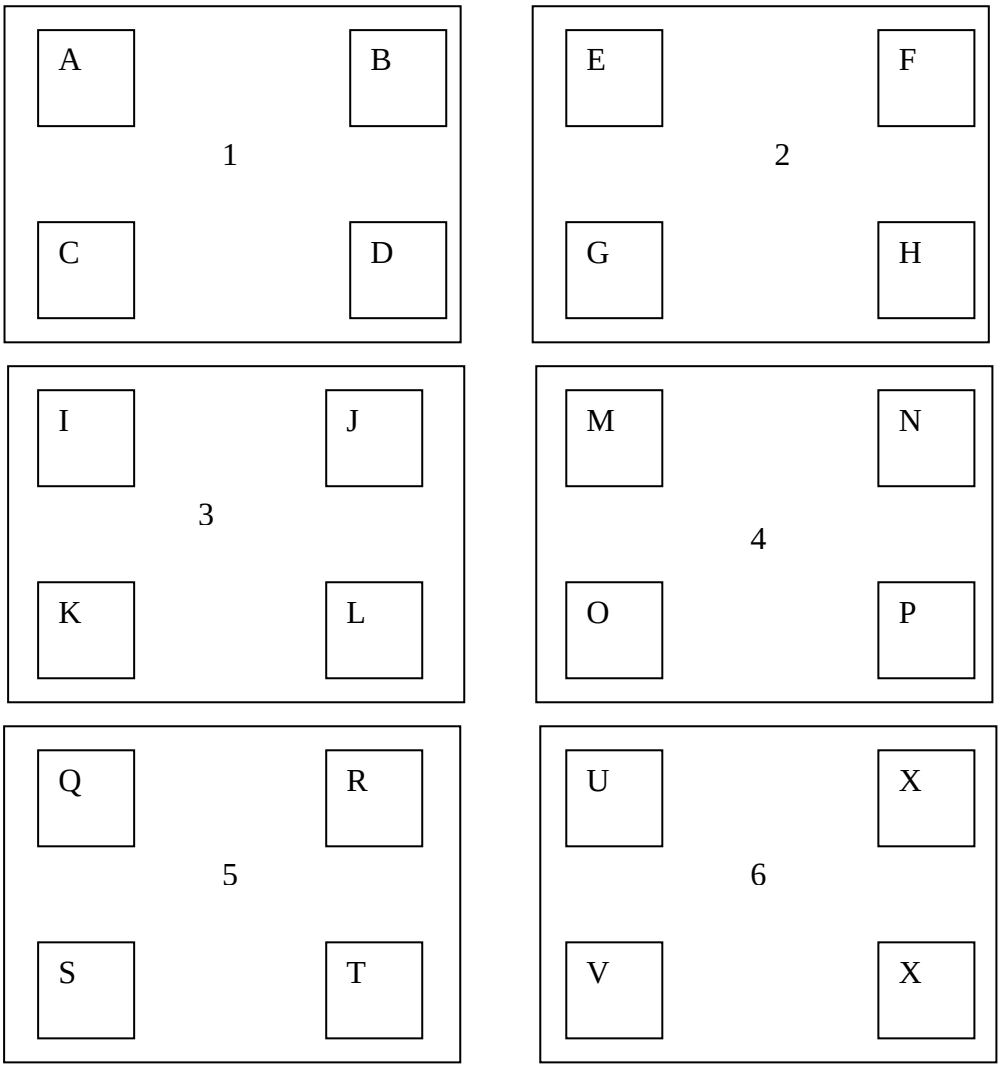


Figure 6.14 Placement of the grating structure as six “images”. Each grating structure is separated by 2 mm in the vertical direction and 1.5 mm in the horizontal. The images are separated by 1 mm so that there is no effect from secondary electron during the exposure process.

6.3 Conclusions

Due to the availability of equipment, electron beam lithography and step and repeat lithography are chosen for the manufacture of static DOEs. Both of these methods are capable of achieving the desired resolution of a 1 μm pixel size.

Electron beam lithography is a direct write technique that does not require the use of a mask; however there are a number of errors associated with it, most notably proximity effects and secondary electron exposure. The technique also requires that samples be charged and any features made on a glass substrate also incur an additional metal coating step.

Step and repeat lithography follows the same initial processing steps as in electron beam lithography, but requires a mask followed by exposure to UV light and is not subject to the same errors. Masks a can be expensive depending on the resolution of the features sizes.

Chapter 7

Manufactured Grating Characterization Methods and Results

7.1 Introduction

The Chapter focuses on a series of experiments that evaluate the structural and optical quality of the manufactured gratings. The techniques used to determine structural quality of the gratings are: white light interferometry [77], atomic force microscopy (AFM) [78], and scanning electron microscopy (SEM) [79]. A beam profiler [80] is used to further characterise the optical quality of the output beam produced by the gratings. CCD images are also used to view the reconstructed images for qualitative purposes only. Sidewall angle tolerances achieved are within $\pm 3^{o}$ from 90^{o} with a sidewall surface roughness of 1 pm. The tolerance achieved on the etch depth was ±5nm about the nominal value of 346nm.

7.2 Structural Characterisation

7.2.1 High Resolution Optical Microscopy

High resolution microscopy allows the fully fabricated grating structure to be viewed and compared to the original design in order to evaluate discrepancies between the two e.g. a region where there should be a void but remains filled in during the

lithographic process (referred to as the proximity effect when electron beam lithography is used).

To acquire a high resolution microscope image requires the sample to be fixed on a stage, with the camera and light source mounted above. The camera utilised is an Olympus with a broadband light source [81] providing a 50x magnification, able to capture most of the grating structure; the amount of structure recorded depends on the desired magnification and the overall size of the grating. If the full grating structure is too large, only a corner needs to be examined as it is replicated in the fabrication process (to minimise speckle).

7.2.2 Atomic Force Microscopy

Atomic Force Microscopy (AFM) assists in determining the structural quality of the fabricated grating, key features being sidewall verticality, surface roughness and etch depth. It is also possible to determine any cracks in the structure, how deep they might be as well as any residual resist. Etch depth and sidewall verticality are important since errors in these parameters contribute to a high DC level and a loss in diffraction efficiency [39-40].

AFM scans a sample using a very fine probe, in the order of 10-20nm in diameter and a length of 30 μm to 500 μm depending on the type of element being examined. The probe is mounted on a tuning fork/cantilever at a height between 50 μm to 500 μm and length of 300 μm to 1000 μm. The force exerted by the probe onto the

sample is in the range of 1-20 N/m, at a resonant frequency of between 20 kHz to 390 kHz. An example of a probe and cantilever is illustrated in Figure 7.1

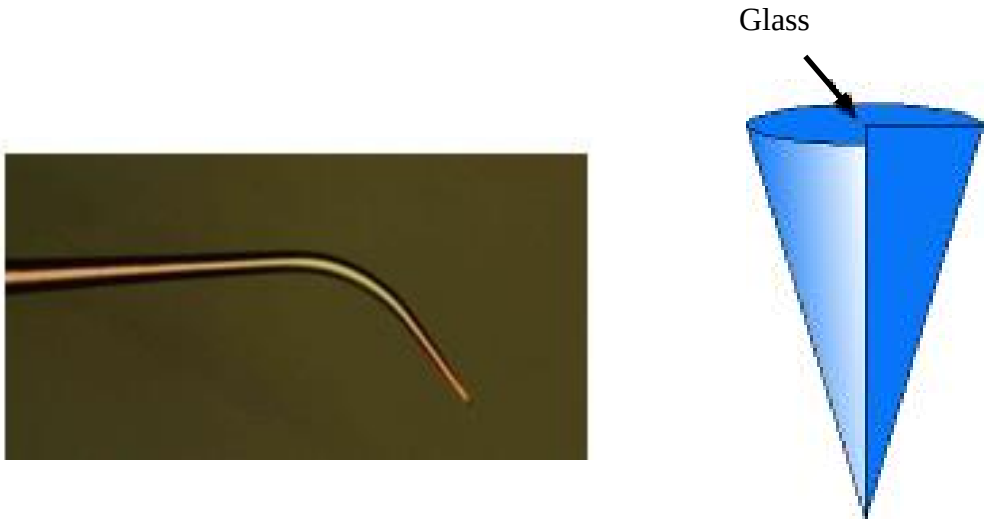


Figure 7.1 Views of the AFM probe made of optical fibre. The optical fibre is heated and manipulated to achieve a bend. The tip is formed by heating the fibre and carefully shaping into a small point on the order of nm [78].

To acquire an image, the probe scans over the surface recording the difference in probe force, which is then translated into a voltage that in turn represents a height. The speed at which the probe moves is important; if the probe moves too quickly the tip could break or the resolution of the measurement degrades. One other impact of probe speed is its effect on the measurement of sidewall verticality. When the probe scans across a sample, it "falls" off edges and "climbs" the other sidewall. If the probe moves too quickly, the edge appears 'rounded', resulting in an inaccurate measurement of the edge verticality.

7.2.3 White Light Interferometry

White light interferometry is a non-contact method used to measure surface roughness. The principle behind white light interferometry is similar to that of a classical interferometry [77], however a broadband source is used instead of a coherent laser source. A basic schematic of a white light interferometer is illustrated in Figure 7.2 . Three to five measurements of the intensity per fringe per pixel are taken so that the phase can be extracted; the phase is used to calculate the height/surface roughness.

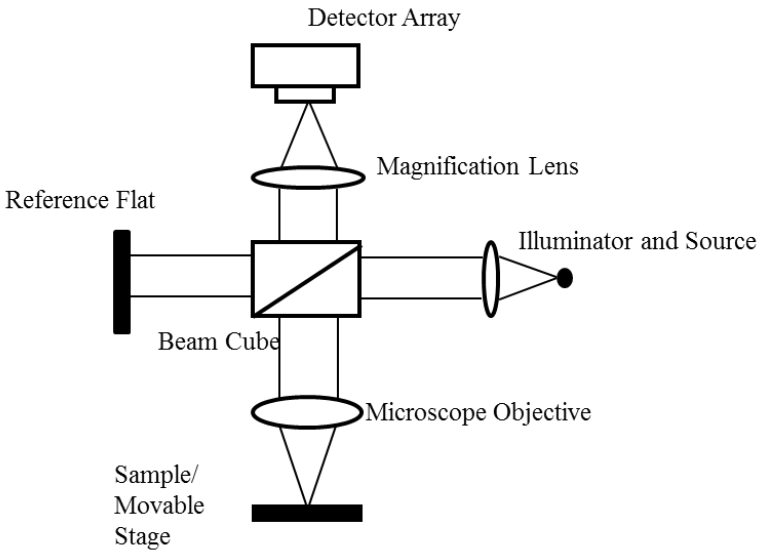


Figure 7.2 Schematic of a white light interferometer. A white light source propagates to a beam splitter where a reference beam splits off whilst the object beam propagates to the sample stage where it is reflected back and collected.

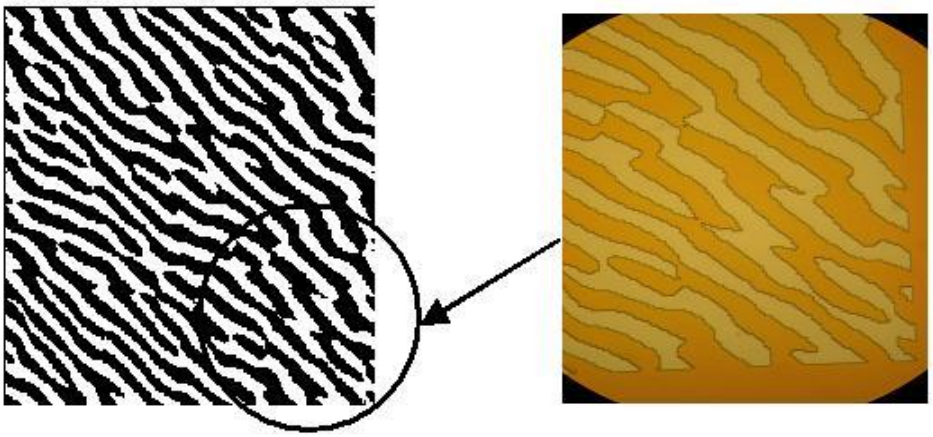
These types of interferometer have become very popular in profiling surface roughness in semiconductor materials [77], and is currently used routinely in many industries.

7.3 Structural Measurement Results

The Section reviews the structural quality of the designed gratings fabricated using electron beam and step and repeat lithography. Errors originating during the fabrication process are identified along with a full interpretation of the characterisation results. Results comprise Scanning Electron Microscopy (SEM) images, AFM images and white light interferometry, the latter used to validate the AFM images.

7.3.1 Electron Beam Lithography Fabrication Results

Figure 7.3 illustrates gratings fabricated using electron beam lithography. The circle in the bottom right hand corner of the simulated grating structure represents the optical image of the corresponding fabricated structures. The dark yellow lines shown in the optical image correspond to the black lines in the simulated structure and the light yellow lines correspond to the white lines. The circled area captures a comparison between the simulated and fabricated grating. Some differences are apparent in this region, directly attributable to fabrication errors.

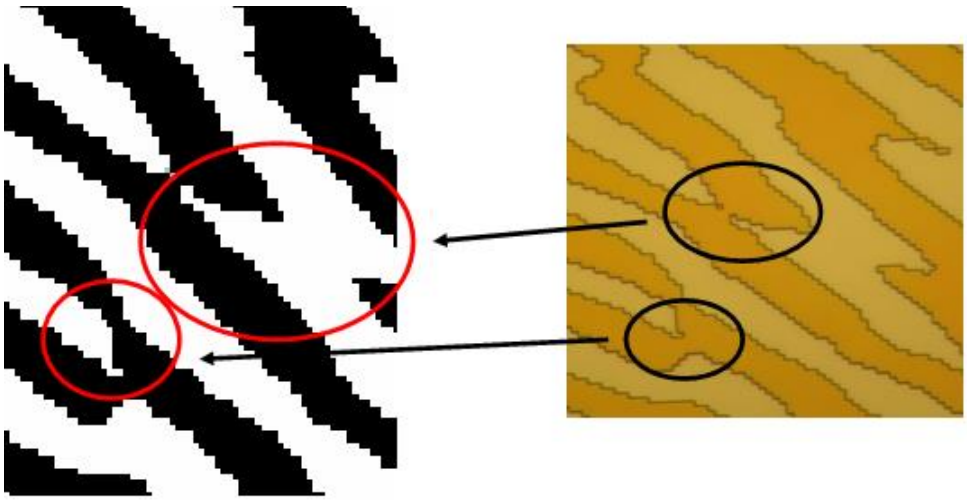


(a) Simulated grating structure.

(b) Optical image of grating structure fabricated by electron beam lithography.

Figure 7.3 Comparison image between the simulated and fabricated grating structure using electron beam lithography where 1 pixel has dimensions of 1μm x1μm.

The differences highlighted by circles in Figure 7.34 clearly accentuate the errors that occurred during the fabrication process. Because these grating structures are fabricated by electron beam lithography, the errors are attributed to the proximity effect, resulting in features that are either too large or too small compared to the design. These types of errors are not uncommon and need to be evaluated on an individual basis as they reduce the optical quality of the grating structure.



(a) Small section of simulated grating fabricated

(b) Section of grating structure by electron beam lithography.

Figure 7.4 Areas with errors. (a) illustrates a region of the grating structure defined by the design tools. (b) Illustrates the same region of the fabricated grating structure. The areas enclosed by circles identify fabrication errors where 1 pixel has dimensions of 1μm x1μm.

AFM images of the same grating (Figure 7.5) demonstrate that both the vertical sidewalls and surface roughness are within tolerance. In these particular images, the pixellation of the grating is clearly visible.

