Free space optical interconnections
for an incoherent correlator

A thesis by

Chun Ho Peter Poon

submitted to the University of London for the degree of Ph.D.

Department of Electronic and Electrical Engineering
University College London

March 1995
ABSTRACT

This thesis describes the design of free-space optical interconnections between two arrays within commercial constraints for use as an incoherent correlator in image recognition or compression applications. Critical reviews are given of computer generated holograms and of refractive microlenses. Photoresist microlenses were chosen due to their ease of fabrication, efficient operation under incoherent illumination and the compactness of the system which results. Novel long focal length (up to 4.7 mm) microlenses with diameters of 120 μm were realised by immersing the microlenses in fluids of appropriate refractive index. They had negligible astigmatic aberrations for off-axis angles of illumination. Two novel configurations are proposed for incoherent correlators based on the image multiplexer and shared microlens interconnect systems. The shared microlens total interconnect was used because it was shorter. Theoretical analysis of a shared microlens system, a shared pinhole system and a shadow casting system showed that the first gave the best combination of high efficiency with good resolution and low crosstalk in the correlation plane. Measurements of the angular responses of two diffusers for generating illumination gave a narrow but efficient response for the microlens diffuser and a wider spread for a bulk plastic diffuser which was then selected. An incoherent correlator was constructed successfully using two high contrast image transparencies giving correlation peaks as predicted and demonstrating translation invariance. The feasibility of using ferroelectric liquid crystal spatial light modulators was assessed. A theoretical computer simulation of the correlation process showed that the poor contrast ratio of the SLM (10:1) seriously degraded the output correlation contrast (2:1). Indeed in experiments using one SLM and a high contrast transparency it was impossible to discern any correlation peaks. We conclude that a higher contrast SLM, which may be in the form of an optically addressed SLM, is required for successful practical operation.
## CONTENTS

Title page 1
System diagram 2
Abstract 3
Contents 4
Acknowledgements 12
Dedication 14

### 1. INTRODUCTION

1.1 Aims 15
1.2 Motivation 16
1.3 Choice of interconnection
   1.3.1 Technology of the interconnection: Optical vs. Electronic 17
   1.3.2 Topology for the optical interconnection: point-to-point vs. optical bus network 19
   1.3.3 Choice of interconnection components: fixed vs. adaptive interconnects 19
   1.3.4 Illumination: Coherent vs. Incoherent optical interconnects 21
1.4 Applications and marketing 22
1.5 Thesis layout 22

### 2. COMPUTER GENERATED HOLOGRAMS REVIEW

2.1 Introduction
   2.1.1 Layout of the chapter 25
   2.1.2 Introduction to computer generated holograms 26
2.2 Dammann grating

2.2.1 Introduction to Fourier gratings 30
2.2.2 Separable Dammann gratings 30
2.2.3 Non-separable gratings 34
2.2.4 Optimisation algorithms 34
2.2.5 Multiple and continuous phase profile gratings 35
2.2.6 Fabrication 36
2.2.7 Conclusions for Dammann gratings 37

2.3 Fresnel zone plates

2.3.1 Introduction 40
2.3.2 Theory of Fresnel zone plates 40
2.3.3 Characterisation of Fresnel lens 43
2.3.4 Generation of Fresnel zone plate arrays 46
2.3.5 Hologram facets 47
2.3.6 Other types of FZP arrays 48
2.3.7 Conclusions for Fresnel zone plates 49

2.4 CGH based interconnect systems

2.4.1 Introduction 49
2.4.2 Space-invariant systems 50
2.4.3 Space-variant systems 51
2.4.4 Conclusions 52

3. REFRACTIVE MICROLENSES REVIEW

3.1 Introduction

3.1.1 Historical developments of microlenses 54
3.1.2 Layout of this chapter 55
3.2 Graded-index profile microlenses
   3.2.1 GRIN rod microlenses 56
   3.2.2 GRIN planar microlens 58

3.3 Surface relief microlenses
   3.3.1 Photoresist reflow microlenses 60
   3.3.2 PMMA microlenses 64
   3.3.3 Semiconductors microlenses using mass transport technique 65
   3.3.4 Photosensitive glass microlenses 66
   3.3.5 Other polymer and plastic microlenses 66
   3.3.6 Complex surface relief microlenses 67
   3.3.7 Other surface-relief microlenses 68

3.4 Fibre microlenses 69
3.5 Comparison of the different types of microlens 69
3.6 Imaging systems
   3.6.1 Introduction 73
   3.6.2 Image multiplexer 73
   3.6.3 Shared microlens system 75
   3.6.4 Sakano's Theory 76
   3.6.5 Hybrid system 77
   3.6.6 Lenslet-pair imaging system 79
   3.6.7 Comparison of the systems 79

3.7 Conclusions 80

4. CORRELATOR SYSTEM: PRELIMINARY SYSTEM DESIGN

4.1 Introduction 82
4.2 Review of correlators
   4.2.1 Definition 83
4.2.2 Fourier-domain correlators 84
4.4.3 Real-space domain (Object-space) correlators 89

4.3 Correlation operation of the optical interconnect systems
   4.3.1 Introduction 92
   4.3.2 Correlation operation of an image multiplexer 93
   4.3.3 Correlation operation of the shared microlens system 94

4.4 Imaging systems comparison - Extension of Sakano's Theory 95
4.5 Optical interconnection elements comparison
   4.5.1 Introduction 98
   4.5.2 Optical interconnection elements summary 98
   4.5.3 Comparison between the CGHs and microlenses 100
   4.5.4 Conclusions 102

5. CORRELATOR SYSTEM: THEORETICAL ANALYSIS

5.1 Introduction 104
5.2 Diffraction theory
   5.2.1 Justification for the use of Rayleigh-Sommerfeld diffraction theory 105
   5.2.2 Diffraction at an aperture 105

5.3 Diffraction theory for the shared microlens system
   5.3.1 Introduction 107
   5.3.2 Derivation of the resolution for a microlens 107
   5.3.3 Resolution of the shared microlens correlator under LED illumination 109

5.4 Diffraction theory for the pinhole system
   5.4.1 Introduction 110
5.4.2 Diffraction theory for the pinhole system 111
5.4.3 Variation of intensity of diffraction with distance away from the pinhole 112
5.4.4 Resolution of the pinhole system 114
5.4.5 Resolution of the simple shadow casting system 116

5.5 Comparison and discussions 117
5.6 Off-axis diffraction
5.6.1 Introduction 120
5.6.2 Shared microlens system 121
5.6.3 Pinhole system 122
5.5.4 Optimisation for off-axis applications 123

5.7 Conclusions 125

6. INCOHERENT ILLUMINATION

6.1 Introduction 127
6.2 Illumination requirements for the shared microlens correlator 128
6.3 Source for incoherent illumination
6.3.1 Definition of photometry quantities 129
6.3.2 Incoherent illumination source 129
6.3.3 Conclusions 132

6.4 Diffusers
6.4.1 Introduction 132
6.4.2 Operation of the diffuser 133
6.4.3 Angular distribution of the scattered light 133
6.4.4 Traditional diffusers 134
6.4.5 Holographic diffusers 135
6.4.6 Refractive diffusers 136
6.4.7 Other diffusers 137
6.4.8 Summary of the different diffusers 137

6.5 Effect of illumination on other system components
6.5.1 Introduction 138
6.5.2. Effect of illumination on contrast ratio of the SLM 139
6.5.3 Sensitivity of the CCD camera 140

6.6 Uniformity comparison between plastic and microlens array diffusers
6.6.1 Introduction 140
6.6.2 Choice of diffusers 141
6.6.3 Properties of the light emitting diode used in the experiment 141
6.6.4 Experimental apparatus 142
6.6.5 Spatial uniformity measurements, results and discussion 143
6.6.6 Angular spread measurements and discussion 147

6.7 Design of a microlens array diffuser with variable diameters 149

7. MICROLENSES: DESIGN, FABRICATION AND PERFORMANCES

7.1 Introduction 152
7.2 Microlens specifications 153
7.3 Fabrication of photoresist reflow microlens arrays
7.3.1 Description of the Fabrication process 155
7.3.2 Factors which control the quality of the photoresist microlenses 156

7.4 Imaging quality of the microlenses 162
7.5 Interferometric measurements of the microlenses
7.5.1 Mach-Zehnder interferometer 165
7.5.2 On-axis aberrations measurements 166
7.5.3 Off-axis aberration measurements 168
7.6 Long focal length microlenses

7.6.1 Introduction 173
7.6.2 Design and fabrication 174
7.6.3 Focal length measurements 176
7.6.4 Interferometric measurements 178
7.6.5 Imaging experiment using an apertured long focal length microlens 182

7.7 Conclusions 183

8. CORRELATOR SYSTEM: OPTIMUM SYSTEM DESIGN, ANALYSIS, SIMULATIONS AND EXPERIMENTS

8.1 Introduction 185

8.2 Computer simulations of the correlator performance

8.2.1 Introduction 186
8.2.2 Simulation techniques 186
8.2.3 Simulation results of binary patterns 188
8.2.4 Simulation of bipolar patterns 191
8.2.5 Simulation of randomised patterns 194

8.3 Ferroelectric liquid crystal spatial light modulator

8.3.1 Introduction 195
8.3.2 Contrast ratio 195
8.3.3 Programming of the SLM 205

8.4 Correlation system results without SLM

8.4.1 Introduction 205
8.4.2 Correlation with transparencies under LED illumination 206
8.4.3 Correlation with resolution targets under LED illumination 208
8.5 System Experiment with the spatial light modulator
   8.5.1 Experiment setup 210
   8.5.2 Result of the experiment 212

9. CONCLUSIONS AND FUTURE WORK

   9.1 Conclusions 213
   9.2 Future research 216

References 218
Appendix I - Re-expression of Sakano’s theory 239
Appendix II - Aspects of diffraction theory (not explained in chapter 5) 241
Appendix III - Seidel coefficients for astigmatism and spherical aberrations 246
Appendix IV - Simulation program 249
Appendix V - SLM Control programs 250
List of publications and patent 259
ACKNOWLEDGEMENTS

I am most grateful to my university and PhD supervisor, David R. Selviah for all his guidance, encouragement, enthusiasm, and for spending a great deal of time in a countless number of stimulating discussions; without him my task would have been impossible. Almost all my new ideas were formulated during the joint discussions.

I would also like to thank my industrial supervisor Dr. Mike G. Robinson for all his advice and for help, especially during the time when I was seconded to SHARP; and for all the time spent in teaching me invaluable practical skills.

I wish to thank SHARP Laboratories of Europe Ltd. for financial support for my PhD studies and Dr. Paul May for his help in many logistic and administration matters when I stayed at SHARP. I would like to thank Dr. Kataoka for his interest in this project. My gratitude also goes to Dr. Craig Tombling for his time and patience in helping me in the clean room at SHARP and for fabricating microlens samples. I would like to acknowledge other researchers at Sharp for their help and the loan of equipment, in particular, Graham Woodgate, Dr. Kathryn Walsh, Dr. Gillian Davis, Dr. Robert Brown; and Graham White for fabricating microlenses samples. I thank Dr. Clive Bradley for permission to spend about one year in total working at the Sharp labs.

In addition, I wish to acknowledge and thank Prof. Mike Hutley and Dan Daly of National Physical Laboratory for their generous help in the many microlens measurements and for useful discussions about the fabrication of photoresist microlenses. I thank Mike Hutley for lending permission to use the facilities of the NPL to test my microlenses. I also acknowledge Lau Hon Wo of DeMontford University for useful discussions on microlenses.

I would like to show my gratitude to Dr. Sally E. Day for useful discussions on liquid crystal SLMs and for commenting on the SLM section of my thesis. I would like to thank Chang Chi Ching for helping with my experiments and Annie Marinopoulou for helping me to obtain the SEM photos. I thank Lawrence Commander for the useful conversations on liquid crystals and Dominic Godwin for helping me with the design of Fresnel zone plates and for the discussions in computer generated holograms. My thanks
also go to Dr. ZQ Mao for his help in controlling the SLM. I would like to show my thanks to Chris Carey who has offered assistance on many occasions and to Dr. S.H. Song who helped in my experiment at UCL. I also wish to thank Madam Tao Shiquan for lending me the equipment and for some very useful discussions. My thanks also go to Tony Rivers who has given some useful suggestions on processing of samples.

I wish to show my gratitude to Prof. Gareth Parry at Oxford for this kindness in letting me use their facilities during my secondment at Sharp. I would like to thank Prof. John Midwinter for his assistance on many matters and for offering other insights to my work. My thanks also go to Dr. Paul Radmore who has helped me with the theoretical analysis in this project. My acknowledgement also goes to Mat Harris for his help on the Sun SPARC network and also for demonstrating the image processing software written by Phoenix Tong. I would also like to thank Jim Rice & Leon Chernac whom I have bothered them frequently with intriguing system problems.

Finally, I would like to thank several of my fellow PhD colleagues in the Digital Optics Group at UCL especially Dr. Evi Zouganelli, Dr. Abid Khan, Dr. David Atkinson, Dr. Marco Ghisoni, Dr. Paul Stavrinou, Chee Fai Tang, Farah Mansoor and Petra Guy. I would also like to acknowledge my personal friends: Felix Ng for providing me with accommodation at Oxford and for stimulating discussions, Chris Ng for technical help at the networks, Chia Choon Chia for support during my PhD and Duncan Watts for also putting me up at Oxford.
To my Mum and Dad
1.1 Aims

The aim of this project is to design a free-space interconnection scheme or system which enables a large number of pixels in an array on a plane to be totally interconnected to a similar large array of pixels on a second plane. The phase 'total interconnection' means that each pixel on the first plane is connected optically to each and every pixel on the second plane. Such an interconnect can be used as part of a correlator or as part of a parallel processing optical neural network which requires massive connections between single processing units. In addition, the interconnection system must be designed with commercial constraints of which the following are the most important.