Since the grating is fabricated in just resist with no etching into the Si layer, cracks appear in the bottom left hand corner, forming an artificial feature in the structure. The origins of these cracks have not been traced but most likely appear when a sample is developed. In addition, scratching of the soft resist surface is possible e.g. from handling by tweezers used to mount the sample or from the AFM probe if the latter comes into contact with the sample.

Incorporation of the fabrication errors in simulations is a powerful approach to verifying their impact on performance; however this is highly challenging because relating the exact fabrication error to a specific impairment is very difficult [39-40, 72]. Etch depth errors result in a higher DC content and lower diffraction efficiency; departures from a 90^{o} sidewall contribute to stray light exiting the system; surface roughness results in scattered light; and stitching errors result in a cross patterning in the reconstructed image.

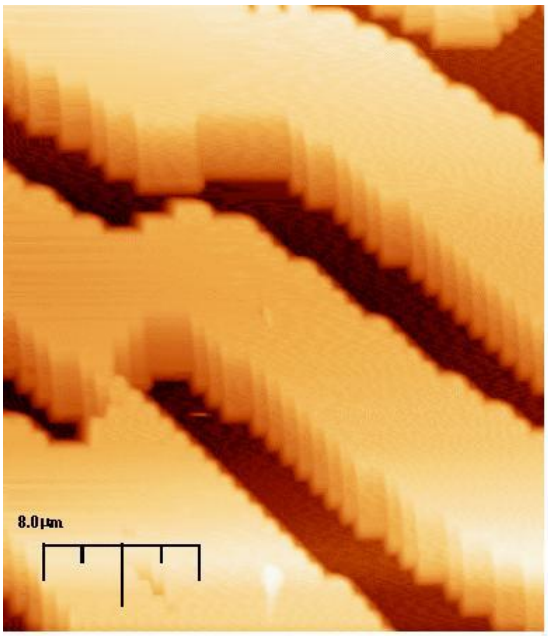


Figure 7.5 AFM image of the grating structure fabricated by electron beam lithography.

Cracks, scratches, proximity errors and errors in the etch depth all contribute to loss of signal quality resulting in a high DC content and scatter.

7.3.2 Step and Repeat Lithography Fabrication Results

Step and Repeat lithography results presented and analysed in this Section characterise the structural quality of the fabricated samples derived from a similar series of measurements viz. SEM, AFM and white light interferometry. SEM measurements determine the etch depth of the fabricated gratings, the uniformity of the etch depth, and permit an evaluation of sidewall verticality. AFM and white light interferometer measurements verify the SEM data.

SEM images show that the gratings are pixelated and uniform in height and structure (Figure 7.6). The etch depth is measured at 345 nm into the SiO2 layer, 1nm different from the designed height indicating an optimum readout wavelength of 635nm, marginally longer than the design wavelength of 632nm (corresponding to an etch depth of 346nm). The sidewall verticality of the gratings is within specified tolerance, independently verified using AFM measurements and white light interferometry. Surface roughness and etch depth uniformity are also determined using the same techniques. Visually, the gratings are well defined and small features are evident, not the case with the gratings produced by electron beam lithography which suffered from proximity problems.

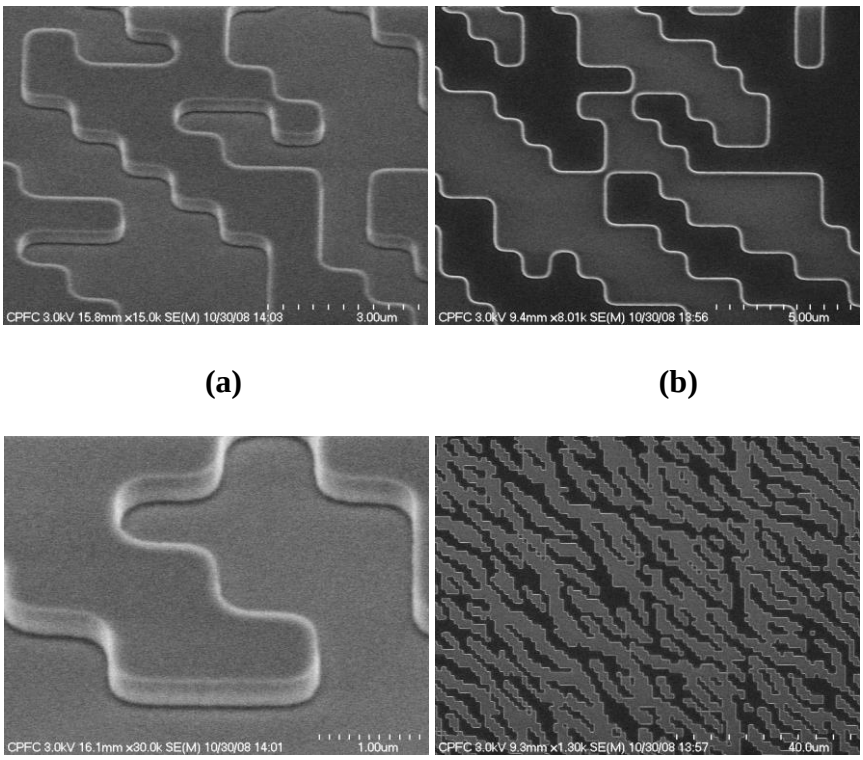


Figure 7.6 A number of SEM images are illustrated. These images demonstrate the design from different angles and at different zoom position (a) and (c) are at different angles illustrating the sidewall verticality and (b) and (d) show the design as it would be viewed at incidence.

7.3.2.1 Atomic Force Microscopy Measurements

AFM measurements are used to verify etch depth and sidewall vertically and yield information on scratches, cracks and leftover resist.

AFM images (Figure 7.7) show regular shapes with some errors owing to fabrication. In the bottom centre of Figure 7.7 (a), some resist remains on the structure, corroborated in the 3D version of the AFM image (Figure 7.7 (b)). The straight lines

running across the AFM images are an artefact of the probe moving across the sample, verified by rotating the sample by 90^{o} when these lines change direction (horizontal to vertical).

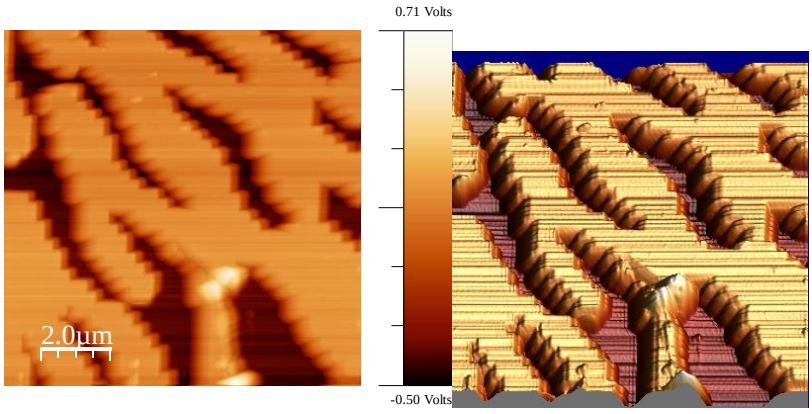


Figure 7.7 AFM images of a grating fabricated by step and repeat lithography. Two views are depicted; a top view (a) where some errors are noted in the centre at the bottom. (b) a 3D view showing the errors more clearly.

Profile measurement (Figure 7.8) records the presence of the raised portions of the grating and further confirmation of leftover resist from the fabrication process. Profile measurements also confirm that sidewalls are vertical, evident on the left hand side of the profile image.

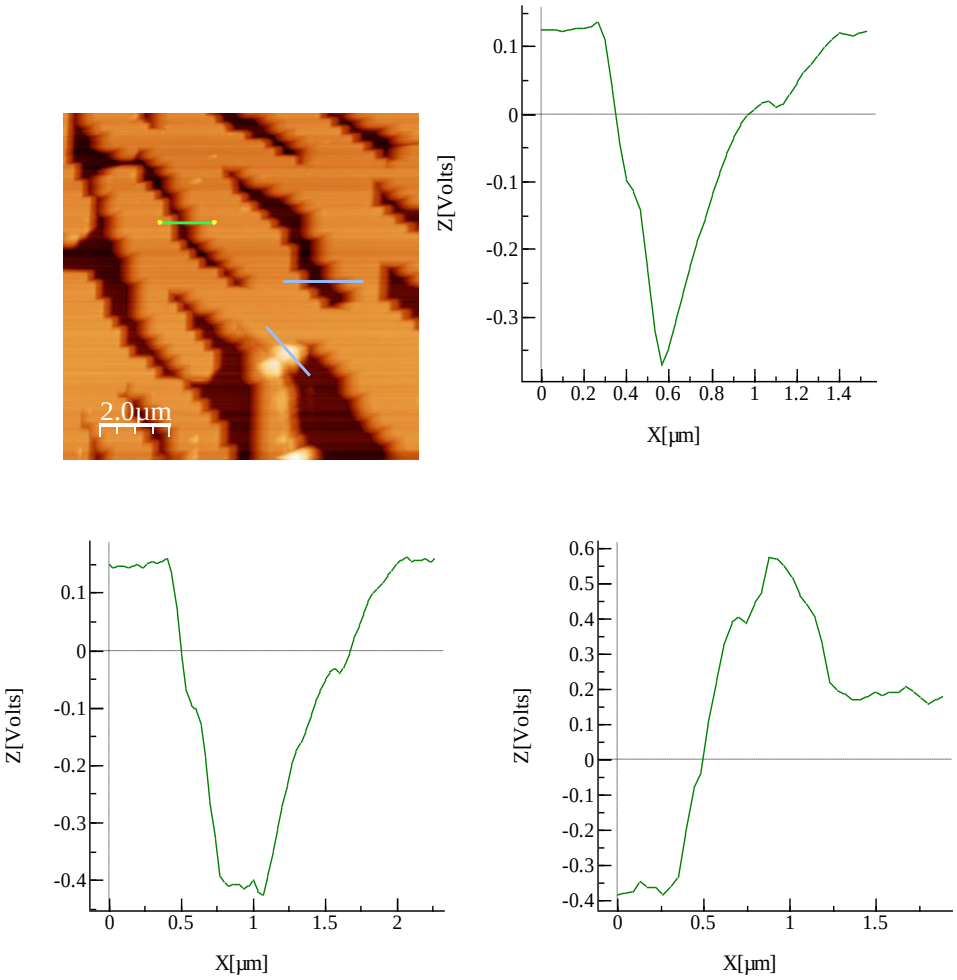


Figure 7.8 Profile measurements from AFM demonstrating both sidewall verticality and the error from the raised portion.

Figure 7.9 highlights two close raised portions owing to incomplete etched areas due to an exposure error or the geometry and separation of the feature.

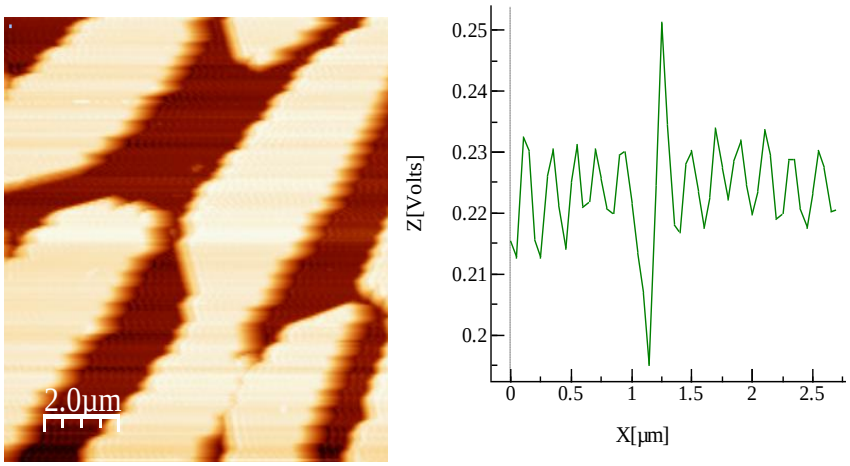


Figure 7.9 AFM measurement profile illustrating an incorrect etch depth..

AFM measurements confirm that both sidewall verticality and surface roughness are within tolerance; however some errors originate from the fabrication. These are common to the fabrication process and are difficult to quantify their effect on the optical quality of the gratings. Most of these errors however result in a higher DC content, un-diffracted light, and scattering of the signal. How much these errors contribute to a loss in signal requires a more in-depth analysis; distinguishing what error contributes to what loss and how by much remains an on-going challenge [39-40, 72]. Some other errors e.g sidewall verticality on one side, lines across the sample are the result of the measurement process and are easily explained and confirmed.

7.3.2.2 White Light interferometry

White light interferometry measurements were performed using a Zygo NewView 5000 [83] in tandem with the analysis software MetroPro [83]. Samples measured

using the white light interferometer are coated with a gold layer of thickness 35nm, necessary since SiO_2 is transparent and since both the SiO2 and the Si layer are being viewed simultaneously. Figure 7.10 depict high resolution optical images of two of the grating structures - circle and a "T"- showing distinct high and low regions (blue: low area; red: high area).

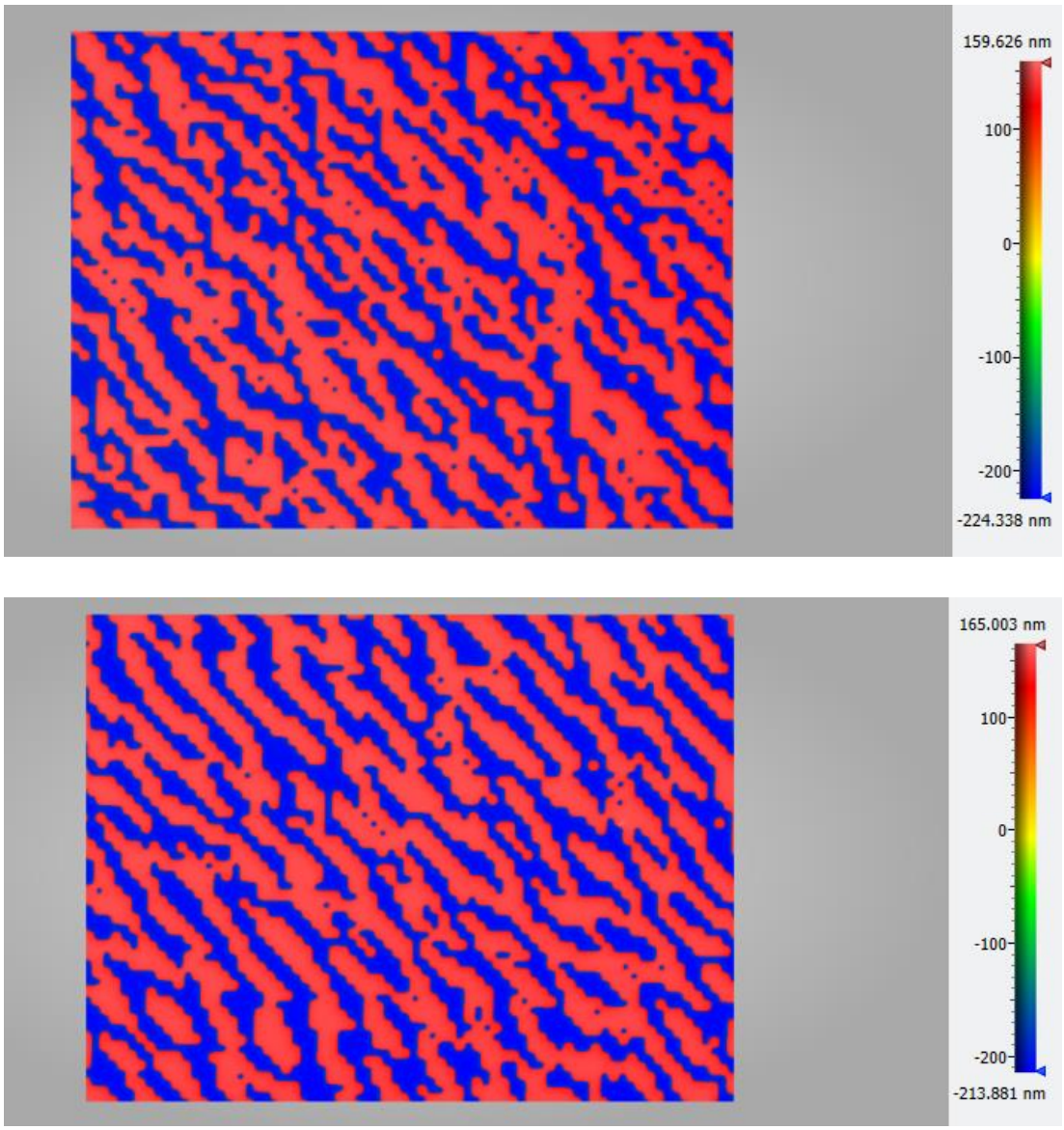


Figure 7.10 White light interferometry images of a circle grating and a "T".