(a) Ultimate system occupy no more space than a few cm$^3$.
(b) System components must be readily available, cheap or capable of in-house manufacture in the sponsor company
(c) System must use LC-SLM or displays which could be made in-house due to the sponsor advanced expertise with liquid crystals and dominance in the market place.
Power efficiency must be maximised.

Number of pixels capable of being totally interconnected ultimately must be large so high resolution images can be used.

1.2 Motivation

Image and pattern recognition can be implemented using many different technologies such as software, electronic and optical hardware etc. Since the process mainly deals with two dimensional images with high resolution and, hence, number of pixels, it is more advantageous to implement pattern recognition using optical technology which is inherently parallel and fast. It also allows direct processing of optical images without conversion into electronics representations thereby removing the electronics bottleneck due to lower speed, bandwidth etc. However, the technology of all optical computing in terms of devices and system architectures although very attractive, requires more research to be carried out. Therefore, in this research project, which aimed at realising commercial product on a short time scale, we utilise all three technologies: software for control and non-linear processing; electronics for interfacing; and optics for the dense, parallel interconnections, to construct an incoherent correlator for the purpose of pattern recognition. As a result of this, the best aspects of each technology are exploited to implement a hybrid optoelectronic system.

In this chapter, we outline possible approaches for realising optical interconnects and explain why we have dismissed some from the outset for use in our system. The remaining approaches are reviewed in the following chapters and our choice is refined with the help of theoretical analysis and experimentation in the rest of this thesis.

1.3 Choice of interconnections

The design and commercial constraints and the system's applications influence the choice of the interconnection in terms of technology, topology and the selection of components. In this section, we will investigate how this choice of interconnection can be made.
1.3.1 Technology for the interconnection: Optical vs. Electronic

The two most popular means of communication for large bandwidth information are light and electricity. There are other types of communication using, for example, radio and microwave frequencies but we will confine ourselves to the optics region of the electromagnetic spectrum. The type of connection is determined largely by the connection distance between the two points, which classifies communication links and networks into six hierarchy levels [Too91]: (the distance quoted is only a rough approximation)

(a) Long haul systems-to-systems (> 1 km)
(b) Direct line of sight communication links (> 1 m)
(c) Local area networks between systems (1 m - 1 km)
(d) board-to-board (1cm - 10 cm)
(e) chip-to-chip (1 mm - 10 cm)
(f) intra-chip (1 μm - 1 cm)
(g) gate-to-gate (0.1 μm - 1 μm)

The optimum choice for long distance communications is optics because of cost and bandwidth, hence the use of optical fibres in telecommunications. Direct line of sight communications had been dominated for the larger distances by microwave links although promising research is being carried out (UCL & BT) using optical infra-red communication links. For shorter distances, infra-red links as well as ultrasonic links are common for remote controls for TV, car alarm etc. Although wires have been the conventional means for local area networks, the use of optics in local area network is becoming more prominent because of the requirements for an ever increasing information capacity. Gate-to-gate interconnections for usual applications are best implemented by electric wires (strip line). However, there is no outright winner on which type of interconnection is most suitable for board-to-board [Bri89, Str93, Jia93, Sak93, Dho92], chip-to-chip [Sch92, Cin93, Fel87a, Ack93, Fel93, Cra92] and intra-chip levels. At the intra-chip level, the trend of decreasing feature size, growing chip size, is posing several limitations on electronics. First, an increase in chip size means that components within a chip are further away from each other. This gives rise to a clock skew effect which cause delay in sending clock signals within a
chip and which becomes serious at high clock rates. Secondly, smaller feature sizes allow more components to be integrated and consequently more interconnections are required to communicate between these components. This will result in problems such as large power dissipation, cross-coupling between wires and capacitive loading in addition to the problems of how to fabricate multiple levels of metallisation for crossing connections. Similarly, at the chip-to-chip level, increasing parallelism between processors means that more and more interconnections are required. The fan-in and fanout of a chip are severely restricted by the number of package pins. As in the case of intra-chip connections, more wires can lead to more interference between the channels and shielding may be required. Furthermore, both GaAs and Si suffer significant speed degradations in transporting signals off the chips because of the high capacitive load. All these problems are also present at the next level: board-to-board interconnections. This is the next area most likely to be surplanted by optics due to the increased industrial research into optical solution.

Optical interconnections can be divided mainly into two groups: waveguide and free space interconnections. Waveguide interconnections include optical fibres and planar integrated optics. Optical fibres and waveguides have the intrinsic advantages of optics in terms of bandwidth. However, they suffer the same drawbacks of electronics that each channel has to be physically separated at the connection stage to avoid crosstalk. This poses a physical constraint on the size of the system. In addition, dispersion and clock skew can occur. Free-space optics [Wu87, Fel87b, Fel89] which includes planar optics (the use of a glass block as the propagation medium), on the other hand, can alleviate all these problems. Like the optical waveguides, they have large bandwidths and are free from severe clock-skew and capacitive loading effects. Unlike them, individual light beams are not confined to a physical guidance medium. They can cross one another without any interference and any cross-talk in the system may only be found at the transmitter and receiver ends. Most importantly, free-space interconnects utilise the third dimension, the volume surrounding the planar electronic geometry and, hence, a large number of interconnections are possible. The only limitation comes from the optoelectronic interface. Consequently, we make use of free-space optics as the means of providing interconnections in our system, which is at the board-to-board level.
1.3.2 Topology of the optical interconnection: point-to-point vs. optical bus network

The two main areas of applications for optical interconnects are communications and optical signal processing. The latter includes applications such as optical correlation using compression and coding, feature enhancement, computing and neural networks. In many cases, the same topology for the interconnect can be used in either application, although the design is usually influenced by the type of interconnect and its functionality. The architecture of the interconnections can be classified into point-to-point (one to one), broadcast or one to all fanout (one to many), fan-in (many to one) and optical bus type of interconnects (many to many) [Kra92]. The point-to-point interconnect consists of connections between one input node and one output node and is usually implemented as an individual stage in a multi-stage system except for the direct line of sight communications. There are many types of point-to-point interconnection from an array of points to an array of points, such as perfect shuffle, crossover, cyclic shift etc. and a combination of these different routing stages give a fully interconnected network [Sch92, Kaw91, Hut92, Jah90,91, Noguc92, Too91]. The optical bus type network [Kra92] has a large number of connections to the output for each input node. Therefore, it is actually a combination of fan-ins and fanouts, and may give rise to a fully interconnected system in one stage.

Multi-stage interconnects are usually favoured for photonic switching purposes in optical communications for their simplicity in the implementation. On the other hand, the compactness of an optical bus system is usually preferred in optical signal processing applications. For the obvious reason of complexity, fibre optics is used in point-to-point networks and free space optics is most suitable for the optical bus system. In our system, we are therefore concentrating on the free-space optical bus interconnection configuration and have decided to orient it to addressing of the problem of image correlation.

1.3.3 Choice of interconnection components: fixed vs. adaptive interconnects

In general, an optical bus type interconnect comprises $M \times L$ connections linking one of the $N \times N$ elements to others (where $N \geq M,L$). One special case is when each element is fully interconnected to all the others, i.e. each has $N^2$ connections. This type of network is called a total interconnect. It is not usually favoured in communication systems because of the
large amount of switching that has to be executed in a single stage. Correlators and neural networks, on the other hand, often require such dense interconnections for robustness although sparse interconnects are sometimes more useful. Nevertheless, a total interconnect is easy to implement in practice and some interconnections can be redundant if required, rather than designing one for a special case.

Interconnections can be fixed or reconfigurable. Reconfigurable interconnects are more useful both for switching in communications and for updating weights which give rise to learning in neural networks. For correlators, the simplest design only requires a fixed interconnect although an intelligent system may need an adaptive one. There are many choices for optical interconnection elements. We consider four major elements and compare their merits in terms of performance and commercial constraints of size, cost and ease of manufacture, as follows:

(a) Thin optical holograms
Holograms can be optically recorded on photographic films. They are cheap and can be easily copied and manufactured. However, thin holograms have poor efficiencies and low signal-to-noise ratios.

(b) Computer generated holograms
Computer generated holograms (CGHs) have design flexibilities and the manufacturing process can be automated. They are fixed holograms and can have high efficiencies depending on the design.

(c) Micro-optical elements
Micro-optical elements such as microlenses have similar benefits to CGHs such as automation and ease of manufacture. In addition, they have high collection efficiencies. Like the CGHs, they are also fixed interconnects.

(d) Volume holograms
Volume holograms have high efficiencies compared to the thin holograms. Materials such as dichromated gelatin and photopolymers are relatively cheap but require processing. Photorefractive materials are adaptive and reusable. However, they are not suitable for this
system because mass manufacture is more costly and time consuming as they are not simple surface relief structures as (b) and (c) which can be copied cheaply using a process akin to the one used to make compact audio discs.

Table 1.1 shows a summary of the different interconnection elements. We can conclude that, in terms of fixed interconnects and under the commercial constraints, CGHs and microlens arrays are the most suitable choices for the fixed optical interconnect element. Further survey and comparison of their properties will be described in the next two chapters. Although they are fixed interconnects, adaption of the system can be implemented separately by introducing another component to the system such as a liquid crystal SLM.

<table>
<thead>
<tr>
<th></th>
<th>CGHs</th>
<th>Microlens arrays</th>
<th>Thin holograms</th>
<th>Volume holograms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantages</strong></td>
<td>- Automation</td>
<td></td>
<td>- Cheap</td>
<td>- very high efficiency</td>
</tr>
<tr>
<td></td>
<td>- Easy processing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- ease of manufacture</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(cheap master or cheap copies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>since surface relief)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Disadvantages</strong></td>
<td>- Fixed</td>
<td>- low efficiency</td>
<td>- need complicated processing</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- low SNR</td>
<td>- may be complicate and costly to manufacture</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.1 Comparison of CGHs, microlenses, thin and thick holograms

1.3.4 Illumination: Coherent vs. Incoherent optical interconnects

Optical interconnects can use coherent light in a laser-based system or incoherent light if using an LED or a white light source. Coherent optical systems provide higher information-carrying capacity but are prone to noise such as speckle which can seriously degrade the image quality. Incoherent systems, on the other hand, introduce redundancy and are more robust to noise. In order to achieve the high information capacity with coherent light, very accurate alignment of the components is required and the use of incoherent light can avoid
this stringent requirement. In addition, LEDs cost less and generally have longer lifetimes than a laser diodes. Therefore we decided to use incoherent illumination.

1.4 Applications and marketing

A further unstated aim was ultimately to make a commercially viable product which could be supplied to a large market application. Ideally the system should use components already familiar to the sponsor or components which the sponsor wished to become more familiar. Similarly, the system ideally should be aimed at a market already of interest to the sponsor into which they were already selling products or into which they had decided to move. The main application considered was for use in direct input of characters on a page into a computer, or photocopier or fax with automatic character recognition optically by scanning the small system across the page. Alternatively, an array of such small systems could cover the page. This is the reason why the simulations and experiments used characters as inputs. Our system is translation invariant so precise alignment with the text is not required. The page scatters light and therefore the recognition system should use incoherent light. One or two liquid crystal displays should be incorporated because of our sponsor's expertise and business in this area. The displays have variations in thickness unlike the liquid crystal SLM. In addition, displays have a much narrower viewing angle than the SLM. The research in this thesis shows that a large angular range of the SLM is required.

A further application considered use for data reduction or image compression in which either a whole image or an array of sub-divided parts of an image could be recognised. This could be achieved by an array of pattern recognition systems. Indeed a recent image compression standard required the image to be divided into an array of $8 \times 8$ pixel sub-array. Each $8 \times 8$ sub-array is correlated with a pattern. This could be achieved using an array of correlator system each using an $8 \times 8$ pixel array.

1.5 Thesis layout

The approaches taken to achieve the aim are outlined as follows:
Introduction

(a) to investigate and review the fixed optical interconnects which includes computer generated holograms and microlenses for use in the incoherent correlator.

(b) to review optical interconnect systems and optical correlators with view to selecting the optimum configuration

(c) to design and analyse the incoherent correlator theoretically.

(d) to fabricate and test photoresist microlenses and long focal length microlenses for interconnections.

(e) to investigate and characterise the illumination and other parameters of the system components.

(f) to demonstrate a compact interconnect for the correlator and compare the performance with simulation.

(g) to outline what advances in devices are necessary to realise a commercially system.

In effect this thesis is a feasibility study report in which a major part is the critical review and comparison of optical interconnections, routing and fanout elements which account for the two chapters devoted to this. Computer generated holograms are reviewed in chapter 2 and refractive microlenses are reviewed in chapter 3. Similarly, the critical review of many types of suitable optical interconnection systems, both coherent and incoherent is a key portion and the associated systems are described at the end of the corresponding chapter. We select the most suitable optical interconnection elements and the optimum system design in chapter 4. The theoretical performance of the shared microlens correlator is analysed in chapter 5 and comparing with pinhole and shadow casting systems. The system is then divided into four main areas, each has been designed for optimum system performances with related parameters. The method of realising incoherent illumination is described in chapter 6 and is an integral part of the design of our system as it dramatically influences its performance. In chapter 7, we report on the design, fabrication and optical quality measurements of photoresist microlens arrays and compare them to an array of graded-index microlens which is commercially available. In addition, the novel long focal length microlenses are described in the same chapter. Another crucial factor to the signal to noise ratio is the contrast ratio of the SLM and this is investigated in chapter 8 and the practical contrast value is used in our simulation of our system. This chapter also describes
an successful experimental demonstration of an incoherent shadow casting system without microlenses. Finally, the conclusions and future work are given in chapter 9.
CHAPTER TWO

COMPUTER GENERATED HOLOGRAMS REVIEW

- Introduction
- Dammann grating
- Fresnel zone plates
- CGH based interconnect systems

2.1 Introduction

2.1.1 Layout of the chapter

In this chapter we review two types of computer generated holograms for implementing an optical interconnect. First, holograms and general computer generated holograms are briefly described in section 2.1.2. Then the first of the two CGH interconnection elements: - Dammann-like gratings which are used for array generation, are reviewed in section 2.2. In section 2.3, we describe the Fresnel zone plates. Finally, some holographic interconnect systems which uses the Dammann grating or FZP are reviewed (section 2.4). The conclusions are given after each section and summarised in tables 2.1 and 2.2. The comparison of the elements will be made together with the refractive microlenses in chapter 4.
2.1.2 Introduction to computer generated holograms

Holograms are formed when a wave scattered by an object interferes with a coherent reference wave and the subsequent interference pattern is recorded. By illuminating the hologram with either wave, the other one is reconstructed by the diffraction of the wavefront at the hologram. The first holograms were formed by Gabor [Gab48] in research on reducing aberrations of electron microscopes. Holograms can be classified into three categories:–

(a) Thin optically recorded holograms, in media such as photographic films (silver halide emulsions)
(b) Volume holograms, also optically recorded, e.g. dichromated gelatin and photorefractive crystals
(c) Computer generated holograms

In this project, we are only concerned about computer generated holograms although both thin and volume optical holograms can also serve the purpose.

Computer generated holograms [Hari84],[Cau85],[Tric87] (CGHs) are interference patterns calculated by a digital computer that simulate the wavefront generation of optically recorded holograms. Given an object description or the complex amplitude of the object wavefront, the computer can be used to calculate the diffraction pattern. CGHs have advantages over conventional optical holograms in terms of more flexibility in design, lower cost and easily reproducible. In addition, CGHs can perform wavefront transformations that are beyond the capabilities of optical holograms. Furthermore, some CGHs can achieve higher efficiencies than thin optical holograms and unwanted coupling and noise can be removed in the design.

The process of generating CGHs is as follows: (Fig. 2.1)

(i) The object, whether it is real or idealised, is digitised by the computer. This is achieved by sampling a sufficiently large number of data points so as to preserve all the information by satisfying the sampling theorem.

(ii) Then the object wave is propagated by the computer from the object plane to the hologram plane. If the object is at a long distance away from the hologram plane, a Fourier transformation (the computer uses the algorithm "Fast Fourier Transform")
(iii) The complex amplitude of the hologram is then encoded as a real non-negative intensity function. Error diffusion [Floy76],[Bam89] is one of the many techniques which are used for binarisation.

(iv) Finally, the hologram is generated from a graphic output device linked to the computer. This may be a printer or plotter which plots a large scale version of the holographic pattern. This mask is later photographically reduced to the suitable size on transparency. An alternative method uses e-beam lithography, in which the data is directly output to the e-beam writer and a real-size mask of the hologram is produced or the hologram itself by a direct write process for achieving higher resolution. The CGHs fabricated represent the transmittance of the hologram but reflection holograms are also possible by metallising one of the surfaces.

One of the first computer generated holograms was described by Brown and Lohmann [Bro66] in 1966. The holograms made were known as "detour-phase holograms"
and were intended for Fraunhofer diffraction. As the mask's transmittance had values of 1 or 0 only, the holograms were called binary. This Fourier hologram is divided into regions of equal size or cells. Each cell has a transparent window, which corresponds to each complex Fourier coefficient. The size of the window is determined by the modulus of the coefficient, whereas its position within the cell represents the phase (Fig. 2.2). It works by the fact that a shift of the transparent area within each cell results in the light from a point source transmitted through it travelling by a longer or shorter path to the reconstructed image. This significant work stimulated much research in this field. Because of the finite number of phase levels that can be achieved in practice, the phases are quantised and quantisation errors can give rise to false image and noise. This problem can be avoided in Lee's holograms. Lee [McC089] devised a method that is based on decomposing the phase of the Fourier components into four quadrature components, which are described by real non-negative functions (Fig. 2.3a). These four functions are represented in a hologram by apertures in equally-sized sub-cells within each cell of the hologram. The amplitude of the Fourier component is determined by the summation of the transmittance of the subcells. Burckhart [Vel86] simplified this method to using three subcells only (Fig. 2.3b). Other earlier CGHs developed include Lee's interferograms, Burch holograms and kinoforms.

Nowadays there are many types of holograms for a large variety of applications. What we are mainly interested in is array generation and, therefore, we will describe the two main types of computer generated holograms in this area - Dammann gratings and Fresnel zone plate arrays - in more detail.
Fig. 2.3 Typical subcell arrangement and phasor diagram for:
(a) Lee's hologram, (b) Burckhart's hologram
2.2 Dammann grating

2.2.1 Introduction to Fourier gratings

A class of array generators used in the far-field are known as Fourier gratings. Fourier transform lenses have until recently been required in a system configuration involving a Fourier array generator. In particular, these diffractive gratings are non-absorptive, phase-only elements which have much higher efficiencies than the amplitude holograms. They can be used as beam combiners and beam splitters for uses including imaging [McC089], coherent communications [Vel86], optical interconnects [Byc90] and neural networks [Vas92]. There are many ways for computing a Fourier grating, depending on the different representations (binary or multi-level), philosophies (e.g. Dammann method) and algorithms (Fienup, Newton-Raphson etc.) [Mai90]. The most general representation of a phase grating is called a kinoform [Tru88a], which has multiple phase levels. A Kinoform can be used to generate any object, just like any computer generated hologram, but we are interested here in its function as an array generator. The phase levels of a kinoform need to be quantised and the step size of a quantised phase is $2\pi/2^M$ where $M$ is a positive integer. When $M = 1$, then we have a binary phase structure with the phases usually designated as 0 and $\pi$. This array generator is known as a Dammann grating.

In the following sections, we first describe the concept and rules for generating separable Dammann gratings (section 2.2.2). We then look at 2-D non-separable gratings in section 2.2.3. This will be followed by a discussion on optimisation (section 2.2.4) and other types of multi-phase gratings (section 2.2.5). Finally, we describe the fabrication techniques for these gratings (section 2.2.6) and summarise in section 2.2.7.

2.2.2 Separable Dammann gratings

In this section, we describe the mathematical representation and design of Dammann gratings in their simplest form. We also mention the figures of merit for describing these gratings.

The idea of using a computer optimised binary-phase grating as a beam splitter came from Dammann et. al. [Dam71],[Dam77]. Consider the system involving two Fourier
Computer Generated Holograms Review

lenses shown in fig. 2.4, The object of amplitude \( a(x,y) \) is Fourier transformed by lens \( L_1 \) to \( A(\xi,\eta) \) and is then multiplied by the transmittance of the grating \( t(\xi,\eta) \). The product is then Fourier transformed again by lens \( L_2 \) to give the output \( b(x,y) \) at the image plane. The Fourier transform of the product at the image plane can be regarded as the convolution of \( a(-x,-y) \) with \( T(x,y) \) where

\[
T(x,y) = \mathcal{F}\{t(\xi,\eta)\}
\]

(2.1)

\( T(x,y) \) is known as the response of the grating. The design of the Dammann grating is such that multiple images of the object form at the output plane which can be described by the amplitude function as:

\[
a(-x,-y)*T(\xi,\eta) = \sum_{n,m} a(x-nx_o, y-my_o) \]

(2.2)

where \( x_o \) and \( y_o \) are the pitches of multiple-image output. It is obvious that:

\[
T(x,y) = \sum_{n,m} \delta(x-nx_o, y-my_o)
\]

(2.3)

\[\]

If we assume that the brightness of the images are equal but the relative phases \( \phi_{n,m} \) can be arbitrary, we can express the grating response, \( T(x,y) \), in a more general form:

\[
T(x,y) = \sum_{n=-N}^{N} \sum_{m=-M}^{M} e^{-j\phi_{n,m}} \delta(x-nx_o, y-my_o)
\]

(2.4)

which is valid in the area of the central block of \((2N+1)(2M+1)\) diffraction orders. The grating description \( t(\xi,\eta) \) can be calculated by the inverse transform of the response \( T(x,y) \). However, this will result in a complicated multiple-level grating. Therefore, we have to look for specific binary functions which satisfy equation (2.4).
First, we assume that the description of the grating can be further decomposed into two 1-D gratings so that less CPU time needed for 1-D optimisation, in the following way:

$$\psi(\xi, \eta) = t_1(\xi) t_2(\eta)$$  \hfill (2.5)

This 2-D phase structure is known as the separable Dammann grating.

The one-dimensional gratings $t_1$ and $t_2$ are binary functions with values -1 and +1 only and have the form shown in fig. 2.5a. The grating period is normalised to 1. Within the period, there are many zeros and different responses result from a period with either an even or an odd number of zeros. In addition, symmetry around $x = 0$ is also a design parameter. For simplicity, we assume that the function is symmetric here. A typical 2-D separable Dammann grating period is shown in Fig. 2.5b.

![Diagram](image)

Fig. 2.5 Structure representation of a single period of:
(a) an 1-D, symmetrical binary Dammann grating
(b) a 2-D Dammann grating

Now, the amplitudes $u_n$ (where $n = 0, \pm 1, \pm 2, ..., \pm N$) of the diffraction orders of this linear grating are given by $N$ equations relating to the $2N+1$ zeros (because by symmetry, $\xi_k = -\xi_k$) and an extra equation relating to the step height. The design of the Dammann gratings has the following requirements:

(i) equal brightness of the diffraction orders:

$$|u_n|^2 = |u_n|^2 \quad \text{for } n = 1, 2, ..., N$$  \hfill (2.6)

Hence, we have $N$ equations with $N$ degrees of freedom.
(ii) The diffraction efficiency of the grating is given by:

\[ \eta_{\text{diff}} = \sum_{n=-N}^{N} |u_n|^2 \]

and we try to achieve the highest possible efficiency

(iii) All the zeros in the solution \( x_k \) must lie within half of the grating period:

\[ 0 \leq \xi_k \leq \xi_{k+1} \leq \frac{1}{2} \]  

(iv) the minimum feature size limited by the fabrication process is the minimum distance between two transition points in the grating structure.

The number of possible solutions increases exponentially with \( N \) and a long computation time is required to find all the solutions for large \( N \). Usually, a root-finding algorithm such as Newton-Raphson method is used. The efficiency and the reconstruction error are similar for different sizes of array. However, the minimum feature size required for the grating increases rapidly as the array size grows which means that larger arrays are usually not limited by constraints from the fabrication process [Vas92].

The typical efficiency achieved for a 1-D Dammann grating is about 70-80% [Vas92] and is consistently over 80% for fanout greater than 14 beams. Smaller fanout has less efficiency because of the fewer degrees of freedom. The efficiency of a 2-D Dammann grating is the product of those for the individual 1-D gratings. For a centro-symmetric 2-D Dammann grating, the typical efficiency is then 50-65%.

Another way of fabricating a large array generator with a fanout of \( NM \times NM \) is to use two Dammann gratings with fanouts \( NxN \) and \( MxM \) [Vas92]. The pitch of the spots generated by the \( MxM \) grating must be bigger than the size of the whole \( NxN \) spot array, or vice versa. The penalties include an increase in the overall dimensions and a drop in efficiency, which is now the product of the individual efficiencies. For a separable design, the efficiency is about 40%.

Besides the diffraction efficiency and the size of the array, two other useful figures of merit are needed for specifying a Dammann grating. They are uniformity of the intensities of the spots generated (or the reconstruction error) and contrast between the spots and the maximum intensity of the background.
2.2.3 Non-separable gratings

A general non-separable binary-phase grating cannot be decomposed into two 1-D gratings. The zeros are indicated by the corners of the \( \pi \)-phase rectangular regions (Fig. 2.6b) and there is no restriction that rectangular regions be isolated, which is the case in separable gratings (Fig. 2.6a). The response is complex in general and has more unknowns and degrees of freedom. The Dammann method of solving the equations, which is described above, can also be used in 2-D non-separable grating designs [Mai89].

![Fig. 2.6 Structure of a single period of:](a) separable Dammann grating (b) non-separable Dammann grating (c) trapezoidal shaped grating

Other types of non-separable gratings include trapezoidal shapes confined in stripes (Fig. 2.6c) [Vas92]. These gratings have improved performance over the rectangular shapes inherent in Dammann gratings but the design complexity is increased. The efficiency of a large array generated by this technique is about 75% (10% better than the separable design). In practice, only an 8 x 8 phase grating has been fabricated.