A slice through the relief maps of the two grating shows that the structures have vertical sidewalls of height 351nm (Figure 7.11(a)) and 357.5 nm (Figure 7.11(b)). A 35 nm gold coating was added to minimise any stray reflection occurring between the SiO2 layer and the Si layer. This layer may have resulted in small errors into the reported etch depth as it varies slightly in the trenches.

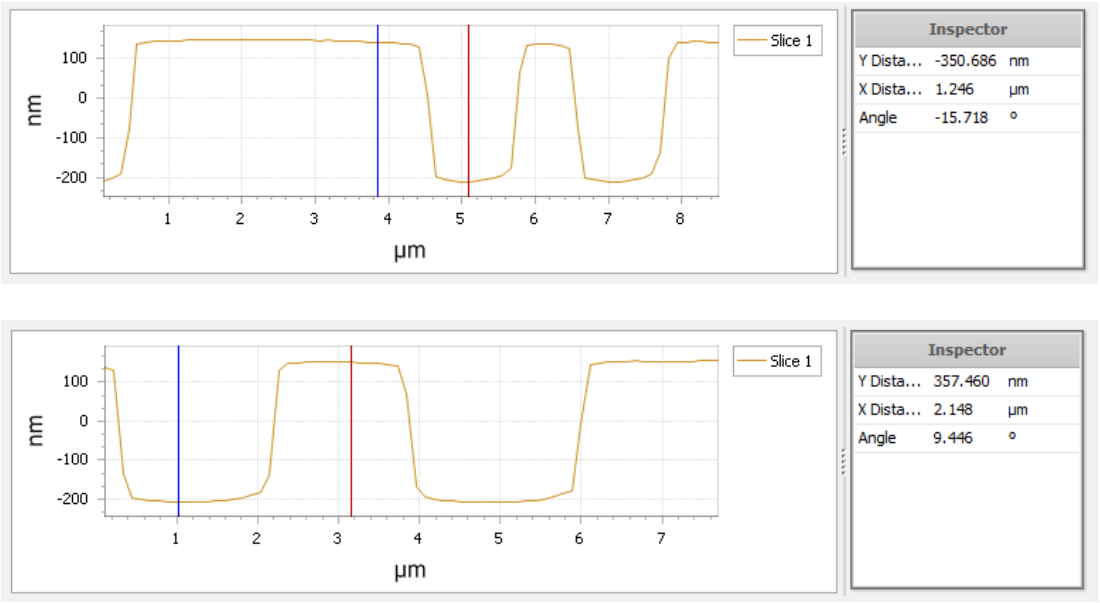


Figure 7.11 Profile measurement for two grating structure measured by white light interferometry.

Figure 7.11 is a line profile of the fabricated gratings illustrated in Figure 7.10. The two profile sidewalls and surface roughness are within tolerances, as specified in Section 7.1. These measurements correlate well to measurements recorded by AFM and SEM. The single spot indicates that the sample has not etched to the correct depth because of the geometry surrounding the spots.

7.4 Optical Measurement

The Section details the characterisation framework to measure the optical performance of the grating structures and reviews results obtained. Optical measurements are targeted towards determining diffraction efficiency, the quality of the output beam and a qualitative measurement of the reconstructed image.

7.4.1 Beam Profiling

A laser beam profiler (Newport LBP1) [80] measures the power of the output beam directly in real time and as it offers a large active sensor area, a large portion of image can be characterised. The active sensor area of the profiler is 6.47mm wide by 4.83 mm high, at a sensitivity of 5nW/cm^2, and is thus capable of viewing very low power output beams, core to the characterisation of errors surrounding the reconstructed image.

The profiler produces 2D and 3D maps of output images. The cross section of the output beam profile is a least-square fit of the Gaussian function; the following is used to calculate the Gaussian fit:

$$I = Ve^{\left(-\frac{(x-c)}{\sigma}\right)^2} \quad \textbf{(7.1)}$$

where I is the intensity of a pixel at location x, V is the maximum intensity of the fitted Gaussian curve (peak intensity), c is the centre of the Gaussian fit peak (centroid), σ is the radius of the Gaussian fit curve at the 1/e² intensity level (diameter).

Figure 7.12 is schematic of the experimental setup for beam profiler measurements, illuminated by a HeNe laser emitting 5mW of power. The laser beam impinges on a beam splitter directing the beam to the sample, the reflection redirected at a 90˚ angle since operation is in reflection mode. An iris is used for alignment purposes to guarantee that the beam enters and returns at the same angle. The beam profiler is mounted on a translation stage to view different section of the reconstructed image.

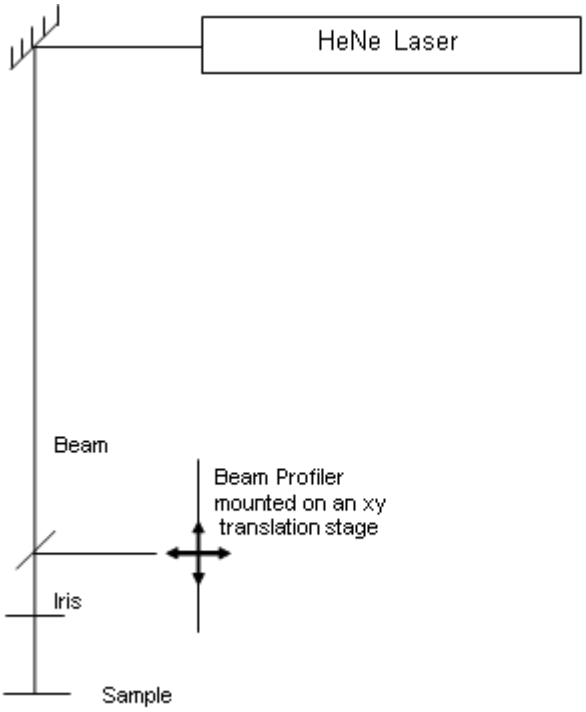


Figure 7.12 Laser beam profiler measurements schematic.

7.5 Optical Measurement Results

7.5.1 Fabricated Grating by Electron Beam Lithography

The optical results for the grating fabricated by electron beam lithography have an etch depth of 390nm. The original image encoded in the grating structure is a “T” to

demonstrate the inversion symmetry of the reconstructed image from a binary grating.

Inspection of the diffracted field reveals a spatially discrete - not continuous -image, a consequence of the 3x3 replication of one unit cell (with pixel number equal to the FFT length used) of the grating structure (Figure 7.13).

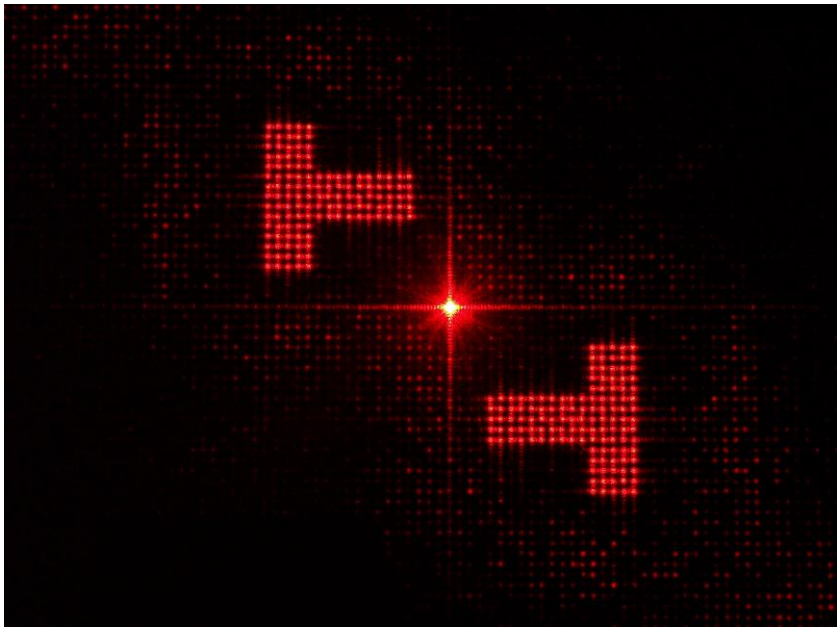


Figure 7.13 Reconstructed “T” image taken using a 6Mpixel camera on a black screen.

The reconstructed image shows a DC component at its centre, between the two “T” structures. The bright spot represents the DC level as a result of Fresnel reflection losses occurring because a metal layer was unintentionally omitted during the fabrication process. These losses occur because as light propagates between two

different media with different indices of refraction, both refraction and reflection can occur, calculated as $R = (n' - n)^2/(n' + n)^2$ [94]. These Fresnel reflections will occur at the interface of the SiO2 layer and the Si bulk layer and is on the order of 16%. Another contributor to the high DC content results from the fact that the illumination beam may have been circumscribed instead of inscribed. The horizontal and vertical lines directly in the center of the image are the result of stitching errors attributed to the electron beam lithography process. Any hysteresis in the scan results in the misalignment of pixels. Some of the errors that are visible in individual beamlets are the result of poor quality optics in the camera used to capture the image (6 Mpixel by Cannon) [80].

Figure 7.154 is an image for a "T" grating structure taken at a distance of 25.3 cm from the sample using a beam profiler, conducted to determine the shape of the output beam and record errors surrounding the beamlets.

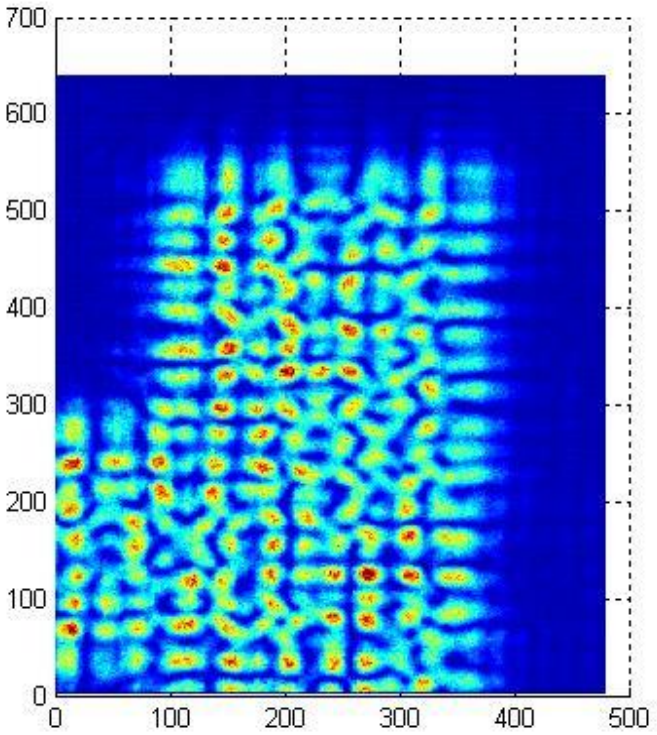


Figure 7.14 Measurement from a diffractive optical element using a beam profiler at a distance of 25.3 cm away.

When a smaller section of the reconstructed image is viewed, the individual beamlets become more visible; measurement of the grating structure at a distance of 18.5 cm from the detector highlights the errors surrounding the beamlets (Figure 7.15). The errors present in the system appear to be defocus errors caused by the distance in which the samples were measured. These samples were measured at a short distance to accommodate the size of the detector. To improve upon these measurements a lens could be used to reduce the focal length of the DOE.

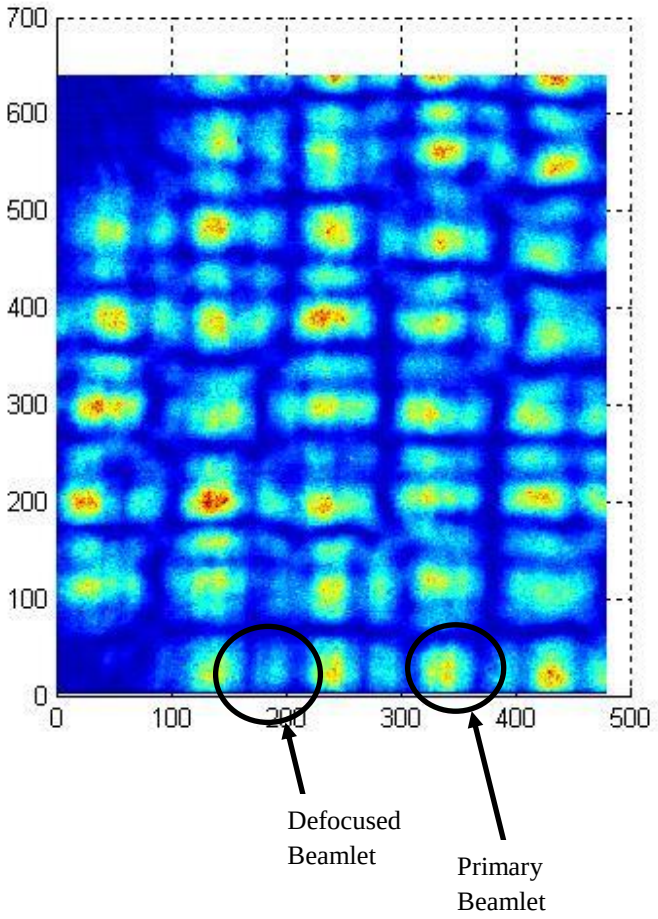


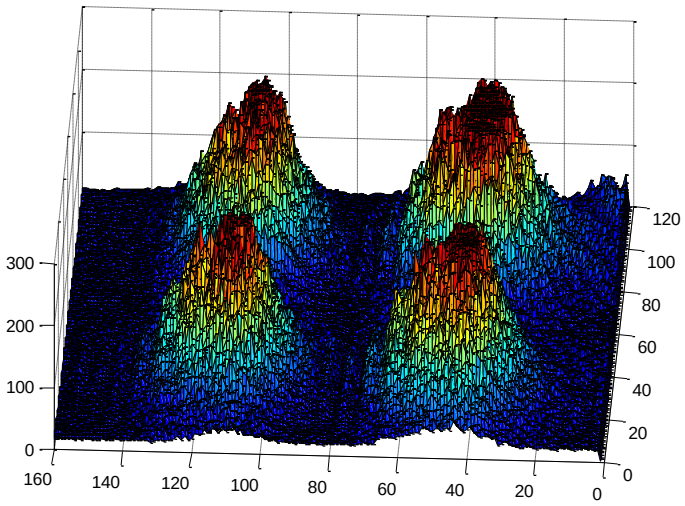
Figure 7.15 Measurement of the grating structure at a distance of 18.5 cm from the detector. The primary beam and a defocused beam are highlighted.

7.5.2 Step and Repeat Optical Results

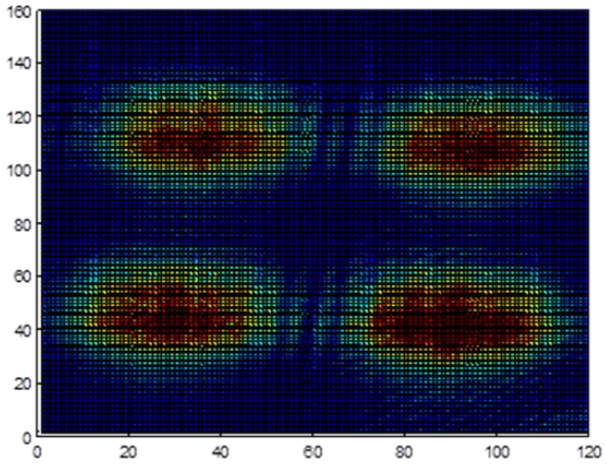
7.5.2.1 Beam Profiler Measurements

Beam profiler results for gratings fabricated using step and repeat lithography are recorded in this Section. The beam profiler facilitates the evaluation of the shape and intensity of output beamlets; owing to the mismatch between the size of the image and that of the detector, only sections of the reconstructed images are evaluated.

Results for a 2×2 binary array generator are recorded at a distance of 18.5 cm to the detector, illuminated using the arrangement described in Section 7.4. Beamlets emitted by the grating are circular in shape, of intensity following a Gaussian radial profile. There are negligible errors in the shape but because of the fabrication constraints on the gratings, small errors such as these are expected. Only half of the reconstructed image is evaluated in order to achieve a more granular view of the individual beamlets, given the size of the output image. The other portion of the reconstructed image produced by a binary grating is identical in shape and output. A DC component, common to most grating structures, is also visible in the overall output and is of relatively high intensity; this high DC level results from Fresnel reflection losses occurring at the interface of the SiO2 to the Si bulk layer and because the illumination beam was circumscribed instead of inscirbed (Figure 7.16).



(a)

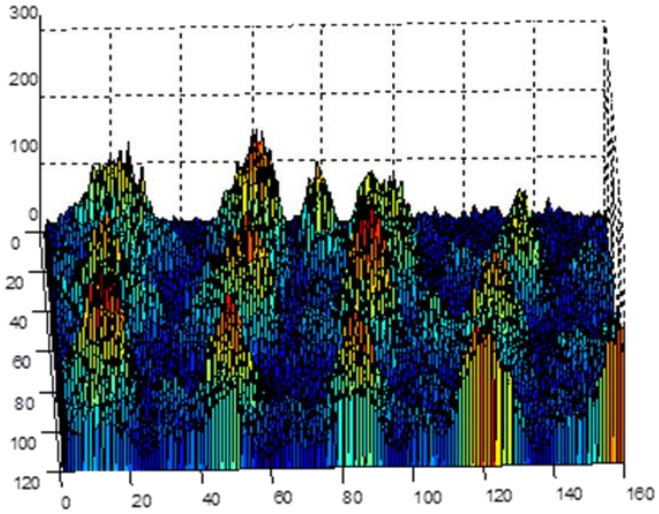


(b)

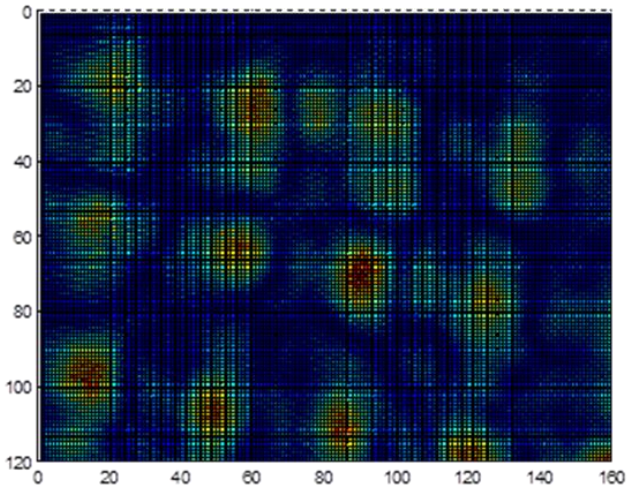
Figure 7.16 Views of the reconstructed image from a 2x2 array generator at a distance of 18.5 cm to the detector. The beamlets that make up the 2x2 array generator can are Gaussian in form. (a) 3D view. (b) 2D view.

A second set of measurements were carried out on a reconstructed image of a section of a circle with multiple circles within it, evaluated because of its size and complexity. The 3D map (Figure 7.17) of the output beam shows many small

beamlets constituting the desired output shape. These beamlets however vary in height and defocused beamleats are evident between the peaks attributed to a defocus error present in the test setup [39-40].



(a)



(b)

Figure 7.17 (a) A 3D output intensity profile of a portion of a reconstructed circle.

(b) A 2D view of a portion of the reconstructed circle.

Results indicate that the 2x2 array generator produces the best quality with uniform images and few errors. The grating structure for the 2x2 array generator is the easiest to produce and the most regular in shape, bringing benefits owing to an easier fabrication process when features are regular and the spacing between the raised and the lower portions is large. These benefits translate into sidewalls that are vertical, an etch depth that is uniform and no errors in the mask fabrication process. All errors result in discrepancies in the reconstructed image and are harder to control when grating structures are fine and the shape of the structure is not as regular as that of the 2x2 array generator.

7.6 Conclusions

It is clear from measurements that the fabricated DOEs produce the designed shapes. Structurally, samples are within the specified tolerance demonstrating good side wall verticality and depth tolerances within the specified range. There is however errors present such as cracks, resist not fully removed, unintentionally omission of the metallisation layer that have resulted in Fresnel losses at the layer transition and stitching errors originating from the raster scanning process of the e-beam lithography and mask fabrication. As the state of the art in fabrication evolves, these errors can be controlled but to fully understand their full effect on image quality requires a full rigorous characterisation of the fabrication process which falls out with the scope of the Thesis.

Chapter 8

Free Space Optical Interconnects (FSOIs)

8.1 Introduction

An application of POGED is in optical beam steering for Free Space Optical Interconnects (FSOIs) [5-11] as it produces high quality reconstructed images with high SNR and diffraction efficiency. The Chapter concerns the demonstration of the application of POGED in a noteworthy application rather than a detailed FSOI performance characterisation.

Although it is acknowledged that a number of FSOI architectures have been reported and indeed deployed [3-12, 85-86], the particular architecture that forms the focus of the feasibility analysis of the use of POGED was reported first by Marsden et al. [85]. The architecture relies on the transposition of inputs to outputs through free space optics, developed further by Milojkovic et al [86] (Figure 8.1). Both architectures employ two sets of meso-lenses in an imaging configuration separated by the distance *2f+2F*, where *f* is the focal length of the meso-lenses and *F* is a distance determined by the number and spacing of beamlets composing the transpose. The input meso-lenses are arranged in the form of a two-dimensional array that steers optical beams to a corresponding output set of meso-lenses which direct the beam to the appropriate output port location. However due to the extreme off-axis geometry of the system, a number of aberrations are introduced that result in increased insertion loss in systems terminated by single mode optical waveguides.

Cross-talk resulting from beam divergence and clipping by the second set of meso-lenses must also be monitored to minimise noise.

8.2 Central Fourier Transform Lens Architecture

The proposed architecture utilises a Fourier Transform lens at its center and two sets of meso-lenses [85-86], one set at the source, the other at the detector/optical fibre (Figure 8.2). A Spatial Light Modulator (SLM) can be deployed at the focal spots of the meso-lenses to perform active beam steering.

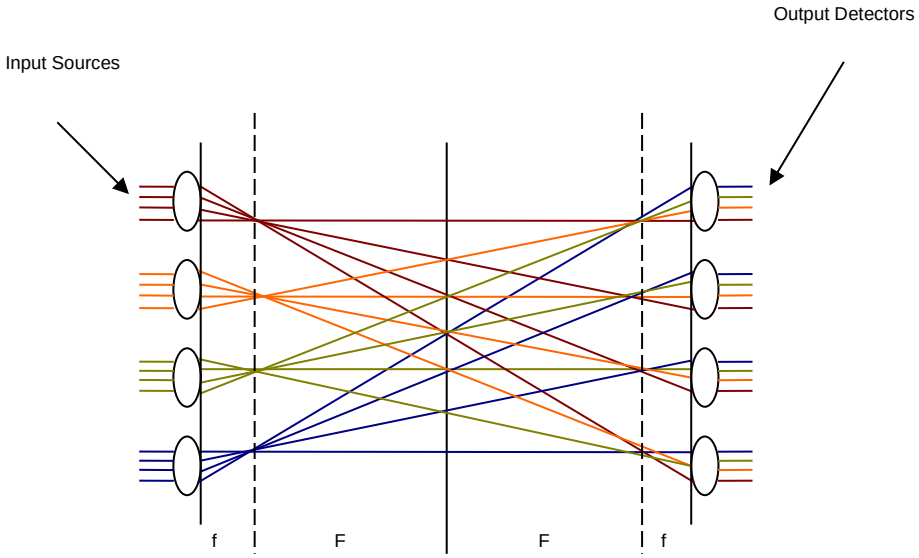


Figure 8.1 Free-space optical interconnect based on the design in [81].

The focus in the Chapter is on the design and evaluation of the central Fourier Transform lens performing the transposition to angularly multiplex beamlets [89]. The meso-lenses at the input array of the system perform the initial space-to-angle multiplex, with the output meso-lenses executing the output angle-to-space de-

multiplex. The architecture grows to a three stage system, the second stage being the central lens [87-88].

The architecture can be designed so that the beam waists at the input and output are equal and matched to input and output devices.

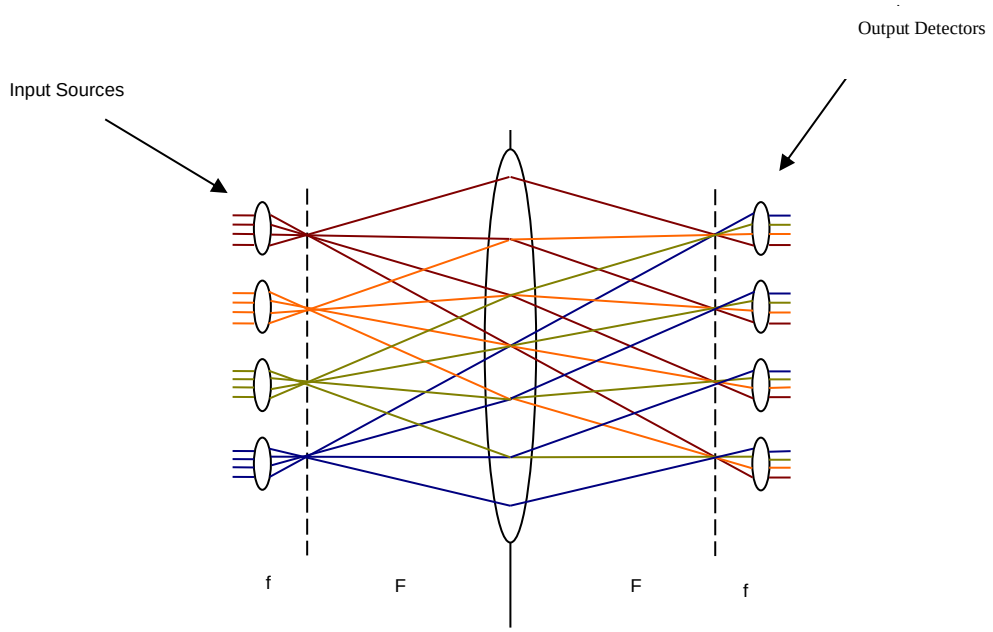


Figure 8.2 Free-space optical interconnect that utilises a Fourier Transform lens at its center.

Due to the limitations inherent to the optical modelling software, Micro-Electro-Mechanical Systems (MEMS) based scanning mirrors [88] are used for the beam steering function in the feasibility evaluation instead of the POGED-generated DOE. The MEMS emulation decreases the complexity of the code required to manageable bounds without loss of significance.

8.3 Central Fourier Transform Lens Design

On embarking on the design of a lens system, best practice dictates the setting of the target parameters viz. focal length, spacing according to the first order paraxial and thin lens approximations [93-94]. Although the approach does not yield a well corrected system, it is nevertheless core to determining the f-number and back focal length [93-94].

The principles from Bieren [92] are used as a starting position for the design of asymmetrical and symmetrical Fourier Transform lenses. The symmetrical Fourier Transform lens comprises six thin lenses capable of six degrees of freedom for correcting spherical, coma, astigmatism and field curvature. The fifth aberration, distortion, is corrected through the sine condition for the image and the symmetry of the lens system. Symmetry also allows the implementation of a reflective geometry [87-88].

Code V [95] is used to produce the first order layout for the Central Fourier Transform lens (Figure 8.3), assuming "perfect" thin lenses at the outset where thickness and the curvature effects are ignored and the first order paraxial calculations are applied.

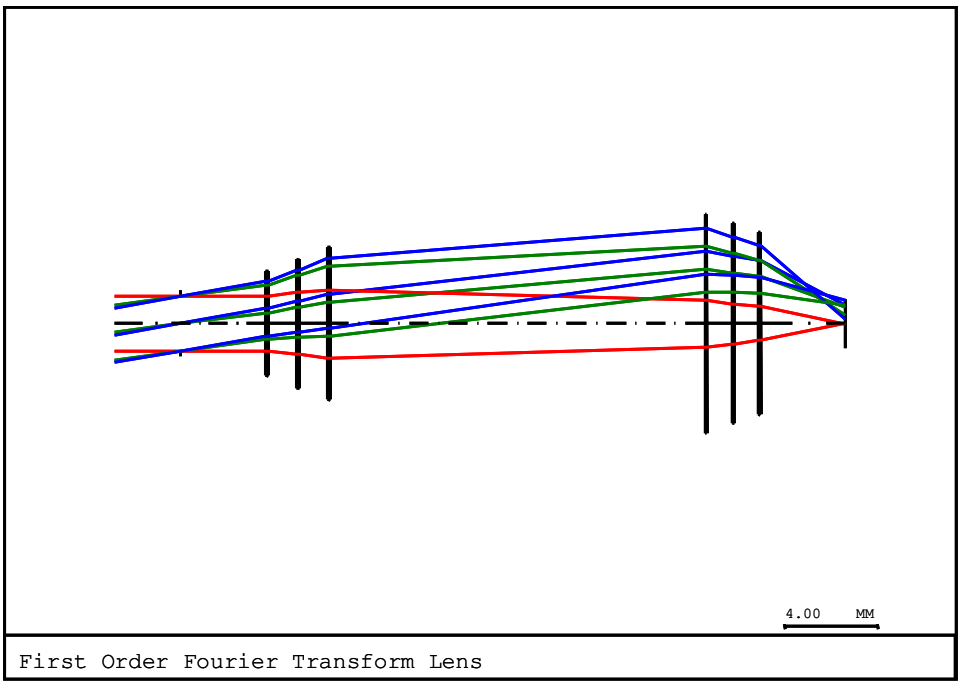


Figure 8.3 First order layout of the center lens system for optical interconnects.

To analyse the quality of the central Fourier transform lens or that of the FSOI, a number of performance metrics are used. The metrics are derived from ray tracing techniques and are described detail in the following references [93-94].

8.4 Central Fourier Transform Lens Design

A triplet design is used as a starting point where the negative elements are in the first and last positions and the positive element lies in the central position [89-90]. The separation between the 'input lens group' and 'output lens group' is approximately equal the focal length of the Fourier Transform lens (FT lens). The entrance pupil for the FT lens is in front of the lens where the micro-lens array is placed. To limit the size of the apertures, a field angle of 20˚ is used; this is an arbitrary value that can be larger or smaller depending on the desired complexity of the device. For

simplicity, this particular design considered a monochromatic source; the HeNe line is assumed to be the system wavelength. However the air gaps can be re-adjusted to wavelengths compatible with the low loss transmission region of optical fibres [97].