2.2.4 Optimisation algorithms

In order to increase the diffraction efficiency of the grating, the number of unknowns in the grating design (and hence the number of degrees of freedom) must be increased. This changes to a problem of optimisation. A popular optimisation algorithm is known as simulated annealing. [Mai90],[Tru88b].
Annealing is a physical process of melting a solid and then letting its molten atoms cool slowly so that they will take the minimum energy configuration as they crystallise. Simulated annealing applied to the optimisation of a grating design is analogous to that process. It involves changing individual pixels sequentially or randomly (from 0 to $\pi$, or vice versa) so that the phase of the grating alters. If the change results in a reduction of the reconstruction error, the inversion of the phase is preserved. If, however, it brings about an increase in the error, then the change is accepted with a probability $\exp(-\Delta E/T)$ where $T$ is the effective temperature of the system and $\Delta E$ is the change in the error. As the error decreases, the temperature $T$ is reduced according to some annealing schedule so that changes that increase the error are accepted with decreasing probability. This procedure continues until no changes are accepted or a suitable error level (or efficiency) is obtained. By the application of simulated annealing, gratings generating 1001 x 1001 have been computed and 201 x 201 spots have been reconstructed experimentally with reconstruction errors less than 10% [Vas93],[Tru90].

Other optimisation algorithms include direct binary search [Sel87] and genetic algorithms. Direct binary search is very similar to simulated annealing except that it is a downhill iterative technique, i.e. if the change does not bring about an decrease in error, it is simply rejected. This has the disadvantage that it may stagnate in one of the local minima, the number of which increases exponentially with fanout [Dam77]. However, it is faster than simulated annealing. Genetic algorithms involve a change in a string of pixels every time and allow the transitions from a local minimum to lower energy minima.

2.2.5 Multiple and continuous phase profile gratings

Owing to the binary nature of the phase levels in a Dammann grating, the sharp transitions contribute to the formation of unwanted high orders. This results in a loss of efficiency to the desired response and an increase in errors. The maximum efficiency obtained from a 2-D binary-phase separable Dammann grating is about 65% which generates 128 x 128 spots. Multiple phase levels are needed to increase the efficiency of the grating further. For a significant increase in efficiency, more than four levels are required [Mil93]. This means that the configuration of the grating approaches that of the kinoform, which can be very
difficult to fabricate for a very large array. Therefore, the improved efficiencies should be balanced against the considerable increase in the complexity of the fabrication process.

Many designs of multi-phase level gratings have been reported [Byc90],[Ima91],[Wong93a]. Until recently, only a small array generator (5x5) with four phase levels was fabricated [Jah89]. The latest achievement was by Miller et. al. who made a 32x16 separable array illuminator with 16 levels [Mil93], with efficiency of 84.2%. Although the efficiency is about 6% lower than the theoretical limit, it is still about 20% better than separable binary gratings. The uniformity is about 11% due to the error introduced in the fabrication process. Non-separable multilevel phase gratings have even higher efficiency but the increase in efficiency is only significant when the array is smaller than 8 x 8. Therefore, for a large array generator, separable designs are adopted.

Continuous surface-relief gratings offer an alternative to multi-level structures and are, in general, capable of giving even higher efficiencies. With the advancing technology of high precision laser beam (or e-beam) direct write on the photoresist, complex profiles can be generated. A 9x9 diffractive element with separable design was fabricated by laser-beam writing, with efficiency of 94% (compared to 91% for a 16-level grating [Mil93]) and uniformity of ±8% [Gale93].

2.2.6 Fabrication

The fabrication technique for binary Dammann gratings involves the conventional process of transferring the binary mask onto the photoresist already spun on the substrate. After developing the resist, the pattern may be etched into the glass substrate in the following way [Jah89] (Fig. 2.7). First, the structure is coated with a thin layer of aluminium. Then, a lift-off step is used to remove those parts covered by resist and aluminium. Finally, reactive-ion etching is employed to etch the glass and also to ensure that the binary depth is about $\lambda/2$ for $\pi$ phase difference. For multiple levels, N binary amplitude masks for $2^N$-level gratings are first fabricated. The mask patterns are then transferred into the photoresist layer and etching or thin-film deposition is applied after each lithography step to create the desired surface-relief profile [Jah89]. Finally, direct laser (or electron) beam writing on positive resist can provide continuous surface relief gratings [Gale93].
A major fabrication problem is the lack of material of sufficiently high optical quality and controllable refractive index and thickness. Photoresist suffers from the lack of stability in thickness, refractive index and durability whereas it is difficult to create accurate structures in glass. Amorphous silicon nitride has been proposed as an alternative material for the fabrication of Dammann gratings [Tag89].

![Fabrication process of a binary phase grating using reactive-ion etching](image)

**Fig. 2.7 Fabrication process of a binary phase grating using reactive-ion etching**

### 2.2.7 Conclusions for Dammann gratings

Fourier phase computer generated holograms are capable of generating an array of point sources. The original Dammann gratings were binary, symmetric and one-dimensional. In a two-dimensional case, the Dammann design can be decomposed into one-dimensional gratings for computation. A Dammann grating is characterised by the number of output beams (fanout), the ratio of incoming light energy to the energy of the desired orders (efficiency), the uniformity of the array and the minimum feature size limited by the fabrication process [Kra90]. The design requires the numerical solution of a system of non-linear equations and there exists a degeneracy in the solutions. After the best solution is selected, the design is usually optimised by an algorithm such as simulated annealing. The largest design can generate more than 1000 x 1000 spots but only a 201 x 201 fanout grating has been fabricated. The Dammann method (summarised in Fig. 2.8) is reasonably good for large array designs, with the typical efficiency being about 65%. The advantage of
this method is that large arrays can be designed easily and fabrication of binary gratings is straightforward. The only drawback is that a smaller array (< 10x10) has lower efficiency because of the fewer degrees of freedom.

If higher efficiency than 65% is desired, non-separable design can be used to generate the phase gratings. Usually, it takes the form of polygon shapes such as trapezium for the ease of fabrication. An increase of about 10% theoretically results for these gratings. In practice, the difference is less than that and it is only worthwhile for small array designs.

An alternative is to use multiple phase level design and usually more than 4 levels are required to give a significant increase in efficiency. A 16-level 32 x 16 array generator was fabricated with efficiency of about 84%. Non-separable multi-phase gratings have higher efficiencies, but again, they are beneficial only for small array generators.

Finally, there is the ultimate design of continuous surface-relief gratings with efficiency of 94% for a 9 x 9 array illuminator. With the advanced technology of electron or laser beam writers, there should be a lesser problem of fabrication errors. However, a large array generator would require complex design, lengthy computation and the high accuracy of fabricating the minimum feature size. In that respect, binary-phase Dammann gratings are still more favourable than the multi-level or continuous phase relief structures. The theoretical efficiencies of the different types of phase gratings are outlined in table 2.1 and the best experimental results are given in table 2.2.
There has been a great deal of interest in utilising Dammann gratings in optical computing systems [Kri92] and optical interconnects [Mor93] as fanout elements in recent years. Other applications include image multiplexing [Dam77], star coupling [Vel86] etc.

<table>
<thead>
<tr>
<th>Phase structure</th>
<th>Design</th>
<th>2x2</th>
<th>4x4</th>
<th>8x8</th>
<th>16x16</th>
<th>32x32</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Binary [Mil93]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>separable (Dammann)</td>
<td>65.7%</td>
<td>50.0%</td>
<td>58.1%</td>
<td>66.6%</td>
<td>67.9%</td>
<td></td>
</tr>
<tr>
<td>non-separable (trapezoidal)</td>
<td>65.7%</td>
<td>77.5%</td>
<td>75.6%</td>
<td>76.2%</td>
<td>74.4%</td>
<td></td>
</tr>
<tr>
<td><strong>Multilevel [Mil93]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>separable</td>
<td>65.7%</td>
<td>84.2%</td>
<td>89.3%</td>
<td>91.6%</td>
<td>91.7%</td>
<td></td>
</tr>
<tr>
<td>non-separable trapezoidal</td>
<td>90.6%</td>
<td>94.1%</td>
<td>90.7%</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td><strong>Continuous [Gale93]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>separable</td>
<td>64.1%</td>
<td>84.6%</td>
<td>92.5%</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>non-separable</td>
<td>91.8%</td>
<td>95.6%</td>
<td>95.1%</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1 Theoretical efficiencies of different types of phase gratings

<table>
<thead>
<tr>
<th>Structure</th>
<th>Type</th>
<th>Fanout</th>
<th>Efficiency</th>
<th>Uniformity</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary [Mil93]</td>
<td>Separable (Dammann)</td>
<td>201x201</td>
<td>65%</td>
<td>± 8%</td>
<td>Easy design</td>
</tr>
<tr>
<td></td>
<td>non-separable trapezoidal</td>
<td>8 x 8</td>
<td>75%</td>
<td>± 4%</td>
<td>More complicated design</td>
</tr>
<tr>
<td></td>
<td>16-level (separable)</td>
<td>32 x 16</td>
<td>84%</td>
<td>± 11%</td>
<td>High efficiency but fabrication needs multiple masks</td>
</tr>
<tr>
<td></td>
<td>9 x 9</td>
<td></td>
<td>94%</td>
<td>± 8%</td>
<td>Very high efficiency &gt; 90% need laser/e-beam direct write</td>
</tr>
</tbody>
</table>

Table 2.2 The properties of various type of Fourier phase gratings fabricated
2.3 Fresnel zone plates

2.3.1 Introduction

A completely different type of computer generated hologram for array generation is called a Fresnel zone plate array. Dammann gratings require Fourier transform lenses and, thus, operate in the far field region. In contrast, Fresnel zone plates (FZP) use near field diffraction and they effectively incorporate diffractive lenses. The focusing ability of the FZP is brought about by diffraction as opposed to the bending of light rays by refraction in the denser medium of a conventional lens.

We first describe the theory used for generating a Fresnel zone plate (section 2.3.2). Then we discuss the different types of FZP according to the structures (section 2.3.3). It is followed by the description of different design methods of incorporating FZP in an array format (section 2.3.4) and in a sub-hologram or faceted array (section 2.3.5). Finally, we summarise this field in section 2.3.6.

2.3.2 Theory of Fresnel zone plates

A Fresnel zone is a concentric ring generated by the near-field diffraction of the primary wavefront surface from an observation point P [Hec87]. The boundaries of the zones correspond to the intersections of the wavefront with the series of spheres centred at P of radius $r_0 + \lambda/2$, $r_0 + \lambda$, $r_0 + 3\lambda/2$ and so on (Fig. 2.9a). By removing either all of the even or the odd zones, there will be a large increase in irradiance and we have a Fresnel zone plate.

![Fig. 2.9 (a) Propagation of a spherical wavefront and (b) the corresponding Fresnel zone plate](image-url)
The radius of the mth zone, $R_m$, can be calculated from [Hec87]:

$$\left( \frac{l}{\rho_0} + \frac{l}{r_0} \right) = \frac{m\lambda}{R_m^2}$$

(2.9)

where $\rho_0$ is the radius of curvature of the spherical wave and $m$ is an integer; this equation is reduced to:

$$R_m^2 = mr_0\lambda$$

(2.10)

if illuminated by a plane wave.

Obviously, the FZP is periodic in $R_m^2$ with a period $r_0^2$. The eqn. (2.10) resembles the thin lens equation and therefore the Fresnel zone plate behaves like a diffractive lens. The primary focal length of the zone plate is said to be:

$$f_i = \frac{R_m^2}{m\lambda}$$

(2.11)

When a collimated beam is incident on the zone plate, a real image is formed at the primary focus at $P$ and a virtual image at the same focal length in front of the zone plate, unlike a conventional lens (Fig. 2.10a). In addition, light also comes to other points at $\pm f_i/3$, $\pm f_i/5$, $\pm f_i/7$ etc. These are the high order foci. These properties are characteristics of in-line holograms.

A Fresnel zone plate is a diffractive optical element and a special class of hologram [Horm67]. Therefore, it is possible to design and fabricate it as a computer generated
hologram. It can be calculated by interfering a point source and a plane wave on the hologram plane (Fig. 2.10b)

A spherical object wave \((E_o)\) can be expressed by the equation:

\[
E_o = \frac{A e^{ikr}}{r} = \frac{A e^{i\pi(x^2+y^2)/z^2}}{r}
\]

(2.12)

where \(A\) is a constant. Assume that \(x^2+y^2 << z^2\), then by Binomial expansion:

\[
r = \sqrt{x^2+y^2+z^2} = z \left( 1 + \frac{x^2+y^2}{z^2} \right)^\frac{1}{2}
\]

(2.13)

and the object wave is approximated by:

\[
E_o = \frac{A}{r} \exp \left( jkr + j\pi(x^2+y^2)/(2z) \right)
\]

(2.14)

A travelling plane wave \((E_r)\) at normal incident on the hologram plane is given by:

\[
E_r = B e^{jkr}
\]

(2.15)

where \(B\) is a constant amplitude.

The irradiance of the interference pattern is then given by:

\[
I(x, y) = \left( \frac{A}{r} \right)^2 + B^2 + \frac{2AB}{r} \cos \left( \frac{\pi(x^2+y^2)}{f_1 \lambda} \right)
\]

(2.16)

where \(f_1 = z\) because the point source is the focus. The phase variation of the interference pattern is (according to fig. (2.9a)):

\[
\frac{\pi(x^2+y^2)}{\lambda f_1} = \frac{\pi R^2}{\lambda f_1}
\]

(2.17)

At the intervals when the phase is an multiple of \(180^\circ\) or \(m\pi\), and \(R = R_m\), the radius of the Fresnel zones, we have an identical equation to eqn. (2.10).
2.3.3 Characterisation of Fresnel lens

Fresnel zone plates can take the form of amplitude or phase holograms, depending on the fabrication technique. Amplitude holograms are simpler but the efficiency is very low (~10%). Therefore, most research is involved with phase Fresnel elements. There are essentially four types of phase diffractive elements: sinusoidal, binary, multi-level and blazed. Sinusoidal gratings have the lowest efficiencies (~34%) [Smi89] and we only concentrate on the other three structures here.

(i) Binary phase Fresnel zone plates

When designing binary Fresnel zone plates for on-axis applications, we must specify three parameters that are interdependent on each other: minimum feature size, focal length and diameter. There are restrictions on the range of focal lengths at a particular diameter.

In fabricating transmission binary gratings, we must ensure that the depth of the grating (d) provides an \( \pi \) phase shift at the particular wavelength so that:

\[
d = \frac{\lambda}{2\Delta n}
\]

where \( \Delta n \) is the difference in index between the substrate and the air. The technique employed is usually standard photolithography with the possibility of etching into the substrate afterwards.

The first order diffraction efficiency of a multi-level approximation of a Fresnel lens using the Fourier theory is given by [Gol90],[Cox90]:

\[
\eta = \text{sinc}^2 \left( \frac{\pi}{N} \right)
\]

where \( N \) is the number of phase levels of the zone plate. For a binary level FZP, the theoretical efficiency is 40.5%. Experimentally, 30% is achieved by neglecting the reflection losses in the interfaces [Tsa90],[Fuj81].

(ii) Multi-phase levels and blazed Fresnel zone plates
Another way to obtain a Fresnel zone lens from a conventional, refractive lens is by the decomposition of the latter. First, a refractive lens is dissected into many layers, each gives a phase delay of $2\pi$ at a particular wavelength, $\lambda$. Then, those layers which do not contribute to the lensing action are removed and we arrive at a lens with shapes shown in Fig. 2.11a [Bri92].

![Blazed zone plate diagram](image)

**Fig. 2.11** Blazed zone plate:
(a) Its cross-section and its resemblance to a conventional lens
(b) Fabrication

This structure is known as the blazed Fresnel zone grating and it works as if it is a normal refractive lens. This surface profile is quite difficult to fabricate and a useful approximation is to use multiple levels with stairstep shapes [Gol89]. The fabrication process of repeated masking and etching (or thin-film deposition) is similar to that for multi-phase Dammann gratings. From eqn (2.19) four, eight and sixteen levels have efficiencies of 81%, 95% and 99% respectively. In reality, a four-level FZP is about 60% efficient [Floo90] and an 8-level one can only achieve 91% of efficiency [Jah90].

The first design of blazed Fresnel zone plate was by Dammann in 1970 [Dam70]. There are three major methods of fabrication of blazed gratings:- electron beam, laser beam direct write systems and the use of binary halftone technology. Both e-beam and laser beam direct write systems fabricate blazed zone plates using the same technique, i.e., apply a predetermined dosage distribution to the resist across the whole substrate followed by the development of the photoresist (Fig. 2.15). A detailed comparison between the e-beam and laser beam lithography is described in [Haru90]. Electron beam lithography is favoured by many researchers [Aoy90],[Yam91],[Shio87,89,91],[Fuj82],[Tan89] because it has a
theoretical resolution of 0.2 μm (c.f. 0.7 μm for laser beam). However, scattering of electrons in the resist causes the broadening of the beam which results in inaccuracy throughout the depth of the resist. For example, the smallest resolution of a grating with the depth of 1 μm in photoresist is 2 μm (c.f. 1.4 μm for laser beam) [Haru90]. In addition, the electron beam lithographic process requires a conductive film of gold or ITO to avoid charging-up during exposure. Moreover, there is a limit on the size of the substrate that can be directly written by the electron beam writer. Finally, a laser beam writer is cheaper and does not require a vacuum environment. Although the theoretical efficiency is approaching 100%, in practice, the blazed FZPs are typically about 65-70% efficient [Shio87]. This could be due to the error introduced in the fabrication process. If we compare this to the 8-level Fresnel lens, it is surprising to find that the efficiency of multi-phase approximation is higher. Reflective blazed zone plates can be made by metallising one side of the diffractive element and elliptical blazed zone plates have been fabricated for off-axis imaging [Shio89,91].

The third method is to use a binary mask with varying density of high resolution pixels or a halftone mask. When applying lithography with the mask, the concentration of the dose is controlled by the grey scale [Mor][O'Sul90],[Pur93]. This halftone method has the advantage that once the mask is fabricated, many blazed Fresnel gratings can be reproduced. Another less well known technique of generating blazed gratings was reported by Hutley et. al. [Hut88] They proposed to manufacture blazed zone plates by recording in the photoresist circular fringes transmitted by a Fabry-Perot interferometer. The triangular groove profile is generated by recording a series of exposures. (Fig. 2.12) However, the numerical aperture is limited by the properties of the camera used to photograph the lens.

![Individual exposures to Fabry-Perot fringes](image)

Resultant groove profile

Fig. 2.12 Synthesis of a blazed profile from Fabry-Perot fringes
2.3.4 Generation of Fresnel zone plate arrays

There are many ways of generating an array of Fresnel zone plates for beamsplitting and multiple imaging purposes. The simplest of which is by joining individual zone plates together in an array format. The array can take the form of square-apertured [Flo90],[Jah90],[Shio87],[Stu], circular-apertured, rectangular-apertured [Shio87] (Fig. 2.13) or hexagonal close packed [Gol89,90]. The rectangular-apertured FZP array is found to be more efficient than normal square-apertured. The largest array of micro-Fresnel lenses made is about 245 x 245 in hexagonal closed packing.

![Arrays of Fresnel zone plates](image)

Fig. 2.13 Array of Fresnel zone plates:
(a) square-apertured, (b) rectangular-apertured and (c) hexagonal closed packed

Simple multiplexing of FZPs to form arrays has the problem of the effect of the aperture on the point spread function on the focal plane. Overlapping the zone plates over the same CGH space may eliminate these effects by [Tan89],[Fel87,89],[Kre93] (Fig. 2.14a). However, the side effect when overlapping the zone plates is the creation of a secondary interference pattern between the zone plates themselves, thus, creating scattering [O’Sul90] and resulting in loss of power. A more elegant and fundamental approach is to design a single element from the start, i.e. calculate the interference pattern between a plane wave and a source array. In this way, interference between sources can be eliminated. [Lee90a, Car92,93a]. In general, the spot size is smaller because the effective diameter of
each FZP is larger. Both analytical solution and numerical approximation has been used for a fanout of four (Fig. 2.14b). Larger fanout results in complicated calculations and longer computation. So far only binary elements generated by this technique have been made, which had an efficiency of 28%, similar to the conventional FZP without fanout.

Fig. 2.14 Fresnel zone plate arrays by:
(a) Overlapping of several zone plates
(b) A single Fresnel holographic element capable for generating four spots

2.3.5 Hologram facets

A hologram facet or a subhologram is a part of a zone plate which has a special beam steering property. It is less efficient than a full FZP because only a part of it is used to diffract light. In general, it gives a poor focus and a larger spot size which result in poor signal-to-noise ratio. The diffractive element is formed by multiplexing an array of facet holograms. Multi-facet holograms are sometimes needed to make the element more compact and to remove the unnecessary diffraction from other parts of the zone plates. Multi-facets can also incorporate multi-functionality in a single element [Kre93, God93] (Fig. 2.15a). In general, they are used for specific applications such as optical interconnects [O'Sul90, Stu, Kaw91, Wu87].
2.3.6 Other types of FZP arrays

A special type of Fresnel hologram can be designed by the spatial modulation of Fourier CGH by a FZP [Kre93] (Fig. 2.15b). First, a Fourier hologram is calculated to perform the fanout in the far field. Then, this function is modulated by a Fresnel zone plate. The imaging properties of these holograms are degraded by aberrations due to the focusing of all the spots by a single FZP.

So far, all the methods for generating the Fresnel holograms are non-iterative methods. The use of an error-reduction iterative algorithm to generate a Fresnel zone was proposed recently by Kress and Lee [Kre93]. The algorithm generates a phase hologram from a set of input (laser beam profile) and output conditions (interconnection pattern). First, the phase of the hologram function is randomised. Then, it is multiplied by the amplitude distribution of the incoming beam. This complex function is propagated to the output by the Fresnel transform, where the amplitude is replaced by that of the

![Fig. 2.15 Other types of Fresnel holograms for array generation](a) Multi-facet Fresnel hologram  
(b) Spatial modulation of Fourier CGH by a FZP  
(c) Fresnel hologram generated by error-reduction iterative algorithm
interconnection pattern. The phase distribution remains unchanged. An inverse Fresnel transform takes the new function back to the hologram plane. Again, only the amplitude is replaced by the laser beam profile. The process repeats until the algorithm converges. The resulting hologram is shown in fig. 2.15c. The hologram produces sharp uniform spots and has a good diffraction efficiency but the computation time is long.

2.3.7 Conclusions for Fresnel zone plates

A Fresnel zone plate is a diffractive lens and the lensing action is based on the near-field diffraction theory. It can be classified into binary phase, multi-phase and blazed zone plates. Although the blazed type is the closest approximation to a refractive lens, multi-phase zone plate with a large number of phase levels eg. a eight-level FZP has a higher experimental efficiency. Arrays of Fresnel zone plates can be generated by spatially multiplexing zone plates by square packing or hexagonal close packing. Other methods include overlapping of zone plates, calculation of the interference pattern of a source array, hologram facets, modulation of Fourier CGH by a FZP and the iterative design. However, we believe that simple multiplexing is still more favourable for a large array (or fanout) in terms of the simplicity in design and fabrication. A 245 x 245 array of hexagonal closed packed FZP has been demonstrated. Apart from array generation [Shio91] and beam steering [Gol89,90],[Flo90], Fresnel CGHs are also used for multiple imaging, optical interconnects [Shio91],[O'Sul90] etc. neural computing, aberration compensation [Vel92] and wavefront transformations [Lee90b] etc.

2.4 CGH based interconnect systems

2.4.1 Introduction

In this section, we will consider the two different configurations of the interconnect system: space variant and space invariant systems. They make use of the Dammann grating and the Fresnel zone plates respectively.
2.4.2 Space-invariant systems

The first and the most popular configuration is the space-invariant system. It comprises two Fourier transform (FT) lenses in a configuration shown in fig. 2.16. Due to the fact that the components are spaced a focal length apart, the total system length is $4f$ and it is also known as a Fourier 4-f system. The input is Fourier transformed to the middle plane by a FT lens. The system is space-invariant because Fourier transform of a translation does not affect the amplitude. Spatial filtering is achieved by using an element especially designed for this system, e.g. the Dammann grating. The Dammann grating provides the fanout and generates $M \times M$ spots from a single spot in the input. If the input spot is replaced by a pattern with resolution of $N^2$ pixels, in other words an input array of $N \times N$ elements, the $M \times M$ replications of the input patterns give $N^2 \times M^2$ elements at the output. This works

![Diagram of 4-f system](image)

*Fig. 2.16 Fourier transform (4-f) system: (a) System setup (b) as array generator and (c) as image multiplexer*
under the condition that the pitch (P) of the M x M array generated is larger than the
dimension of the input array, i.e. \( P \geq Np \) where \( p \) is the pitch of the N x N array of input
elements. In this way the input array is fully interconnected to the output plane.

The use of Dammann gratings in space-invariant interconnects are demonstrated in
optical switching systems [Mor93] and neural networks [Bam92],[Kri92],[Yay92] (an 8 x 8
neural network was built). In optoelectronic neural networks, however, the use of spatial
filtering is not just limited to beam splitting. The convolution of the input with the FT of
the filter at the output gives the possibility to a correlation of the input pattern with the
stored patterns whose FT is located at the filter plane [Heg90].

The limit of growth of the system comes in both the size of the Dammann grating
and the aperture of the lens. As the fanout of the Dammann grating becomes larger, the
physical size of the element is also increased. This in turn puts restrictions on the aperture
of the FT lens which must be large enough to cover the Dammann grating. The largest
fanout element fabricated can generate 201 x 201 number of spots [Vas92].

2.4.3 Space-variant systems

Space-variance means that an element is not necessarily interconnected with the rest of the
\( N^2 - 1 \) elements but is arbitrary connected to some. This is entirely opposite to a space-
invariant system which specifies total parallelism. At the highest level, a space-variant
system is a permutation network, such as perfect shuffle, butterfly etc., in which a few
connections are linked to each element [Sch91,92]. At the other end, a total interconnect
can also be implemented by this method with great complexity. In a fully space-variant
system, the aperture is partitioned between the channels, i.e., a certain facet of the aperture
belongs to each channel [Kra92]. A simple setup shown in fig. 2.17a uses a holographic
optical element to implement the dense interconnections. Each input element is linked to an
output by a particular hologram. Therefore, a \( N^2 \times N^2 \) array of holograms is required to
fully interconnect two N x N planes. These holograms may be faceted, spatially
multiplexed or overlapped on each other [Kre93]. Several neural networks based on this
configuration have been built [Whi88],[Rob90],[Kel91] but the largest is only 8 x 8 [Kel92].
An alternative configuration is to implement the system in a reflective mode as in the case of a folded multi-facet system proposed by Ozaktas et. al. (Fig. 2.17b) [Oza91]. Another implementation makes use of several optical elements such as microlens arrays and fan-in and fanout holograms (Fig. 2.18).