The aberrations of the lens system are minimised using RMS wavefront error, defined as the difference between the ideal and the aberrated wavefronts - these aberrations are described in Section 8.3 - with the entrance pupil set at 20 mm and fields set at 0 degrees, 7.5 degrees and 10 degrees. The focal length of the system is set to 20 mm and the overall length is also constrained to 60 mm; these two constraints are fixed to restrict system extent. If they are not fixed, the system or the focal length grow too large, yielding a system that may have high aberrations or it may not fit in the desired mechanical layout. Another dimension of the design is the separation between the two sets of lenses which must remain large, in the order of the focal length of the system [92].

Curvatures are optimised first and when the design is close to being diffraction limited, the thicknesses are then optimised. The optimisation progresses until only fractional improvements to the error function are being made. The design for the Fourier Transform lens is detailed in the data listed in Appendix A and illustrated in Figure .

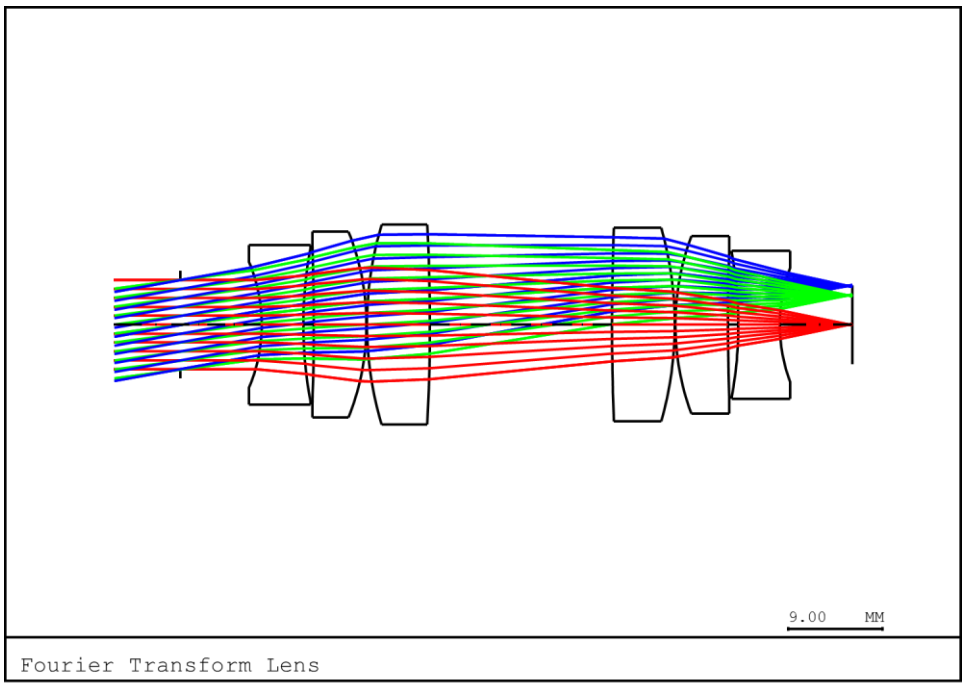


Figure 8.4 Symmetric Fourier Transform lens.

8.4.1 Performance of the Fourier Transform Lens

The system is evaluated using three techniques; ray aberration curves, the shape and magnitude of the curve determining the aberrations present in the system; RMS spot diagrams indicating aberrations through the shape of the focal spot the system produces; and the Modulation Transfer Function (MTF) which determines the resolution of the system [93-94].

The MTF for the Fourier Transform lens is presented in Figure 5, indicating that the system is diffraction limited out to 120 lp/mm. If multiple wavelengths are considered in the design, the MTF degrades and the system needs to be re-optimised due to chromatic aberrations. The SAG and TAN where the sagittal plane is the vertical plane and tangential is the horizontal plane, are illustrated in Figure 4 for different field positions. A separation of these two curves represents a number of

different aberrations such as astigmatism, lateral colour [93-94] and are only properly evaluated using ray aberration or through RMS spot size plots. In this case no separation is apparent.

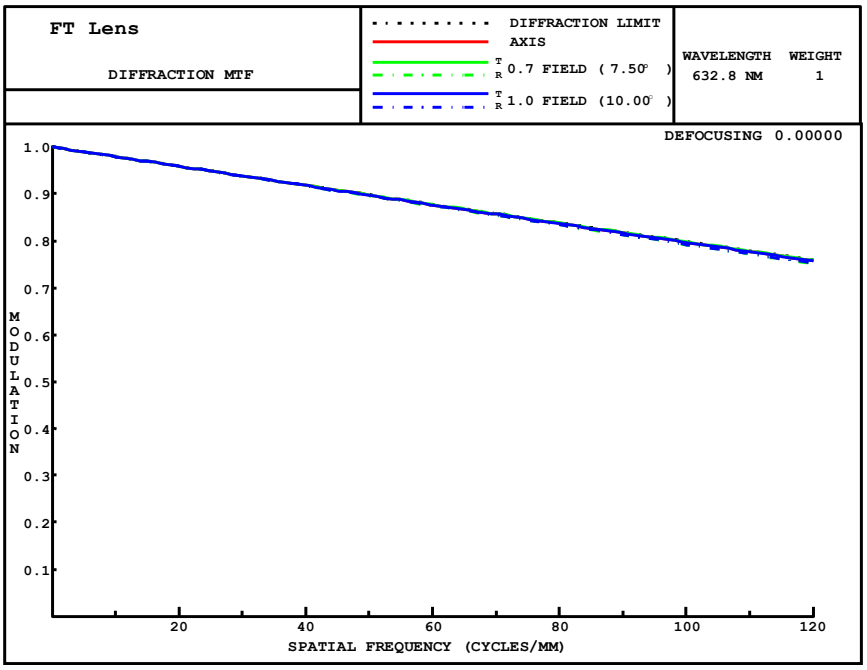


Figure 8.5 MTF for the Fourier Transform lens. The curve represents a diffraction-limited system.

Inspection of the ray aberration curve at different field angles (Figure 8.6) indicates that the fifth order spherical aberration dominates the system. However, the scale of these aberrations is less than one micron and does not significantly affect the quality of the transpose. Some astigmatism is also present but again, at a scale considered to be negligible.

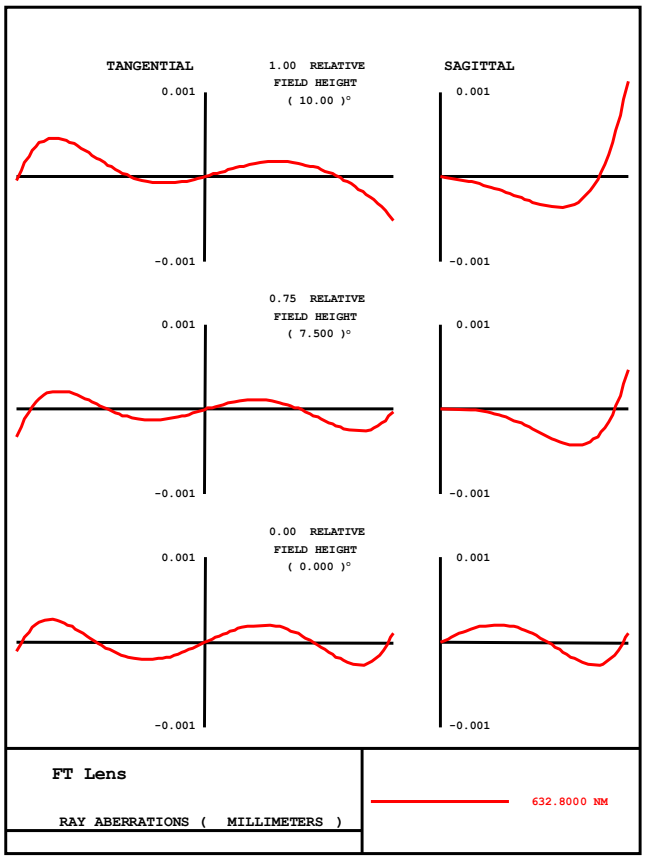


Figure 8.6 Ray aberration curves for different field angles of the Fourier Transform lens.

Figure 8.7 illustrates that the RMS spot sizes are free from monochromatic aberrations and are tightly focused, indicative of negligible levels of aberrations. For significant aberrations the focal spots are not as tightly focused e.g. for coma the spot follows a distinct comet like pattern that grows in size as the field angles become larger and for astigmatism the focal spot follows a distinct linear pattern in x or in y. The determination of lateral colour is not relevant in the design since monochromatic inputs are considered.

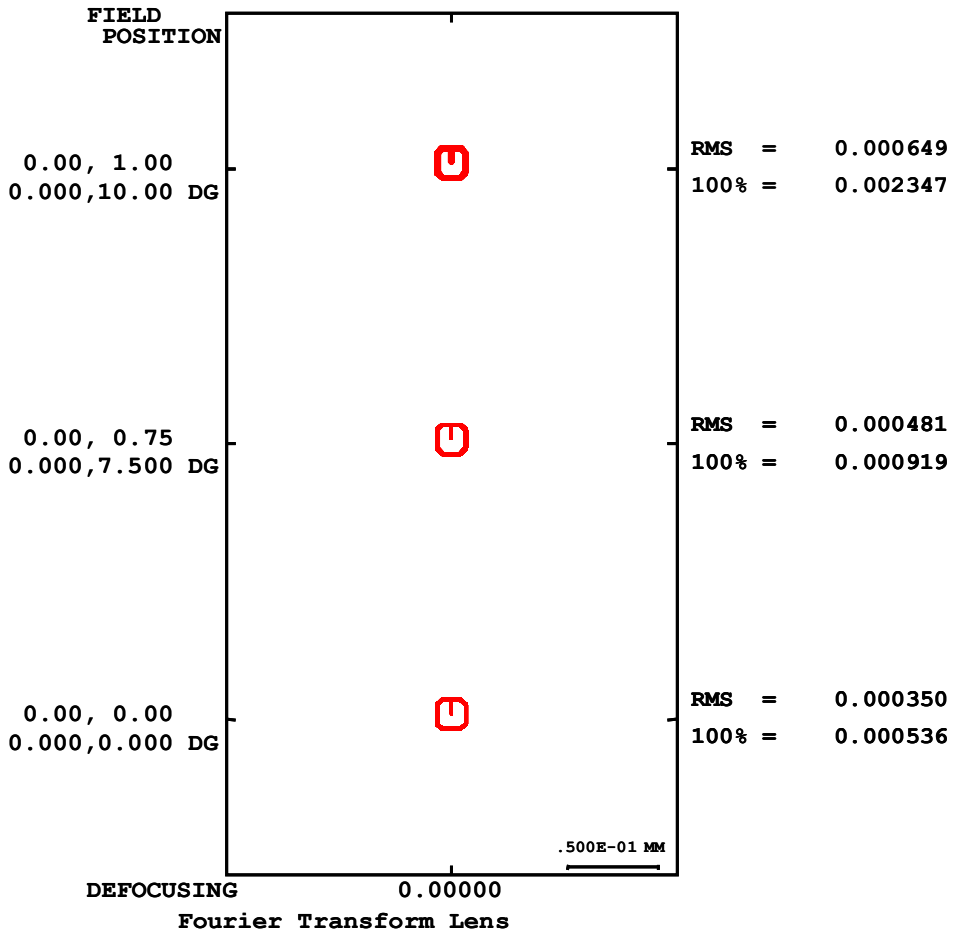


Figure 8.7 Spot diagram of Fourier Transform lens at different field points.

A symmetric diffraction limited Fourier Transform lens of 1 um focal spot size following the design principles detailed in [90] has been successful designed and its performance verified. Even though the system has been designed and analysed for an operational wavelength of 632 nm, it is possible to achieve a well corrected system at 1550 nm, most readily accomplished through changing the air gaps within the system to refocus.

8.5 Free Space Optical Interconnects (FSOIs)

The validation of the performance of the Central Fourier Transform lens generated using POGED and the ability of the DOE to effectively steer and shape an optical beam is the foundation for the modelling of the representative FSOI application. For modelling purposes in Code V, a MEMS scanning mirror based approach is adopted to demonstrate the feasibility of system operation. MEMS devices steer the input beam at the required angles to match the second set of meso-lenses and guide the outputs to either an optical fibre for onward transmission or to a detector. The second set of MEMS is also used to correct misalignment losses induced by the Fourier Transform lens but are not able to shape the beam or filter any stray light that may plague the system.

MEMS devices provide a way of emulating the beam steering functions within the FSOI through changes in the angle at which the beam is directed through the Fourier Transform lens. The mirrors are placed at the focal point of the meso-lens to maintain system symmetry. The second set of mirrors is placed at an equal spacing to the mirrors of the centre lens system, which in this case is also the location of the focus. The use of the DOE allows the beam to be reshaped. The mirrors are then re-oriented angularly to align with the optical axis in order to strike the detector at its required output port location (Figure 8).

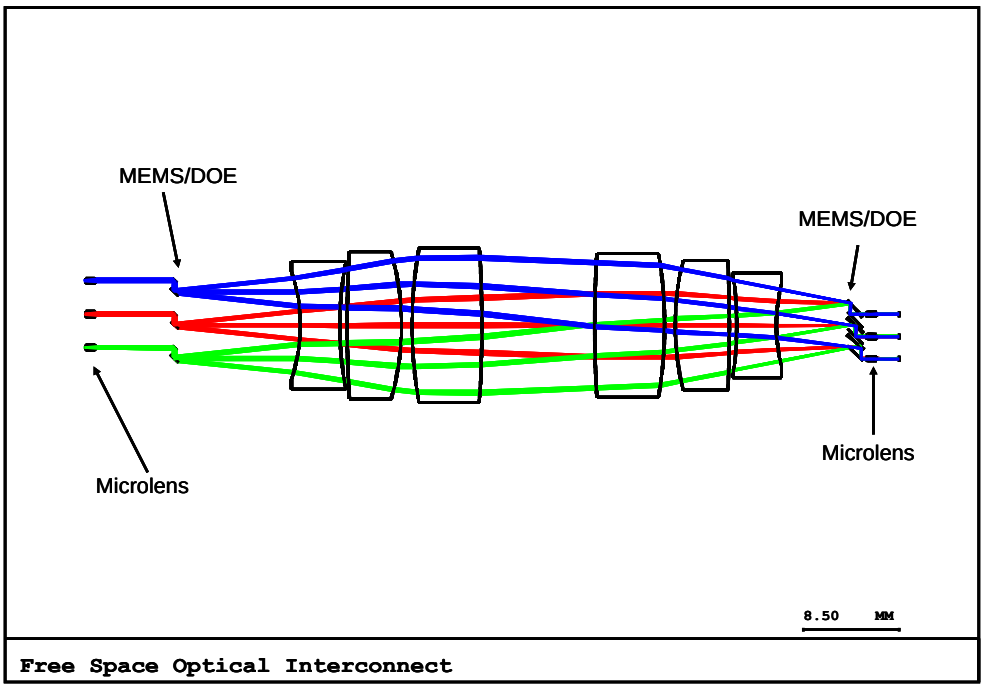


Figure 8.8 Layout for an unfolded FSOI.

Micro-lenses emulate Gaussian beam propagation, their design governed by both the type of micro-lens and the output optical fiber; in this case a Corning SMF-28e fiber is assumed [97]. Because the selected optical fiber operates at 1550 nm, the design is refocused to operate at this wavelength. The optical fiber has a mode field diameter of 10.4 µm at 1550nm and a Numerical Aperture (NA) of 0.14 [97]. The micro-lenses [98] are fused silica of thickness of 0.9mm, radius of curvature of 0.33 and NA of 0.17. The lenses are in an array form; but for the purpose of the emulation, only one is used.

8.5.1 Gaussian Beam Propagation

The beam synthesis software is used to propagate a Gaussian beam through the system from the output of the first to the input of the second micro lens array to verify beam shape and size.

Figure shows a Gaussian beam traced through the system. As it encounters the second set of MEMS mirrors, the beam has reached its focused position and is beginning to diverge. The second set of mirror lenses is placed just before the diameter of the beam becomes too large, causing the apertures of the mirrors to grow in size and interfere with mirrors adjacent to it. Figure shows the beam just before it enters the second set of micro lenses.

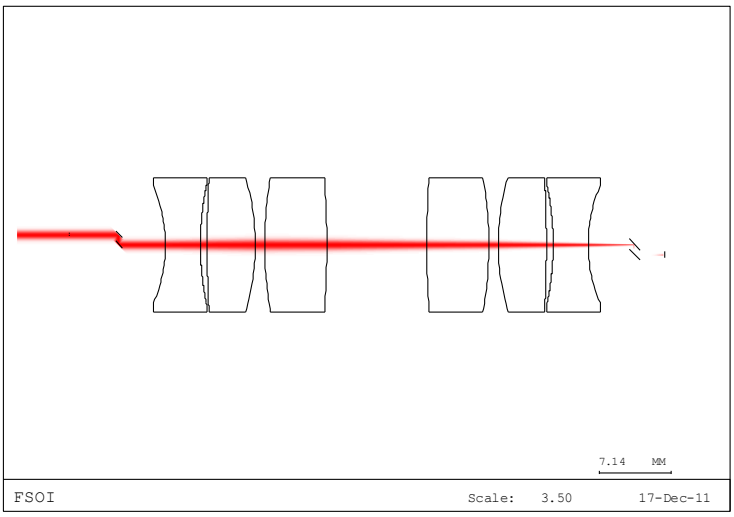


Figure 8.9 Gaussian beam propagating through the lens system starting from the output of the first to the input of the second micro lens.

The beam has a strong central peak surrounded by a large ring with an intensity of half the central peak. Figure 8.10. The ring around the central peak is produced by small aberrations in the system and the MEMS mirrors. With the current version of Code V, the coupling efficiency is calculated to be 25%; this however could be improved with a different program designed to optimise the coupling efficiency.

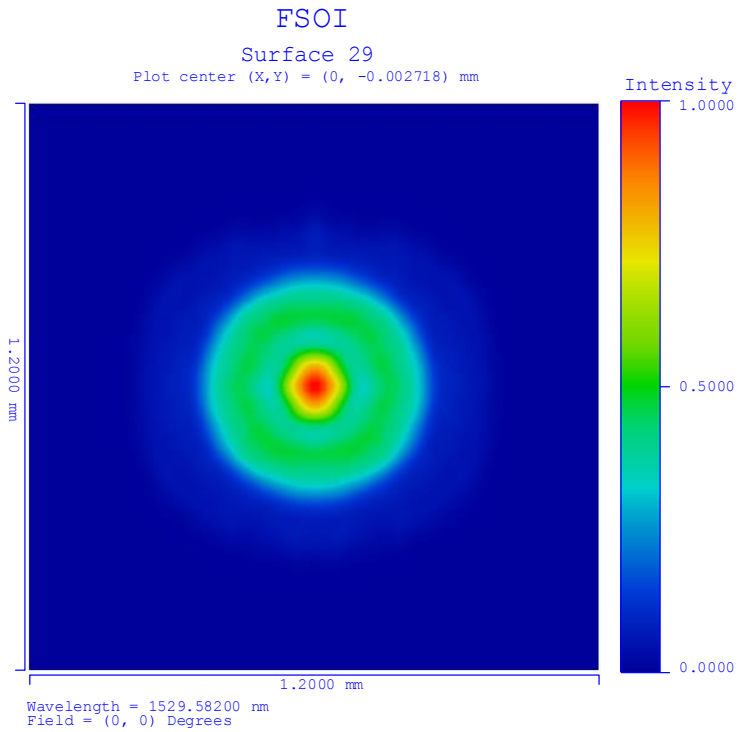


Figure 8.10 Output beam before the second micro lens array.

Overall the system is well corrected with few aberrations present in the Central Fourier Transform lens. The MEMS mirrors are steering the beam into the desired output positions and although slightly modulating the shape of the beam, it is nevertheless effectively coupled through the second micro lens array and in turn into the output optical fiber.

8.6 Conclusions

A symmetric Central Fourier transform lens has been designed and its performance validated. The lens is corrected for all primary aberration through the six lenses tandem and is diffraction limited with a 1 um focal spot size.

The lens is modelled within an FSOI application utilising MEMS as the beam steering mechanism in place of the desired DOE due to the limitation imposed by the available optical design software. The FSOI architecture is capable of steering the input beams to desired output locations with minimal aberrations present. Even if the aberrations present in the emulated system proved to be excessive, the use of the DOE located in the output plane would enhance performance since it provides the additional capability to reshape the beam correctly minimising existing aberrations.

Chapter 9

Conclusions and Future Work

9.1 Conclusions

The Thesis centres on the development of the Phase Optimised General Error Diffusion (POGED) algorithm for creating Diffractive Optical Elements (DOEs). These DOEs can be used in number of different application such beam steering, beam shaping, image projection, and phase retrieval, applications that require the reconstructed image to be of high quality and established in a relatively short time scale.

Simulated Annealing (SA) and the Direct Binary Search (DBS) algorithms have been extensively analysed and their associated limitation are well recorded viz. the DBS algorithm produces low SNR resulting from trapping within a local minimum and even though the SA algorithm is capable of finding a superior local minimum, the time required to find it is on the order of tens of minutes (Table 9.1). Although these algorithms are similar to POGED, their calculations are not executed in the same plane; POGED is calculated in the Fourier plane whilst the other algorithms are calculated in the plane of the target image. POGED also incorporates the use of a filter function that is continually updated as the algorithm progresses, whereas the DBS and SA algorithms utilise a fixed signal window.

Table 9.1 Comparison Table for DBS and SA for a binary DOE.

Algorithm	Time	Computed Diffraction Efficiency (%)	SNR (dB)
DBS	0:1:38	68%	10 dB
SA	0:21:38	69%	26dB

POGED has also been developed for operation on the far-field (Fraunhofer) and near-field (Fresnel) regions. In the far field region, characterisation through simulation for a number of different phase levels, ranging from binary to multi-level grating structures for a number of shapes for two performances metrics, diffraction efficiency and SNR has been carried out. Results indicate that the diffraction efficiencies are comparable to theoretical maximums and the SA algorithm, coupled to high fidelity manifest through high SNRs. A fourfold improvement in the calculation times is observed for POGED compared to SA (Table 9.2).

Table 9.2 Comparison Table for SA and POGED for a binary DOE.

Algorithm	Time	Computed Diffraction Efficiency (%)	SNR (dB)
SA	0:21:38	69%	26dB
POGED	0:01:00	72%	33dB

POGED has been modified for operation in the near-field region using the convolution/angular spectrum method for performing the transform. The method is selected because it eliminates high oscillation in frequency space caused by the *1/z* factor embedded within the integral method. The near field POGED produces

holograms comparable in diffraction efficiency with higher SNR to the benchmark Gerchbert-Saxton (GS) algorithm.

The time required to implement POGED in the near-field region is an order of magnitude larger than for the far-field region. The increase in the calculation time stems from the need to update function *ph*, describing the "diffusion" of errors created by the phase change at a specific pixel location in respect to its neighbours. If the function cannot be stored in memory via a lookup table, calculation times increase. The extent of the lookup table is in turn, limited by memory; a 32bit computer imposes limits to a 128x128 pixel structure. An analytical method for the update of the *ph* function would decrease calculation times.

Measurable improvements in fidelity and diffraction efficiency occur within the first 100 iterations, the remainder of the time spent correcting a small number of pixels with little further improvement in performance. The behaviour allows flexibility in controlling the exit condition, with concomitant decreases in calculation times. The same trend is observed with GS algorithms, however the POGED offers higher diffraction efficiency and fidelity.

The POGED mask function was investigated to determine optimal shape, size and offset value; the shapes investigated were a square, circle and the object. A range of mask sizes and offset values have been tested. Results identify trends; the larger the offset value, the higher the diffraction efficiency at the expense of a degraded SNR. The size and shape of the filter also plays a role in setting the diffraction efficiency

and SNR. For objects considered in this work, a square filter 10 pixels larger than the object improves performance. The filter must be carefully evaluated for the appropriate application, no matter the region of interest.

The enhanced features of POGED offering high fidelity, high diffraction efficiency in shorter time scales, act as a stimulus to a number of new and enhanced applications. An example of one application is Free Space Optical Interconnects (FSOIs) where beam steering and shaping DOE's can be utilised because time scales and quality of the reconstructed image are critical to performance. The FSOI feasibility evaluation utilised a Fourier Transform lens at its center to reduce aberrations and clipping caused the by the extreme off-axis geometry of some of the output ports. The Fourier Transform lens design is proven to be diffraction limited and corrected for primary aberrations, allowing for the minimisation of any aberrations induced by the lens system. Even though it is only modelled for monochromatic inputs, it is easily adapted for polychromatic operation. Due to incompatibilities with the optical design software it was not possible to include any DOEs generated by POGED; however an emulation to characterise the Fourier Transform lens through a MEMS scanning mirror based approach provides an indication of the feasibility of a DOE generated by POGED.

9.2 Future Work

The availability of multiple cores on graphical processing units (GPU) and parallel computing toolboxes brings great benefit through significant reductions in POGED calculation times. The utilisation of the advanced hardware/software will however require some modifications to the algorithm in respect of the appropriate execution

of the loop used in verification of the error function, stemming from the fact that GPUs must be programmed in parallel to take full advantage of the multiple cores. Other changes to the algorithm could also be investigated to take full advantage of the enhanced processing capabilities; one route is to partition the problem into smaller segments and recombine the results periodically for evaluation.

An application area where POGED could be potentially bring significant impact is in the field of optical metrology where algorithms are required to produce high quality data. The Iterative Fourier Transform Algorithm (ITFA) [26-27] and the Conjugate Gradient (CG) method [50] are two methods in current use. In practice however, the ITFA suffers from stagnation, leaving CG as the only practical option. It would be worthwhile investigating if POGED can be adapted to this application and compare performance with the incumbent CG method.

The ability to generate any high quality DOE in shorter time scales opens up a raft new applications where they can be effectively used.

References

[1] “Cisco Visual Index: Forecast and Methodology, 2012-2017” Tech Report, Cisco Inc, 2012

[2] R. Ramaswami and K. N. Sivarajan, *Optical Networks (Second Edition) A Practical Perspective*, San Francisco, CA, Academic Press, 2002

[3] K. Wang, A Nirmlalthas, C. Lim, E. Skafidas and K. Alameh “High-speed free-space based reconfigurable card-to-card optical interoconeects with broadcast capability,” *Optics Express Annual Meeting of the IEEE*, vol. 2 no. 13, pp. 15395-15400 (2013)

[4] B. Ciftcioglu, R. Berman, J Zhang, Z. Darling, S. Wang, J. Hu, J. Xue, A. Garg, M. Jain, I. Savidis, D, Moore, M. Huang, E. G. Friedman, G. Wicks and H. W, “A 3-D integrated intrachip free-space optical interconnect for many core chips,” *IEEE Photonics Technology Letters*, vol. 23 no. 3, pp. 164-166 (2011)

[5] G. Tian, R. Nair, M.W. Haney, “Integrated free-space optical interconnects: All optical communications on- and off-chip,” *Optical Interconnects Conference, 2012 IEEE* , pp., 20-23 (2012)

[6] C. Henderson, D. Gil-Leyva and T. D Wilkinson “Free Space Adaptive Optical Interconnect at 1.25 Gb/s, With Beam Steering Using a Ferroelectric Liquid-Crystal SLM,” *Journal of Lightwave Technology*, vol. 24 no. 5, pp. 1989-1997(2006).

[7] H. P. Kuo, P. Rosenberg, R. Walmsley, S. Mathai, L. Kiyama, J. Straznicky, M. Mclaren, M. Tan, and S.-Y. Wang, “Free-space optical links for board-to-board interconnects,” *Applied Physics A*, vol. 95 no. 4, pp. 955-965 (2009)

[8] D.A.B. Miller, “Rationale and challenges for optical interconnects to electronic chips,” *Proceedings of the IEEE*, vol. 88, no. 6, pp. 728-749 (2000)

[9] S. Sultana, F. M. Shahriar, M. K. Hasan, “Chip-to-chip Free-space Optical Interconnection Using Liquid-crystal-over-silicon Spatial light modulator”, *Technological Developments in Networking, Education and Automation*, pp. 507-510-435 (2010)

[10] Michael J. McFadden, Muzammil Iqbal, Thomas Dillon, Rohit Nair, Tian Gu, Dennis W. Prather, and Michael W. Haney, “Multiscale free-space optical interconnects for intrachip global communication: motivation, analysis, and experimental validation”, *Applied Optics*, vol.45, no. 25, pp. 6358-6366 (2006)

[11] J. Chou, K. Yu, D. Horsley, B. Yoxall, S. Mathai, M. R. T. Tan, S.-Y. Wang, M. C. Wu , “Robust free space board-to-board optical interconnect with closed loop MEMS tracking”, *Applied Physics A*, vol. 95 39, no. 4, pp. 973-982 (2009)

[12] A.G. Kirk, “Free-Space Optical Interconnects”, *Optical Interconnects, Springer Series in Optical Sciences*, vol. 119, 5, pp. 344-377 (2006)

[13] Boulder Nonlinear XY 512 Nematic SLM, available at: http://www.bnonlinear.com/

[14] J. R. Fienup, "Iterative method applied to image reconstruction and to computer-generated holography," *Optical Engineering* vol. 19 no. 3, pp. 297-305 (1980)

[15] J. R. Fienup, “Phase retrieval algorithms: a comparison”, *Applied Optics* vol. 21 no. 15, pp. 2758-2769 (1982).

[16] J. R. Fienup, "Phase retrieval algorithms: a personal tour [Invited]," Applied Optics vol. 52, 45-56 (2013)

[17] J. H. Crocker, H. C. Ford, G. F. Hartig, and R. I. Jedrzejewski, "Optical performance of the Corrective Optics Space Telescope Axial Replacement (COSTAR)," *Proceedings SPIE: Instrumentation in Astronomy VIII*, vol. 2198, pp. 1170-1180 (1994).

[18] J. S. Liu and M. R. Taghizadeh, "Iterative algorithm for the design of diffractive phase elements for laser beam shaping," *Optics Letters*, vol. 27 no. 16, pp. 1463-1465 (2002).

[19] J. S. Liu, A. J. Caley, and M. R. Taghizadeh, "Diffractive optical elements for beam shaping of monochromatic spatially incoherent light", *Applied Optics* vol. 45 no. 33, pp. 8440-8447 (2006)

[20] J. S. Liu, A. J. Caley, M. R. Taghizadeh, E Gu, J. M. Girkin and M. D. Dawson, "Design of diffractive optical elements for beam shaping of micro-pixellated LED light to a tightly focused spot," *Journal of Physics D: Applied Physics*, vol. 41 (2008)

[21] J. Cordingley, "Application of a binary diffractive optic for beam shaping in semiconductor processing by lasers," *Applied Optics* vol. 32 no. 14, pp. 2538-2542 (1993).