As the system size grows, more holographic interconnections are required and the system becomes very complex as it is difficult to diffract enough light into the large number of channels from one aperture. Consequently, it is not possible to build very large systems.

2.4.4 Conclusions

The space-invariant system is a common Fourier 4-f system. The system length is fixed at 4 times the focal length of the FT lenses and the limitation of the system growth comes from
the aperture of the FT lens. The single-stage space-variant system suffers from the complexity of the holographic interconnections required, of which $N^4$ elements are needed. Therefore, a large single-stage space-variant interconnect are not possible. The comparison of the two systems are given in table 2.3.

<table>
<thead>
<tr>
<th>Type of system</th>
<th>System length</th>
<th>Optical element</th>
<th>No. of elements for interconnecting $N^2 \to N^2$</th>
<th>Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space-invariant</td>
<td>$4 \times f_{FT}$</td>
<td>Dammann grating</td>
<td>single hologram with $N^2$ fanout</td>
<td>8x8 Neural Net</td>
</tr>
<tr>
<td>space-variant</td>
<td>depend on design</td>
<td>Hologram facets</td>
<td>$N^4$ facets</td>
<td>8x8 Neural Net</td>
</tr>
</tbody>
</table>

Table 2.3 Comparison between Space variant and space invariant systems

$f_{FT}$ - focal length of the Fourier transform lens
3.1 Introduction

3.1.1 Historical developments of microlenses

Although there have been a lot of recent developments in different types of microlens, they are by no means new optical elements. In fact, those lenses made by Hooke and Leeuwenhoek [VZu81] in the 17th century were small (diameter φ 0.7 mm and resolving power of 1.4 μm). One of the early applications that required lens arrays was Lippmann's work on integral photography [Lip08]. The limitations of the technology and rather few areas of application prevented the full development of this field. The first of the modern microlenses appeared in the form of graded-index (GRIN) rod lenses [Kob79], which behave like a section of graded index fibre with self-focusing ability. But it was not until the development of technology on microscopic scale which brought a significant resurgence of interest in microlens arrays about 10 years ago.
The whole subject of microlenses was reborn in the 80's when Oikawa et. al.[Oik81b] successfully applied the graded-index technology to fabricate an array of microlenses on a planar substrate in 1981. This opened up a wide variety of potential applications by stacking the planar microlenses onto other planar optical components to form 3-D systems, though the speculations were mainly involved with optical fibre communications at that time [Iga82]. However, miniaturisation of optoelectronic components together with advances in achieving uniformity across substrates enabled arrays of components to be fabricated. This in turn required microscopic lenslet arrays of good quality and uniformity which were reproducible. Although the planar microlens arrays had good quality, they involved a long process of ion-diffusion and suffered from disadvantages such as poor integration with other semiconductor components and were difficult to copy and mass-produce. Therefore, other methods were developed which produced surface-relief lenses on glass, semiconductors, polymers and photoresist. The most significant advance was by Popovic [Pop88] et. al. in 1988 who made microlens arrays by standard lithographic techniques on photoresist. The microlenses themselves can be etched into different substrates or copied onto different materials. Now, microlenses are used for a wide variety of applications, including collimation of laser beams from surface-emitting lasers [Cra92], to photodetectors [Mak88], beam coupling between optical fibres [Leg91], incorporation into liquid crystal displays [Lew93], imaging in photocopying machines [Bor91], 3D integral imaging [Dav91] and retro-reflective screens [Stev91], active control in telescopes to reduce atmospheric wavefront aberrations [Art91], optical interconnects [Hut92, McCo92] and optical neural networks [Agu90], and many others.

3.1.2 Layout of this chapter

The scope for microlenses over the past few years has increased tremendously and it is difficult to include all in this survey. However, we try to cover all the microlenses known to the author to date. Different types of microlens use different lens materials and a variety of fabrication processes. Microlenses can be roughly divided into three main groups:- graded-index profile (section 3.2), surface relief (section 3.3) and fibre microlenses (section 3.4). These will be described in more detail in the following sections.
We compare the properties of most of the microlenses in section 3.5. Section 3.6 describes of the imaging systems that use microlenses and the conclusions are given in section 3.7.

### 3.2 Graded-index profile microlenses

These microlenses comprise an inhomogeneous medium which provides the focusing properties as opposed to the shape of the air-glass interface for conventional lenses. The most successful fabrication involves an ion-exchange technique [Ham69, Hen75] which alters the refractive index gradient in the glass. There are two types of microlenses: rod-shape lenses and microlenses with hemispherical index profile in a planar substrate. Both types of lenses are manufactured commercially by Nippon Sheet Glass (NSG), Japan.

#### 3.2.1 GRIN rod microlenses

The GRIN (graded-index) rod microlens (or SELFOC lens, the tradename of NSG) is a glass rod of 1 to 3 mm in diameter, with a radial distribution of refractive index [Newport, Moo80, Mar82]. The index is at a maximum on the axis of the lens, and can be expressed as a parabolic relationship:

$$n(r) = n_0 \left(1 - \frac{Ar^2}{2}\right)$$

(3.1)

where $n_0$ is the refractive index on the lens axis and $\sqrt{A}$ is the quadratic gradient constant. The GRIN rod lenses are essentially the same as sections of graded-index fibres except that the fibre dimensions are much smaller and, hence, the fibre is used to guide signals as opposed to the imaging provided by the GRIN rod lens.

The GRIN rod lenses are fabricated using an ion-exchange technique, in which thalium and sodium ions in the silica glass diffuse out and potassium ions in the KNO$_3$ ion bath diffuse into the glass at 500°C (Fig. 3.1). Ti$^+$ ions are bigger than K$^+$ ions. Therefore, when the K$^+$ ions replace the Ti$^+$ ions, the difference in the polarizability causes a reduction in the refractive index of the glass. As the K$^+$ ions diffuse into the glass, they spread out so that the concentration is the greatest near the interface and there is little index change deep in the core of the glass rod. The resulting graded-index profile in the rod provides the
lensing action in the following way. A light beam incident on the end of the rod lens is bent by the gradual decrease in index and propagates inside the rod through a sinusoidal path with respect to the lens axis (Fig. 3.2). This is very similar to how light is guided by the graded-index fibre. A difference from the conventional lenses is that the numerical aperture (NA) of a GRIN lens varies with the position of the beam entering the lens with respect to the lens axis. It is largest on the optical axis and decreases to zero at the lens periphery. This is because the further away from the axis, the less angles at which the beam incident on the lens will be successfully guided by the lens medium.

![Fig. 3.1 Ion-exchange process of a GRIN rod lens](image)

![Fig. 3.2 Ray path in a GRIN rod lens](image)

The length in which the beam describes one full sinusoidal period is defined as the pitch of the length of the lens. One of the most useful lens lengths is 1/4-pitch because light travels exactly one quarter of the period. This means that a collimated beam is focused at the opposite end of the lens and vice versa. (Fig. 3.3)

![Fig. 3.3 1/4-pitch GRIN rod lens: (a) focusing of a collimated beam and (b) collimation of a point source](image)

GRIN rod lenses can also be made in plastic material [Koi82] but glass rod lenses are much preferred for many applications because of robustness. The aberrations of the microlenses can be evaluated by either an imaging technique [Yam80] or using a Twyman-Green interferometer [Cli82].
In a GRIN lens, the focusing is achieved by the lens medium, rather than by the air-lens interface as in the case of the conventional lens. This is very beneficial because the optical elements at each end can be held in place by index-matching glue or epoxy, making the system very robust [Kir91]. In addition, Fresnel reflection losses are eliminated. Therefore, one of the primary applications of these lenses is in optical fibre coupling with other optical components [Kaw80]. However, they can also replace conventional Fourier transform lenses in a Fourier-system, rendering it more rugged and compact [Ham91]. When the rod lenses are made into arrays, as in the case of the SELFOC lens array (SLA) [Nippon], they can be used in photocopiers [Tom80], LED/LCD/CRT printers and colour facsimiles.

3.2.2 GRIN planar microlens

The graded-index (or distributed-index) planar microlenses (PML) were pioneered by Oikawa, Iga and Sanada of Tokyo Institute of Technology. They first made a plastic planar microlens by means of monomer exchange diffusion [Oik81a]. Then they fabricated an array of 5 x 5 GRIN planar microlenses on a glass substrate with \( \phi = 1.2 \) mm and \( f = 9.4 \) mm [Oik81b]. The standard microlens arrays now made by NSG are 32 x 32 microlenses with \( \phi = 250 \) mm and \( f = 560 \) mm [NSG] although sheets of microlenses with diagonal of 17" can be made [Wong93b].

The advantages of planar microlenses arrays over GRIN rod lens arrays are that PML arrays can be fabricated monolithically and the precise formation can be controlled by using photolithographic techniques. There are three methods of fabrication:

(i) The first and the earliest method is to employ monomer exchange technique to make plastic lenses [Oh73]. The process requires shorter diffusion times than that for glass substrates and the resulting lenses usually have long focal lengths (e.g. \( f/8 \)) but fast lenses (i.e. small \( f/# \)) are necessary for some applications.

(ii) Planar microlenses are fabricated on glass substrates using the ion-exchange technique which is similar to that of GRIN rod lenses [Oik82, Nis91]. An array of circular windows is patterned photolithographically onto the metal mask which is already prepared...
on top of the substrate (Fig. 3.4a). The masked substrate is immersed in molten salts, in which ion-exchange take place between the higher index of Ti\(^+\) ions in the bath and the K\(^+\) and Na\(^+\) ions in the substrate. The process is slow (165 hrs) and the diffusion occurs both axially and radially, resulting in a microlens with a diameter bigger than the aperture of the mask. The higher the numerical aperture required, the longer the diffusion process. While the hemispherical profile of the lens refractive index (Fig. 3.4b) in the substrate is produced in the process, a curved surface profile is also created (Fig. 3.4c). This effect becomes more and more pronounced as the diffusion gets longer and longer. This swelling may be due to the high concentrations of Ti\(^+\) salt present which has a larger ion radius than the ions being replaced in the substrate. The usual planar microlenses have NA up to 0.3 and the swelled type planar microlenses have maximum NA of 0.57 [Oik90]. Lenses with higher NA’s are not possible without introducing significant aberrations. In general, the diameter of the microlenses range from 10 mm to 1mm with NAs from 0.05 to 0.57.

(iii) An alternative fabrication technique on glass substrates is electromigration. It can be used to produce lenses with short focal lengths in reduced fabrication time compared to the ion-exchange technique [Iza72]. However, the latter process is more popular.
The off-axis aberration properties of planar microlens were investigated by oblique ray tracing [Mis88]. It was found that at angles up to $5^\circ$ off axis, spherical aberrations, field curvature and coma were the most significant.

When used for optical communications [Oik90], PMLs have a particular advantage over other surface relief lenses in that anti-reflection coating can be easily applied on their planar surfaces.

Another benefit is that non-circular apertured lenses can be made with no dead space between them in the case of small NA microlens arrays. Other forms of graded-index microlens were also fabricated by ion-beam sputtering for specific applications, such as integration with an individual semiconductor laser diode [Mis88]. Other applications include collection of light for CCD arrays and imaging in optical parallel processing systems [Agu90, Aki90, Ham90a,b, Ham91b].

### 3.3 Surface relief microlenses

Surface relief microlenses behave like conventional lenses, with the lensing properties depending on the curvature of the lens and the lens medium index. There are several methods of fabricating these lenses on different substrates and once they are made, they can be copied onto other materials as well. These will be discussed in the following sections.

#### 3.3.1 Photoresist reflow microlenses

The aim of integrating of a microlens with a LED brought about the first photoresist microlens in 1981 by O. Wada et. al. [Wada81]. However, it was not until 1988 when Z. Popovic et. al. [Pop88] applied the technique for the fabrication of microlens arrays. These microlenses can be subdivided into five kinds:

(i) Binary-patterned photoresist microlens

The earliest and the most common method, as proposed by Wada and Popovic, uses a photolithographic process to deposit patterned circular photoresist discs on a glass
substrate. During subsequent melting at about 140-160°C, the photoresist is pulled into a spherical shape by the surface tension (Fig. 3.5). The resulting microlenses have excellent qualities and diffraction limited performances. In order to obtain a more robust lens, the photoresist lens shape can be transferred to the substrate beneath by reactive-ion etching. This means that microlenses of any material such as Si, GaAs, InP as well as glass can be fabricated. In addition, lens shapes other than spherical are possible by just altering the mask. For example, cylindrical microlenses can be made by using long strips on the mask [Ish93] and toroidal / ellipsoidal microlenses have been fabricated by us using elliptical apertures [Car93b]. Furthermore, once an array with good quality lenses is made, they can be replicated by epoxy resin for mass production. This simple and versatile monolithic fabrication technique has attracted a tremendous amount of interest because the microlenses can be made in-house and integrated with other devices. There is, however, a disadvantage of having gaps between microlenses which constitute dead space. They are necessary to prevent the photoresist of neighbouring lenses joining up during reflow. Usually, gaps of at least 5 μm are required.

Hutley et. al. [Hut90a,b, Dal91, Hut91] extended the study on this technique by reporting in detail the design, fabrication and measurements of these microlenses. They devised a formula between the thickness of the photoresist, the radius of the lens and the
desired focal length [Dal91] which is very important in the design of microlenses. The photoresist used in fabrication is viscous to achieve thick lenses with short focal lengths. The action of the surface tension works well if the resist is sufficiently thick. Thicker lenses are possible with multiple layers of resist spun on the substrate. This means that lenses with short focal length are easier to make than those with long focal lengths. Indeed, the range of the f-number was from 0.8 to 2.5, although now it is possible to make f/4 photoresist microlenses [Dal93]. This is very advantageous for integration of lens with an optoelectronic device.

Hutley et. al. also proposed that longer focal lengths could be achieved by enclosing index-matching fluid between the lens and a cover glass although they did not report experimental results. In addition to the standard type of microlens, they made three types of lens array: cylindrical, square aperture lenses and circular lenses with hexagonal close packing (HCP). Another way of making square lenses is by applying photolithography with the striped mask for cylindrical lenses twice orthogonally. Square-apertured lenses and HCP circular lenses have much less dead space than other square packed lenses, thus, reducing the optical power loss which may be very important in certain applications. However, square-apertured lenses have poorer focal concentration of energy.

(ii) Mask-preform photoresist microlens

An extension to the standard technique is to use several masks to create multiple levels on the photoresist cylinders before annealing [Dal91]. This process is known as "Pre-forming" (Fig. 3.6). This early approximation of the final lens shape can determine the focal length and microlenses with longer focal lengths than those generated by the standard technique are possible. This is because a shallow profile of the lens can be created by preforming.

\[\text{Fig. 3.6 Preforming of the photoresist microlenses}\]
(iii) Electron-beam direct write photoresist microlens

This technique uses the variation of electron-beam dosage on the positive resist and lenses can be directly written on the photoresist without the need for a mask. It can be used in two ways. First, a multiple level photoresist island can be written with a few levels and then thermal annealing is used to smoothen the surface, just as in "preforming" [Bir91a,b]. An array of 16x16 10-level preformed microlenses were made.

The second type of microlenses is generated by refining the resolution of the variable dosage so that the actual lens profile can be written directly. An array of 20 x 20 f/4 square microlenses (50 x 50 mm each) were made [Yam91, Aoy89]. The lens shape can be further smoothened by the thermal reflow process [Jay94] and a f/4.5 microlens (φ 250μm) with less than 1 λ of aberration is fabricated. Such good quality microlens with this f/# this is almost impossible to achieve with the standard binary patterned reflow technique. Aspherics such as elliptical microlenses can be generated by this method [Shio92]. The elliptical microlens was (NA = 0.14, lens size is about 100mm) coated in a reflective layer for use in reflection mode at 20° off-axis.

(iv) Optical generation of photoresist microlenses

It is possible to generate lenses with hexagonal packing by exposing the photoresist to three equally inclined beams and then developing it [Cow84]. Hutley [Hut90c] showed that arrays of diverging lenses can be fabricated in this way (Fig. 3.7a). Equally, positive lenses can be made from replications. Besides the spherical diverging lens, a series of quasi-cylindrical lenses of positive power are also made. They give rise to a 'chicken-wire' image (Fig. 3.7c). An elegant way of fabrication is to use the chicken-wire image as a mask onto a second surface with photoresist coating, to give converging lenses after using the standard technique of photoresist reflow.
Fig. 3.7 (a) Cross-section of the microlens array showing spherical divergent lenses and quasi-cylindrical concave lenses
(b) Virtual images of point source
(c) Real "chicken-wire" images formed by the convex lenses

(v) Substrate microlenses by etching

Finally, the photoresist microlenses can be etched into different substrates beneath the resist coating by using reactive ion etching. The resulting profile of the substrate microlens depends on the differential etching rates of the photoresist and the substrate by the gases chosen. The gas mixture for etching into fused silica is oxygen and CHF$_3$ [Mer92, Eis93]. f/5 lenses have been made, which cannot be obtained from the conventional photoresist microlens. It was observed that, for an array of square lenses, the etching process causes the lenses eventually to touch each other, and eliminates all the dead space [Eis93].

Semiconductor (Si) microlenses are also fabricated by reactive ion-beam etching [Mak88, Wad88, Nak93, Ster94] for use in longer infra-red wavelengths.

3.3.2 PMMA microlenses

Microlenses can be fabricated in polymethyl methacrylate (PMMA) by deep proton irradiation [Fra60, Kuf93] (Fig. 3.8). A mask is used to define circular areas of the substrate to be irradiated by a high energy proton beam. Afterwards, these areas have less molecular weight because the polymer chains are split. Then, the substrate is placed into an atmosphere of monomer vapour. The diffusion of monomer causes the volume of irradiated domains to swell and by surface tension, these form spherical shapes. The resulting
structures can be fixed by photoinitiated polymerisation. By controlling the energy of the protons, the depth of penetration can be fixed. This in turn determines the resulting curvature. This method can produce very high numerical aperture, e.g. > 0.5 (or smaller than f/1). Replication can be achieved by casting.

![Diagram showing the fabrication process]

**Proton beam irradiation**

**Monomer diffusion**

**Volume expansion of the irradiated domains**

Fig. 3.8 Fabrication of PMMA microlenses

### 3.3.3 Semiconductors microlenses using mass transport technique

Another technique for producing semiconductor microlenses monolithically is to use the mass transport mechanism. This is analogous to the preforming of microlens shapes in photoresist. Multilevel mesa structures are created on the substrate by repeated application of photolithography and wet chemical etching. Then the structures are annealed to form a smooth lens shape by mass transport in an environment with $\text{H}_2$ and $\text{PH}_3$ flowing at high temperatures (800 - 880 °C) for 2 - 20 hours (Fig. 3.9). The advantage of this technique is that the surface profile and hence the focal length can be controlled precisely. Microlenses were made in InP [Lia88] and GaP [Lia89] and were integrated monolithically with diode

![Diagram showing the fabrication process]

**Chemical etching**  **Mass transport**

InP (or GaP) substrate  multilevel mesa structures  InP (or GaP) microlens

Fig. 3.9 Fabrication of semiconductor microlenses by mass transport
lasers [Lia90a,b,c]. An improved version of this technique is to make use of the variation of mesa width to control the curvature. [Lia92].

### 3.3.4 Photosensitive glass microlenses

Corning first developed the photothermal technique using photosensitive glass substrates for the fabrication of spherical microlenses [Bor85,88,91,93 Bel88, Bara89] (under the tradename SMILE). The photosensitive glass has Ce$^{3+}$ ions which reduce the silver ions in the glass to silver if irradiated by u-v light. The traces of silver and other metals such as gold and copper formed are known as noble metal colloids and give a coloured appearance to the glass. During the u-v radiation exposure, a mask of opaque circles on a transparent background is placed on top of the substrate so that the circles masked off in the substrate remain unaffected while the surrounding regions are coloured. During the subsequent thermal development at 600° C, the noble metal colloids serve as nuclei for a lithium metasilicate microcrystalline growth from the homogeneous glass. As a result, the crystallised regions become harder and squeeze the soft undeveloped glass beyond the original surface into spherical shapes (Fig. 3.10). The microlenses formed are bi-concave and the dead spaces are opaque which reduce loss and cross-talk. Applications include use in photocopiers, optical fibre interfaces and auto-focus cameras.

![Fabrication of SMILE lenses](image)

**Fig. 3.10 Fabrication of SMILE lenses**

![Fabrication of polymer microlens array](image)

**Fig. 3.11 Fabrication of polymer microlens array**

### 3.3.5 Other polymer and plastic microlenses

Plastic lenses of diameters greater than 2 mm are traditionally made by injection moulding [Trib91]. This technique has the disadvantage of high initial cost of the metal mould and
the optical quality is not very good. However, if the mould is made using diamond turning technology [Pas92], then diffraction-limited plastic lenses are possible. The drawback is that the metal mould is even more costly.

Another type of plastic microlens described by Pantelis et. al. [Pan92,94] uses a much simpler and cheaper technique. It is known as the hot pressing method. A sheet of thermoplastic such as polycarbonate, PMMA or polystyrene is heated until it softens sufficiently to be permanently deformed (~ 200°C). It is then pressed against a perforated stainless steel sheet. The circular apertures are created by chemical etching techniques. The softened plastic forms a protrusion at the apertures, thus, forming the lenses (Fig. 3.11). The steel sheet is then removed and the resulting microlenses have reasonable qualities because the curved surface of the lenses was not in contact with any solid surface. Lenses of diameter 1.5 mm have focal lengths from 1.9 mm to 3.8 mm depending on the pressing time (30 - 60s). The lenses do not have good sphericity over the whole surface. The degree of sphericity is dependent upon the diameter and the focal length.

3.3.6 Complex surface relief microlenses

A new and elegant technique devised by Purdy [Pur93] for the fabrication of microlenses with any shape and size, is known as the Halftone micro-optic technique. By varying the pixel size and/or the pixel density, the grey scale information is encoded spatially as continuously varying neutral densities (Fig. 3.12a). A digital mask has been fabricated with over 10,000 different grey levels with a sub-micron pitch size by using an e-beam reticle writer with positional increments of 0.1 μm in both x and y directions (Fig. 3.12b). With such mask, a single e-beam dosage is all that required to generate near analogue patterns on the positive photoresist. This method is clearly much better and more versatile than using multiple exposures with a mask set or direct-write e-beam which were discussed earlier. Spherical microlenses have been fabricated using this technique (Fig. 3.12c). However, its elegance does not stop here. It can be used to fabricate surface relief diffractive optical elements such as blazed Fresnel lenses. Blazed gratings are normally made using direct-e-beam dosage but the Halftone technology can create gratings with steep vertical angles and sharp corners. In addition, once the mask is created, the components can be reproduced
Refractive Microlens review

with great ease. Lenses that have never been made before are possible, such as the
dispersive microlens (a grating on top of a surface-relief microlens) [Gal93].

Fig. 3.12  Microlenses fabricated using halftone micro-optic technology
(a) Design options of a unit cell of the halftone mask
(b) Halftone mask
(c) Microlenses

3.3.7 Other surface-relief microlenses

Another company which fabricates microlens arrays is called Adaptive Optics Associates.
They can fabricate arrays on epoxy, zinc selenide, elastomers, plastic, sol-gel quartz and
glass [Ada93]. Although there is no publication on the fabrication process, we believe that
they use the photoresist technique. The lenses usually have very high fill factor (> 98%).

An interesting technique by Lohmann et. al. [Loh93] uses a punched hole as a seed
for etching into a spherical shape. The holes are dented by the diamond tip of a hardness
tester and are etched by fluoric acid. Plano-concave lenses result, which can be optical
elements on their own right or serve as a mask for replicating convex lenses. Lenses made
are very small, of the order of a few μm. The formation of a doublet was proposed by
applying a thin optical glue on the concave surface. As the solvent evaporates, the glue will
then form a spherical droplet.

Yet another technique produces microlens arrays on materials such as
phosphosilicate glass (PSG) and polysilicate [Ens93]. First a PSG layer is deposited on a
silicon wafer by a chemical vapour deposition process. Then photolithography is employed
to lay down a circular pattern on the photoresist. The patterns are subsequently plasma
etched into the PSG. Reflow of the PSG occurs when treated with wet oxygen at 950 - 1100° C. Microlenses with sizes of a micron can be formed in this way. By depositing a layer of polysilicon on top of the etched PSG cylinders and repeating the same process of photolithography, etching and reflow, microlenses in polysilicon are formed.

Microlenses are also fabricated in other materials such as silica using sol-gel processing [Nogue92] and in photopolymer [Phi91] (see section 6.4.6)

3.4 Fibre microlenses

(a) (b) (c)

Fig. 3.13 Fibre microlens: (a) tapered-core lens, (b) tapered-cladding lens and (c) asymmetrical hyperbolic microlens

Individual microlenses can be fabricated at the end of optical fibres for coupling between fibres and lasers. The simplest technique is to pull and melt the fibre end, which results in a tapered-core lens (Fig. 3.13a) [Barn93]. Alternatively, the fibre may be tapered by acid etching which gives a tapered-cladding lens (Fig. 3.13b). A third method is to use laser micromachining of the end of the fibre to form a microlens [Pre90]. Hyperbolic microlenses (Fig. 3.13c) can be fabricated in this way to couple the highly elliptical beam of the laser diode into the fibre [Pre93, Edw93].