[22] D. Grewell, and A. Benatar, "Diffractive optics as beam-shaping elements for plastics laser welding," *Optical Engineering*, vol. 46 no. 11, (2007)

[23] B. W. Kolpatzik and C. A. Bouman "Optimized error diffusion for image display," J. Ele. Img. Am. A vol. 1 no. 3, pp. 277-292 (1992).

[24] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of the phase from image and diffraction plane pictures," *Optik* vol. 35, (1972)

[25] E. G. Johnson and M. A. G. Abushagur, “Microgenetic-algorithm optimization methods applied to dielectric gratings,” *Journal Optical Society of America A*, vol. 12 no. 5, pp. 1152-1160 (1995)

[26] H. Kim, B. Yang, and B. Lee, “Iterative Fourier transform algorithm with regularization for the optimal design of diffractive optical elements,” *Journal Optical Society of America A*, vol. 21 no. 12, pp. 2353-2365 (2004)

[27] F. Wyrowski and O. Bryngdahl “Iterative Fourier-transform algorithm applied to computer Holography,” *Journal Optical Society of America A*, vol. 5 no. 7, pp. 1058-1065 (1988).

[28] M. A. Seldowitz, J. P. Allebach, and D. W. Sweeney, "Synthesis of digital holograms by direct binary search," *Applied Optics*, vol. 26 no. 14, pp.2788-2798 (1987).

[29] B. K. Jennison, J. P. Allebach, and Q. W. Sweeney, "Efficient design of direct-binary-search computer-generated holograms," *Journal Optical Society of America A*, vol. 8 no. 4, pp. 652-660 (1991).

[30] A G Kirk, T J Hall, “Design of Computer Generated Holograms by Simulated Annealing: Coding Density and Reconstruction Error”, O*ptics Communications*, vol. 94 no. 6, pp. 491-496 (1992)

[31] A. G. Kirk and T. J. Hall, “Design of binary computer generated holograms by simulated annealing, Observation of metastable states,” *Journal of Modern Optics*, vol. 39 no. 12, pp. 2531-2539 (1992).

[32] F. Fetthauer, S. Weissbach, O. Bryngdahl, Equivalence of error diffusion and minimal average error algorithms, Optics Communications, vol. 113, no. 4–6, pp. 365-370 (1995)

[33] K. Heggarty and R. Chevallier, “Signal window minimum average error algorithm for computer-generated holograms,” *Journal Society of America A*, vol. 15, no. 3 pp. 625-635 (1998)

[34] R.W. Floyd, L. Steinberg, “An adaptive algorithm for spatial grey scale,” *Proceedings of the Society of Information Display*, vol. 17, no. 2 pp. 75-77 (1976).

[35] M. Born, and E. Wolf, *Principles of Optics 6th Ed*., Cambridge University Press, N.Y., (1980).

[36] E. Hecht, *Optics 3rd Ed*., Reading Massachusetts, Addison-Wesley, Inc., 1987.

[37] G. Tricoles, “Computer generated holograms: an historical review,” *Applied Optics*, vol. 26 no. 20, pp.4351-4357 (1987).

[38] J. W. Goodman, *Introduction to Fourier Optics 2nd Ed*., New York, McGraw-Hill, Inc., 1996.

[39] Kress, B.; Meyrueis, *Digital Diffractive Optics: An Introduction to Planar Diffractive Optics and Related Technology*, New York, Wiley-VCH, 2000.

[40] D. C. O’Shea, T. J. Suleski, A. D. Kathman and D. W. Prather, *Diffractive Optics – Design, Fabrication and Test*, Bellingham, Washington , SPIE Press, 2003

[41] B. R. Brown and A. W. Lohmann, "Computer-generated binary holograms," *IBM Journal of Research and Development*, vol. 13, pp. 160-168 (1969).

[42] B. R. Brown and A. W. Lohmann, "Complex spatial filtering with binary mask," *Applied Optics*, vol. 5 no. 6, pp. 967-969 (1966).

[43] J. A. Jordan, Jr., P. M. Hirsch, L. B. Lesem, and D. L. Van Rooy,” Kinoform Lenses,” *Applied Optics*, vol. 9 no. 8, pp. 1883-2887 (1970)

[44] W. H. Lee, “Binary computer-generated holograms,” *Applied Optics*, vol. 18 no. 21, pp. 3661-3669 (1979).

[45] W. H. Lee, “Binary synthetic holograms,” *Applied Optics*, vol. 13 no. 7, pp. 1677-1682 (1974).

[46] W. H. Lee, “Sampled Fourier transform holograms generated by computer,” *Applied Optics*, vol. 9 no. 3, pp. 639-643 (1970)

[47] W. H. Lee, “Computer-generated Hologram: Techniques and applications,” in *Progress in Optics, XVI, E. Wolf, ed*, pp. 119-232 (North-Holland, Amsterdam, 1978)

[48] A. W. Lohmann, D. P. Paris, “Binary Fraunhofer holograms generated by computer," *Applied Optics*, vol. 6 no. 10, pp. 1739-1748 (1967).

[49] RPC Photonic, available at: http://www.rpcphotonics.com/

[50] B. Bhuian, R.J. Winfield, S. O’Brien, G.M. Crean, “Pattern generation using axicon lens beam shaping in two-photon polymerisation,” *Applied Surface Science*, vol. 254 no. 4, pp. 841-844 (2007)

[51] B. Dean, “*Hybrid Diversity Method Utilizing Adaptive Diversity Function for Recovering Unknown Aberration in an Optical System*,” 22 December 2009, United State Patent US 7,635,832 B2

[52] J. R. Fienup, "Gradient-Search Phase-Retrieval Algorithm for Inverse Synthetic-Aperture Radar," Optical Engineering. vol. 33 issue 10, pp. 3237-3242 (1994).

[53] G. Brady, *Application of Phase Retrieval to the Measurement of Optical Surfaces and Wavefronts*, PhD Thesis, University of Rochester, New York, 2008

[54] A.G. Kirk, K. Powell, and T. J. Hall, "Error diffusion and the representation of the problem in computer generate holograms," *Optical Computing and Processing*, vol. 12 no. 3, pp. 199-212 (1992)

[55] A. G. Kirk, K. Powell and T. J. Hall "A generalization of the error diffusion method for binary computer generated hologram design," O*ptics Communications*, vol. 92 no. 1-3, pp. 12-18 (1992)

[56] J. L. Ramsey and T. J. Hall "POGED: A fast alternative to Computer Generated holograms design by simulated annealing," *Diffractive Optics, Warsaw, Poland* (2005)

[57] F. Wyrowski, and O. Bryngdahl, "Digital Holography as part of diffractive optics," *Report on Progress in Physics*, vol. 54 no. 12, pp.1481-1571 (1991)

[58] C. I. Podilchuk and R. J. Mammom, " Image recovery by convex projections using least square constraints", *Journal Optical Society of America A*, vol. 7, no. 3 pp. 517-531 (1990)

[59] S. – J. Yeh and H. Stark, "Iterative and one-step reconstruction from nonuniform samples by convex projections," *Journal Optical Society of America A*, vol. 7, no. 3 pp. 491-499 (1990)

[60] H. Bauschke, P.L Combettes, and D. Russell Luke, "Phase retrieval, error reduction algorithm and Fienup variants: a view from convex optimization," *Journal Optical Society of America A*, vol 19 no. 7, pp. 1334-1345 (2002).

[61] H. H. Bauschke and J. M. Borwein, "Dykstra's alternative projection algorithm for the two sets," *Journal of Approximation Theory*, vol. 79 no. 3, pp. 418-443 (1994).

[62] Matlab R2007b available at: http://www.mathworks.com/products/matlab/

[63] P. Senthilkumaran, F. Wyrowski and H. Schimmel, "Vortex stagnation problem in Iterative Fourier Transform Algorithms," *Optics and Lasers in Engineering*, vol. 43 no. 1, pp. 43-56 (2005).

[64] S. Morucci, P. Noirard, J.-C. Grossetie, "Improvement of Fresnel computer-generated holograms." O*ptics Communications*, vol 134 issue 1-6**,** pp. 25-30, (1997)

[65] M. Makowski, M. Sypek, A. Kolodziejczyk, G. Mikuła, J. Suszek, "Iterative design of multiplane holograms: experiments and applications," *Optical Engineering*, vol. 46 no. 4, pp. (2007)

[66] M. Sypek, "Light propagation in the Fresnel region-New numerical approach," O*ptics Communications*, vol. 116 issue 1-3, pp. 43-48 (1995)

[67] A. VanderLugt, "Optimum sampling of Fresnel transforms," *Applied Optics*, vol. 29 no. 23, pp. 3352-3361 (1990).

[68] J. L Ramsey, V.P. Sivokon, T.J. Hall, I. Andonovic and C. Michie "Fresnel Based Phase Optimised General Error Diffusion Algorithm for Optical Beam Shaping", *Proceeding SPIE: International Optical Design Conference 2010*, vol. 7652, (2010)

[69] P. Rai-Choudhury, *Handbook of Microlithography, Micromachining, and Microfabrication. Volume 1: Microlithography*, Washington State, SPIE Press, Book, 1997

[70] R. C. Jaeger, *Introduction to Microelectronic Fabrication: Volume 5 of Modular Series on Solid State Devices (2nd Edition),* New Jersey, Prentice Hall, 2001

[71] M. J. Thomson, M.R. Taghizadeh, “Design and fabrication of Fourier plane diffractive optical elements for high-power fibre-coupling applications,” *Optics and Lasers in Engineering,* vol. 43 no. 6, pp. 671–681 (2005)

[72] M. R. Taghizadeh, P. Blair, B. Layet, I. M. Barton, A. J. Waddie, and N. Ross, “Design and Fabrication of Diffractive Optical Elements,” *Microelectronic Engineering*, vol. 34 no. 3-4, pp. 219-242 (1997).

[73] M. Ekberg, M. Larsson, S. Hard, and B. Nilsson, “Multilevel phase holograms manufactured by electron-beam Lithography,” *Optics Letters*, vol 15 no. 10, pp.568-569 **(**1990)

[74] F. Watt, A. A. Bettiol, J. A. Van Kan, E. J. Teo and M. B. H. Breese, "Ion Beam Lithography and Nanofabrication: a Review,” *International Journal of Nanoscience*, vol. 4 no. 3, pp. 269 (2005).

[75] G. Della Valle, R. Osellame, and P. Laporta, , "Micromachining of photonic devices by femtosecond laser pulses," *Journal of Optics A: Pure and Applied Optics*, vol. 11, (2009).

[76] ISP Optics, available at: http://www.ispoptics.com/articles/6/IR%20OPTICAL%20MATERIALS

[77] R. T. Blunt, “White Light Interferometry – a production worthy technique for measuring surface roughness on semiconductor wafers”, *International Conference on Semiconductor Manufacturing*, Vancouver BC, pp. 59- 62 (2006)

[78] G. Binnig, C. F. Quate and Ch. Gerber‡, "Atomic Force Microscope". Physics Review Letters, vol. 56 no. 9, pp. 930-933 (1986).

[79] K. C. A Smith and C. W Oatley, "The scanning electron microscope and its fields of application". *British Journal of Applied Physics* vol. 6 no. 11 (1955).

[80] Newport LBP1 available at: http://search.newport.com/?q=*&x2=sku&q2=LBP-1-USB

[81] Olympus Upright Microscope available at : http://www.olympusamerica.com/seg_section/seg_microscopes.asp?section=upright

[82] Nanonics Multiview 2000 AFM available at: http://www.nanonics.co.il/

[83] Zygo whitelight interferometer and Metropro available at: http://www.zygo.com/?/met/profilers/

[84] Canon PowerShot 6 Mpixel camera available at: http://www.canon.ca/inetCA/subCategoryHome?msegid=2&catid=17&scatid=18

[85] G. C. Marsden, P. J. Marchand, P. Harvey, and S. C. Essener, "Optical Transpose Interconnection System Architectures," *Optics Letters*, vol. 18 no. 13, pp. 1083-1085 (1993)

[86] P. Milojkovic, M. P. Christensen, and M. W. Haney, "Minimum Lens Complexity Design Approach for a Free-space Macro-optical Multi-chip Global Interconnection Module," *Proceedings SPIE: In Optics in Computing*, vol. 4089, (2000)

[87] T. J. Hall, '*Optical Transpose System*', 16 August 2000, UK Patent Application GB0020046.9, International Patent Application PCT/GB01/03643,.Published as WO02/15633 A1. USA Patent 6,898,013 granted 24 May, 2005

[88] T. J. Hall, '*Optical crossbar switch*', 7 November 2001, UK Patent Application, GB0126822.6, International Patent Application PCT/GB02/04990, Published as WO 03/041421 A3.

[89] W. Pijitrojana, *Optical Transpose Interconnection System Design in Optical Phase Space*, PhD Thesis, Kings College London, London 2003

[90] J. L. Ramsey, W. Pijitrojana and T. J. Hall, "Novel Lens Design for Free-Space Optical Interconnects", *Proceedings SPIE: International Optical Design Conference*, vol. 6342, (2006)

[91] S.-S. Lee, L.-S. Huang, C.-J. Kim, and M. C. Wu, "Free-space fiberoptic switches based on MEMS vertical torsion mirrors," Journal of Lightwave Technology, vol. 17 no.1 , pp. 7-13 (1999).

[92] K. von Bieren, "Lens Design for Optical Fourier Transform Systems", *Applied Optics*, vol. 10 no. 12, pp. 2739-2742 (1971).

[93] M. J. Kidger, *Fundamental Optical Design*, Bellingham, Washington, SPIE Press, 2002.

[94] W. J. Smith, *Modern Optical Engineering 3rd Ed.*, N.Y., SPIE Press McGraw-Hill Inc., 2000.

[95] Code V, version 9.6, developed by Optical Research Associates, available at: http://optics.synopsys.com/codev/

[96] A. J. Dragt, "A Lie connection between Hamiltonian and Lagrangian optics," *Discrete Mathematics and Theoretical Computer Science*, vol. 1, pp. 149-157 (1997)

[97] Corning Datasheet PI1446, April 2005, available at: http://www.corning.com/index.aspx

[98] SUSS Micro-Optics FC-Q-250 Microlens array, available at: http://www.suss-microoptics.com/

Appendix A

```
%function poged%function poged
% GED algorithm
% Multi-levels
% Phase optimisation
% Greedy quantisation
% FTs with origin at array centre

% The PCS algorithm is run first to construct a mask
% In general for a fixed signal window this can be done off-line and
the mask read-in

clear all
close all

%--------------------------------------------------------------------
% Algorithm setup parameters
%--------------------------------------------------------------------
pi     = 4*atan(1);
n      = 128;
m = 128;
nc     = fix(n/2)+1;   %central pixel
dia    = 2;        %Grid size, dimensionless
a_beam = 3;        %beam radius, dimensionless
dx     = dia/n;    %Grid step = pixel size
df     = 1/dia;    %Grid step in Freq domain = Freq pixel size in
Fourier plane
z      = 0.05;     %dimensionless propagation distance
%NOTE:
% x_dimensionless = X/a_beam, where a_beam = effective radius of the
beam
% z_dimensionless = Z/Zd, where Zd=k*BeamRad^2 and k=2pi/lambda
eps    = 0.1;      %regularazation parameter
nit    = 1000;     %iterations limit for Filter generation
it_max = 50;      %iterations limit for Hologram generation
%amplitude at z=0 and z=Z
beam0type = 1;     %0=uniform, 1=gauss, 2=rect super-gauss, 3=circle
super-gauss
beamztype = 3;     %0=gauss, 1=rect super-gauss, 2=circle super-
gauss,
                   %3=letter T on-axis, 4= letter T off axis
init_phase= 0;     % 0=zero, 1 = random in  +/-(init_range)*pi
init_range= 0.5;   % range of random phase: +/-(init_range)*pi

debug     = 0;     % print (=1) or not intermediate results

%parameters for Target
t_type     = 0; % t_type = define type of target to create (0=cross,
1= letter T)
adjustL    = 1; %if 0 use nBorder=3, nHalfWidth=10 for 21x21 grid of
Target
nHalfWidth = fix(a_beam/dx); %
nBorder    = 3; %border width in poxels

% hoe phase parameters
levels = 16;       % number of quantisation levels in n-ary case
```

```
step   = 2*pi/levels;
tol    = 0.5*step; % tolerance

[x,y]   = meshgrid((-0.5*dia) : dx : 0.5*dia-dx);
[fx,fy] = meshgrid((-0.5*n*df): df : 0.5*n*df-df);
qphase  = z*(fx.^2+fy.^2);
qph     = ifftshift(qphase);
qf      = exp(1i*2*pi^2*qph);
Cqf     = conj(qf);

%defines Filter window size at z=0.
widthx=n/4; %16; % recognisable
widthy=n/4; %16; % recognisable
% co-ordinates
bigorgx   = nc;
bigorgy   = nc;
smallorgx = fix(widthx/2)+1;
smallorgy = fix(widthy/2)+1;
lx        = bigorgx-smallorgx+1;
ly        = bigorgy-smallorgy+1;
ux        = lx+widthx-1;
uy        = ly+widthy-1;


%initialize arrays for history of Hologram error and Diffraction
efficiency
errH_norm = (1:it_max);
errH_norm = errH_norm*0.0;
errPH_norm  = (1:it_max);
errPH_norm  = errPH_norm*0.0;

%h0=field amplitude at z=0, F0=required (target) amplitude at z=Z
% h0    = zeros(n,n);
F0    = zeros(n,n);
x2    = (x/a_beam).^2;
y2    = (y/a_beam).^2;
rad2  = x2+y2;
if beam0type == 0
  h0 = ones(n,n);
elseif beam0type == 1
  h0 = exp(-rad2);          %gauss
elseif beam0type == 2
  h0 = exp(-x2.^10).*exp(-y2.^10);     %rect super-gauss
else
  h0 = exp(-rad2.^10);    %circle super-gauss
end
% normalize energy to 1
% h0 = h0/sqrt(sum(sum(h0.^2)));
% h0 = ones(n,n); %for test only
% h0 = phase;
fprintf('\n <h,h> = %g', sum(sum(h0.^2)));

offnc   = 0;

[T,nL]    = CreateTarget(t_type,nHalfWidth, nBorder, adjustL);
% load psf_irr2
% T = psf_irr2;
```

```
[dxx,dyy] = size(T);

% if beamztype == 0
%   F0 = exp(-rad2);          %gauss
% elseif beamztype == 1
%   F0 = exp(-x2.^10).*exp(-y2.^10);  %rect super-gauss
% elseif beamztype == 2
%   F0 = exp(-rad2.^10);      %circle super-gauss
% elseif beamztype == 3      % Letter T embedded in larger array
(on-axis)
% %   co-ordinate of LLH corner of signal window relative to origin
%   offx          = -fix(dxx/2);
%   offy          = offx;
%   offnc         = 0; %fix(dxx/2)+1;
% %   offx & offy = co-ordinate of LLH corner of signal window
relative to origin
%   nearx         = bigorgx+offx;
%   neary         = bigorgx+offy;
%   farx          = nearx+dxx-1;
%   fary          = neary+dyy-1;
%   F0(nearx:farx,neary:fary)=T;
% else                         % Letter T embedded in larger array
(off-axis)
% %   co-ordinate of LLH corner of signal window relative to origin
%   offx=0;
%   offy=0;
%   offnc         = fix(dxx/2)+1;
% %   offx & offy = co-ordinate of LLH corner of signal window
relative to origin
%   nearx         = bigorgx+offx;
%   neary         = bigorgx+offy;
%   farx          = nearx+dxx-1;
%   fary          = neary+dyy-1;
%   F0(nearx:farx,neary:fary)=T;
% end
[D, F0, h0, phi] = Zer_FT(n, dia, -0.006, 0.006, -0.01, 0.014, -
0.335, -.305, .21, -.505, -.166, -.014, .425, .11, -.079, -.09,
.193, -.007, .069, .051, .043, -.11, -.216, .118, .015, -.065, -
.036, .002, .003, -.033, -.07, .037, .075, .092, -.066, .004, .034,
.000);
%
phi = angle(D);


% normalize energy to 1
F0 = F0/sqrt(sum(sum(F0.^2)));
fprintf('\n <F,F> = %g', sum(sum(F0.^2)));
Fcutoff      = 0.01; %cutoff at 1%
a            = zeros(n,n);
a(F0>Fcutoff*max(max(F0)))= 1-eps;
a = a + eps;
d=a;
% display hard filter
figure(1);
D=a+eps;
D = 255*D/max(max(D));
image(D);
colormap(gray(256));
```

```
axis square;
axis xy;

%-----------------------------------------------------------------------
% construction of soft mask & filter function module
% Dykstra's alternating projections algorithm
%-----------------------------------------------------------------------

% acc=eps/10; % slow
acc=eps/2; % faster
nit=1000;
c=0;
it=0;
finished=0;
while ( not(finished) & (it<nit))
   it=it+1;

   % project a onto set of functions > 0
   b=a+c;
   k=find(a<0);
   c=b;
   b(k)=0;
   c=c-b;

   % project onto the subspace of bandlimited functions
   p=ift(b);

   q=zeros(n,m);
   q(lx:ux,ly:uy)=p(lx:ux,ly:uy);

   a=real(ft(q));

   if (min(min(a))>-acc)
      finished=1;
   end
end

% check the similarity
similarity=sum(sum(a.*d))/(sqrt(sum(sum(a.*a)))*sqrt(sum(sum(d.*d)))
)

P=a+eps;
peak=max(max(P));
trough=min(min(P));

Pmax = peak;
Pmin = trough;
fprintf('\n Filter function generated in %d iterations',it);
fprintf('\n Pmax = %g, Pmin = %g, Similarity =
%g',Pmax,Pmin,similarity);
fprintf('\n');

% extract mask p

q=ift(P);
p=q(lx:ux,ly:uy);
```

```
% display soft filter and mask
%
% % soft filter
% figure(2);
% D = P;
% D = 255*D/max(max(D));
% image(D);
% colormap(gray(256));
% axis square;
% axis xy;
%
% % soft mask with supressed central peak
% figure(3);
% D=abs(p);
% D(smallorgx,smallorgy)=0;
% D = 255*D/(max(max(D)));
% image(D);
% colormap(gray(256));
% axis square;
% axis xy;
% %--------------------------------------------------------------------
% % poged module
% %--------------------------------------------------------------------
% t1=cputime;
% rand('state',sum(100*clock));
%
% % fixed parameters

pi=4.0*atan(1);
levels = 16; % number of quantisation levels in n-ary case
tol = pi/64; % tolerance (~ 128 levels)
step = 2*pi/levels;

p00=p(smallorgx,smallorgy);
x=(1:widthx)-smallorgx-1;
y=(1:widthy)-smallorgy-1;

% initialise target
F = F0;
% F = abs(F).*(exp(complex(0,2*rand(n)*pi))); % complexify with a
random phase
F = abs(F).*(exp(complex(0,1))); % complexify with a random phase
F = F/sqrt(sum(sum(P.*(F.*conj(F))))); % normalise too simplify
expression for alpha

f = n*ift(F); % new target object

% initialise hologram as hard-clipped target

% binary

h=sign(real(f));

% quaternary

%for i=1:n
```

```
    %for j=1:n
         %hr=real(f(i,j));
         %hi=imag(f(i,j));
         %if abs(hr)>abs(hi)
             %h(i,j)=complex(sign(hr),0);
         %else
             %h(i,j)=complex(0,sign(hi));
         %end
    %end
%end

% n-ary

% h = h0.*exp(complex(0,step*mod(round(atan2(imag(f),real(f))/step),
levels)));
% h = h0.*exp(complex(0,step*mod(round(phi/step), levels)));
% continuous
% h=h0.*exp(complex(0,atan2(imag(f),real(f))));

% initial reconstruction

H =  ft(h)/n; % the normalisation is to ensure transform is
isometric
              % note that this will require multiplication by n
using
              % the inverse transform h=n*ift(U)


it = 0;
flip = 1;

while flip && (it<it_max) % iterate until no pixel flips are
accepted

    it = it+1;
    flip = 0;
    flips = 0;

    % initialise / update smoothed target
    pf=n*ift(P.*F);

    % initialise / refresh smoothed hologram and alpha
    ph=n*ift(P.*H);

    alpha = sum(sum(h.*conj(pf)));
    pff   = sum(sum(pf.*conj(f))); % constant
%     alpha = sum(sum(ph.*conj(f))) / pff;
%     alpha = sum(sum(ph.*conj(f)));
%     alpha = sum(sum(h.*conj(pf)));

    % optional report of  energy

    beta = sum(sum(ph.*conj(h)));
    energy = real(beta-alpha*conj(alpha))

    % report of total normalized error in z=Z (target) plane
```

```
    sH        = sum(sum(H.*conj(H)));
    sF        = sum(sum(F.*conj(F)));
    a_en      = sum(sum(H.*conj(F)))/sF;
    errH_norm(it) = 1 - abs(a_en)^2 * sF/sH;
    diff_eff(it)  = 1 - errH_norm(it);

    alphaH  = sum(sum(P.*(H.*conj(F))))/sum(sum(P.*(F.*conj(F))));
    fprintf('\n Iter # %d,  errH_norm= %g,  diff_eff=
%g',it,errH_norm(it),diff_eff(it));
    fprintf('\n          pff = %g,  alpha = %g, alphaH =%g', pff,
alpha, alphaH);

    % optional display of evolution

    % hologram

    %figure(4);

    %D = atan2(imag(h),real(h));
    %D = D-min(min(D));
    %D = D/(max(max(D)));

    %D = 255*D;
    %image(D);
    %colormap(gray(256));
    %axis square;
    %axis xy;

    %pause

    % reconstruction

    %figure(5);
    %D = abs(H);
    %D = D/(max(max(D)));
    %D = 255*D;
    %image(D);
    %colormap(gray(256));
    %axis square;
    %axis xy;

    %pause

    % quantisation in object domain

    % visit every pixel

    for i = 1:n
        for j = 1:n

            hold =h (i,j);
            h0ij = h0(i,j);  % field amplitude at pixel = (i,j)

            % quantise

            c=ph(i,j)-alpha*pf(i,j);
```

```
            d=p00-conj(pf(i,j))*pf(i,j);
            b=c-d*hold;

            % binary

            hnew=sign(real(-b));

            % quaternary

            %br=real(-b);
            %bi=imag(-b);
            %if abs(br)>abs(bi)
                %hnew = complex(sign(br),0);
            %else
                %hnew = complex(0,sign(bi));
            %end

            % n-ary

%             hnew = h0ij*exp(complex(0,step*mod(round(atan2(imag(-
b),real(-b))/step), levels)));

            % continous

%             hnew=h0ij*exp(complex(0,atan2(imag(-b),real(-b))));

            % update state information on a pixel flip

            dh = hnew - hold;
            b2     = mod(round(atan2(imag(-b),real(-
b))/step),levels);
            b1     =
mod(round(atan2(imag(hold),real(hold))/step),levels);
            b3     =
mod(round(atan2(imag(hnew),real(hnew))/step),levels);

            if i==nc && j==nc
              fprintf('\n          At i=j=nc: abs(dh)/abs(hold)= %g,
tol= %g, abs(hnew)=%g',abs(dh)/abs(hold),tol,h0ij);
              fprintf('\n          At i=j=nc: c=%g, d=%g, b=%g,
arg(-b)= %d, arg(hold)=%d, arg(hnew)=%d',c,d,b,b2,b1,b3);
            end


            if (abs(dh) > tol*h0ij)

                flip = 1;
                flips = flips+1;

                % update hologram
                h(i,j) = hnew;

                % update alpha
                alpha = alpha + dh*conj(pf(i,j));
```

```
                % update ph

                u=mod(x+i,n)+1;
                v=mod(y+j,n)+1;

                ph(u,v)=ph(u,v)+dh*p;

            end
        end
    end

    iterations = it;
    number_of_flips = flips;
    fprintf('\n          Flips = %d',number_of_flips);

    % phase optimisation in Fourier domain

    % update reconstruction

    H =  ft(h)/n;

    % update target reconstruction phase

    F = abs(F).*(exp(complex(0,atan2(imag(H),real(H)))));

    % update target object

    f = n*ift(F);

end

% computation_time = cputime-t1

%-----------------------------------------------------------------------
% reporting module
%-----------------------------------------------------------------------
% signal to noise ratio & diffraction efficiency

total = real(sum(sum(H.*conj(H)))); % this is a known number and
really does not need to be calculated

alpha = sum(sum(H.*conj(F)))/(sum(sum(F.*conj(F))));
G = H-alpha*F;
error=G.*conj(G);

support=find(F);
noise=sum(error(support));

signal= real(alpha*conj(alpha)*sum(sum(F.*conj(F))));


snr = signal/noise

difraction = signal/total
```

```
% display hologram

figure(4);
D = h0.*atan2(imag(h),real(h));
% D = 255*D/(max(max(D)));
% D = D-min(min(D));
imagesc(D);
% colormap(gray(256));
axis square;
axis xy;

% display reconstruction

figure(5);
D = abs(H);
% D = 255*D/(max(max(D)));
imagesc(D);
% colormap(gray(256));
axis square;
axis xy;

% figure(6);
% plot(errH_norm(1:it));
% figure(7);
% plot(diff_eff(1:it));

return
```

Appendix B

FABRICATION DATA

Fourier Transform Lens

ELEMENT NUMBER	RADIUS OF CURVATURE FRONT	RADIUS OF CURVATURE BACK	THICKNESS	APERTURE DIAMETER FRONT	APERTURE DIAMETER BACK	GLASS
OBJECT	INF		INFINITY			
			APERTURE STOP	8.4932		
			7.7595			
1	-15.4492 CC	38.7187 CC	4.0000	10.8806	13.5628	FK5 Schott
			1.0000			
2	-209.6547 CC	-24.2127 CX	5.0000	13.8478	15.9495	SF4 Schott
			0.1000			
3	33.4579 CX	-188.4024 CX	6.0000	17.1217	17.2167	SF4 Schott
			17.4280			
4	188.4024 CX	-33.4579 CX	6.0000	16.6266	16.3539	SF4 Schott
			0.1000			
5	24.2127 CX	209.6547 CC	5.0000	15.1687	12.9474	SF4 Schott
			1.0000			
6	-38.7187 CC	15.4492 CC	4.0000	12.5486	9.8429	FK5 Schott
	IMAGE DISTANCE =		6.9225			
IMAGE	INF CX			7.5080		

NOTES - Positive radius indicates the center of curvature is to the right
Negative radius indicates the center of curvature is to the left

- Dimensions are given in millimeters

- Thickness is axial distance to next surface

- Image diameter shown above is a paraxial value,
it is not a ray traced value

- Other glass suppliers can be used if their materials are
functionally equivalent to the extent needed by the design;
contact the designer for approval of substitutions.

REFERENCE WAVELENGTH = 632.8 NM

INFINITE CONJUGATES

EFL	=	21.2330
BFL	=	6.4431
FFL	=	1.3164
F/NO	=	2.5000
IMAGE DIST	=	6.9225
OAL	=	57.3875
PARAXIAL		
IMAGE HT	=	3.7439
SEMI-FIELD		
ANGLE	=	10.0000
ENTR PUPIL		
DIAMETER	=	8.4932
DISTANCE	=	0.0000
EXIT PUPIL		
DIAMETER	=	136.9883
DISTANCE	=	348.9137

NOTES - FFL is measured from the first surface
- BFL is measured from the last surface