3.5 Comparison of the different types of microlens

A summary of all the properties of different types of microlens are given in table 3.1. The three parameters (diameter f, focal length f and NA) are represented in chart forms in fig. 3.14(a), (b) and (c). Finally, the benefits and drawbacks of individual methods are outlined in table 3.2 for comparison.
<table>
<thead>
<tr>
<th>Type of microlens</th>
<th>Lens diameter (μm)</th>
<th>Focal length (μm)</th>
<th>Numerical aperture</th>
<th>Technique/shapes/substrates</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELFOC (NSG)</td>
<td>1000 - 3000</td>
<td>1500 - 6250</td>
<td>0.1 - 0.6</td>
<td>- Ion exchange technique&lt;br&gt;- rod-shaped lens with gradient index profile&lt;br&gt;- glass</td>
</tr>
<tr>
<td>Planar (NSG)</td>
<td>10 - 1000</td>
<td>30 - 50,000</td>
<td>0.01 - 0.5</td>
<td>- Ion/monomer exchange, electromigration&lt;br&gt;- planar with spherical index profile&lt;br&gt;- glass, plastic</td>
</tr>
<tr>
<td>Photoresist (NPL)</td>
<td>5 - 760</td>
<td>15 - 1300</td>
<td>0.13 - 0.6</td>
<td>- Photolithography + photoresist reflow, preform&lt;br&gt; + photoresist reflow, direct e-beam write, optical generation, halftone mask, etching into substrate&lt;br&gt;- spherical, square, hcp, cylindrical&lt;br&gt;- photoresist, substrate e.g. glass, semiconductors</td>
</tr>
<tr>
<td>PMMA (CNRS)</td>
<td>50 - 1000</td>
<td>50 - 5000</td>
<td>0.1 - 0.5</td>
<td>- Deep proton irradiation&lt;br&gt;- spherical&lt;br&gt;- PMMA</td>
</tr>
<tr>
<td>Semiconductor (MIT)</td>
<td>67 - 130</td>
<td>85 - 200</td>
<td>0.39 - 0.75</td>
<td>- Preform by etching + Mass transport technique&lt;br&gt;- spherical&lt;br&gt;- InP and GaP</td>
</tr>
<tr>
<td>SMILE (Corning)</td>
<td>70 - 1000</td>
<td>100 - 85,000</td>
<td>0.003 - 0.35</td>
<td>- Photothermal technique&lt;br&gt;- spherical&lt;br&gt;- photosensitive glass</td>
</tr>
<tr>
<td>Polymer and plastics (BTRL)</td>
<td>1000 - 2000</td>
<td>1900 - 3800</td>
<td>0.2 - 0.55</td>
<td>- Hot pressing method&lt;br&gt;- spherical&lt;br&gt;- thermal plastics e.g. PMMA, polystyrene</td>
</tr>
<tr>
<td>Monolithic lenslet modules</td>
<td>100 - 1000</td>
<td>200 - 260,000</td>
<td>0.002 - 0.1 (spherical) &lt; 0.5 (aspherics)</td>
<td>- not known&lt;br&gt;- Square, rectangular, circular or hexagonal&lt;br&gt;- epoxy, ZnSe, plastics, sol-gel quartz and glass</td>
</tr>
</tbody>
</table>

Table 3.1 Comparison of the properties of different types of microlenses
### Table 3.2 Comparison of the merits of different types of microlens

<table>
<thead>
<tr>
<th>Microlens type</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
</table>
| SELFOC        | - No air gap between components  
               - as FT lens → compact system  
               - AR coating can be applied easily | - Non-monolithic fabrication  
               - large sizes & long fs  
               - Long process of diffusion |
| Planar        | - can have very long fs  
               - can eliminate dead space | - Long process of diffusion  
               - Poor quality high NA lenses |
| Photoresist (binary) | - Very simple method & versatile  
                         - good quality large NA lenses | - cannot form long fs lenses  
                         - have gaps between lenses |
| Preforming / direct write | - can design for any curvature | - complex technique  
                           - costly equipments involved |
| Etched substrate | - any material depending on λ | - Etching process is involved |
| PMMA          | - High NA lenses can be made | - complicated technique |
| Semiconductor | - monolithic integration with other components  
                 - control over curvature | - complicated technique |
| SMILE         | - bi-convex lens  
                 - can have very long fs  
                 - opaque dead space | - Very expensive  
                 - cannot have large NA lens |
| Plastic       | - large lens and long f  
                 - simple technique  
                 - cheap to reproduce | - do not have excellent quality  
                 - have dead spaces |
| Halftone      | - versatile technique  
                 - can design any shape to great precision  
                 - no dead spaces | - limited by resolution of mask writer |
| Adaptive Optics | - a variety of shapes, substrates and sizes  
                          - very high fill factor | - expensive for what it is worth |
Fig. 3.14 Comparison of (a) diameter, (b) focal length and (c) NA of various microlens
3.6 Imaging systems

3.6.1 Introduction

Systems which make use of microlenses as the interconnection elements are invariably imaging systems. The microlens array can be one of the refractive types described in this chapter or the diffractive Fresnel zone plate array of chapter 2. The system usually consists of one or a series of microlens arrays to image the input in such a way that the input is fully interconnected with the output.

3.6.2 Image multiplexer

The simplest multiple imaging system is the image multiplexer (Fig. 3.15) [Aki90], in which each microlens in a N^2 array images the entire N x N input plane to the output. Ideally, the resolution of the microlens is chosen to give N^4 output elements and the SLM (with N^4 pixels) is placed next to the output for controlling the N^4 connection weights independently. However, it is difficult to realise such a high density of output elements as it will have similar dimensions to the input array and microlens array due to short microlens focal lengths. So a common variant of this system is one having the same number of N x N pixels of the same size as those in the input array. Each microlens form an image of the whole input plane within a single output pixel resulting in a total interconnect of all input and output pixels.
The parameters which are important in the design of optical array systems are the type of illumination, the pitch of the microlens and the input pixel size. One of the requirements for the system to work is that light from the far corner of the input array must reach a microlens in the opposite far corner so that each microlens can image the whole input array. Under coherent illumination, care must be taken to allow this to happen and a diffuser may be required to send illumination beams in random directions. In addition, coherent illumination brings about the problem of speckle and, hence, the image quality is degraded. On the other hand, there is no such problem with incoherent illumination and diffusers may not be required depending on the choice of the incoherent source. The system size is determined by the size of the images at the output. By changing the separation distance between the input and the microlens array, different demagnification of the input results. The optimum distance between the input and the microlens array is obtained when the image just fills up the whole of the output N×N subpixels. If the input plane and the microlens array have similar dimensions, then by geometrical considerations, their separation must be Nf while the image plane is at a distance f away from the microlens array.

The image multiplexer suffers from off-axis imaging aberrations as the size of the input elements and the microlenses are comparable. Consequently, off-axis aberrations, field curvature and image shifts at the output are important issues to be considered in the design of the system.

This simple system has been investigated theoretically [Sak90] and is favoured by many researchers in optical bus interconnects [Sak91a,b,c] and optical neural networks e.g. [Yu90a]. In particular, Yu et. al. have demonstrated an 8 x 8 optoelectronic neural network...
using an array of 8 x 8 lenslets (somewhat larger than microlenses) and two liquid crystal TV screens, for the input and the weights (fig. 3.16). They implemented a Hopfield net [Yu90a], an inter-pattern association net [Yan90,91,Yu90b] a neural net with shift and rotation invariance [Yu91a], a Kohonen self-organising net [Lu90] and two Hamming nets [Yan92, Uan93, Yu91b,92]. In the Hamming net implementation, they use 12 lenslets for the classification with 12 stored patterns and they have demonstrated both monochromatic and polychromatic (colour) models.

3.6.3 Shared microlens system

In a shared microlens system, an array of \((2N-1)^2\) microlenses are used to image the input onto each output element (Fig. 3.17), in contrast to the \(N^2\) microlenses in the image multiplexer. It works in such a way that a subarray of \(N \times N\) microlenses images the \(N \times N\) input onto one output pixel. By moving the subarray to the next row or column, the image is focused onto the adjacent pixel on the next row or column respectively. Therefore, each microlens links several inputs and outputs. In other words, the microlenses are shared and, hence, the name of the system.

![Diagram of shared microlens system](image)

Fig. 3.17 Shared microlens system

In the special case when the distance between the input and the microlens plane is equal to that between the microlens plane and the output, one-to-one imaging is achieved. This means that each separation must be \(2f\). The microlenses in this system effectively displace the full-size multiple images and overlap them at the output. In the analysis given
by Sakano et. al. [Sak90] of this system, the optimum ratio of the separations is 1:1, i.e. a
2f:2f system.

One of the advantages of this system is that it is very compact. Since the whole of
the N x N input is now imaged by an equal number of microlenses, the off-axis angles
involved are smaller than the previous system and therefore the aberrations should be
considerably less severe. In addition, other types of aberration are also reduced because of
the one-to-one imaging. Furthermore, overlapping of images at the output pixels is inherent
in this system, thereby, averaging the aberrations.

The disadvantages associated with this design are the loss of light and the loss of
several degrees of freedom in controlling the weights. An SLM can be implemented next to
the microlens plane to allow limited switching. However when a particular microlens is
switched, the whole group of interconnections passing through that microlens will be
switched together. Consequently, we have a system which is totally interconnected (N^4) but
only have of the order of N^2 independent weights and, therefore, a novel learning algorithm
is required to implement a neural network.

This system has been theoretically analysed [Sak90], however, only one system
using a 5 x 5 pinhole array rather than microlenses has been practically implemented as a
Hopfield neural network [Noguc90]. The pinholes serve to select only those rays of interest
but we show later (in chapter 5) that this reduce the overall efficiency.

3.6.4 Sakano's Theory

Sakano et. al.[Sak90] analysed both the image multiplexer and the shared microlens
systems under ideal conditions. They argued that the system size is limited by astigmatic
aberrations and by the diffraction limited spot size. The total maximum number of
elements, N_{max}, in the image multiplexer is:

\[ N_{\text{max}} = \left( \frac{L}{9.76\lambda f} \right)^{\frac{1}{3}} \]  (3.1)

while that for the shared microlens system is:

\[ N_{\text{max}} = \left( \frac{L}{9.76\lambda f} \right)^{\frac{1}{2}} \]  (3.2)
where \( L \) is the system length, \( I = (3n+1)/4n \) and \( n \) is the refractive index of the microlens. The relationship between \( L \) and \( N_{\text{max}} \) is shown in fig. 3.18. Clearly, the shared microlens system is more compact for the same array size. This follows from the facts that it is a one-to-one imaging system and less off-axis aberrations exist.

![Graph showing \( N_{\text{max}} \) versus system size for the two systems from Sakano's theory.](image)

Fig. 3.18 Plot of \( N_{\text{max}} \) versus system size for the two systems from Sakano's theory

### 3.6.5 Hybrid system

There is another unifying way to explain the operation of an image multiplexer. When the waves propagate from the input plane to the vicinity of the microlens array, Fraunhofer diffraction of the waves occurs because the input lies in the far field region of the microlens array. The Fourier transform of the input is then inverse-FT-ed by individual microlenses to form multiple images. Therefore, a FT lens can be inserted into the system to perform the Fraunhoffer diffraction without any change in the function of the system (Fig. 3.19). This is the basis of the hybrid imaging system introduced by Hamanaka et. al [Ham90a,b].

![Diagram of hybrid imaging system](image)

Fig. 3.19 The equivalent configuration for the image multiplexer
In order to achieve the multiple Fourier transformations, the input must be illuminated at different angles. In the system proposed by Hamanaka, this special illumination is obtained by making the setup symmetrical (Fig. 3.20a) Another microlens array L1 is used to generate multiple beamlets from a collimated beam. Then a Fourier transform lens L2 turns the beamlets into a series of angular plane waves which are used to illuminate the input, which is now in the middle of the setup. As before, multiple Fourier transforms form at a focal length away from the microlens array L4 and multiple images are obtained on the other side of the array. This setup has the advantage that coherent illumination, which is not possible with the image multiplexer, can be used. In addition, the problems of off-axis imaging are reduced in this setup. The system can be made more compact by the use of SELFOC GRIN-rod lenses as FT lenses [Ham91a,b] (Fig. 3.20b). A neural network was proposed using this setup [Ham91].
Fig. 3.20 (a) Hybrid (Fourier-imaging) system
(b) implementation using GRIN rod lenses and planar microlens arrays
(c) Folded hybrid system (reflection type)

The aperture of the FT lens is the major factor limiting the growth of the system size as in the case of the space-invariant system.

So far, the systems described operated in transmission but they can be designed to work in reflection. A folded setup is shown in Fig. 3.20c and has half the system length. Besides their use for multiple imaging in neural networks, multiple Fourier transforms are also used in parallel matched filtering for pattern recognition in neural networks [Agu90]. Other uses of the setup are found in discrete correlators [Tan91],[Miy93]

3.6.6 Lenslet-pair imaging system

A system recently proposed by Marchand et. al. [Marc93],[Mars93] makes use of a pair of microlens arrays for the total interconnect. Two N x N microlens arrays form a tele-centric pair for multiple imaging (Fig. 3.21). An 8 x 8 interconnect has been built [Vas92].

![Fig. 3.21 Lenslet-pair imaging system](image)

3.6.7 Comparison of the systems

A brief summary of all the imaging systems in terms of system length and the number of microlenses are given in table 3.3.
Refractive Microlens review

<table>
<thead>
<tr>
<th>Type of system</th>
<th>System length</th>
<th>Optical element</th>
<th>No. of elements for interconnecting $N^2 \rightarrow N^2$</th>
<th>Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image multiplexer</td>
<td>$(N+1)f_m$</td>
<td>1 microlens array</td>
<td>$N^2$ microlenses</td>
<td>8x8 Neural Net</td>
</tr>
<tr>
<td>Shared microlens</td>
<td>$2f_m$</td>
<td>1 microlens array</td>
<td>$(2N+1)^2$ microlenses</td>
<td>5x5 Neural Net</td>
</tr>
<tr>
<td>Hybrid</td>
<td>$4 \times f_m + 4 \times f_m$</td>
<td>2 microlens arrays</td>
<td>$2N^2$ microlenses</td>
<td>330 interconnect</td>
</tr>
<tr>
<td>Lenslet-pair</td>
<td>depends on design</td>
<td>2 microlens arrays</td>
<td>$2N^2$ microlenses</td>
<td>8x8 interconnect</td>
</tr>
</tbody>
</table>

Table 3.3 Comparison between different imaging systems

- $f_{FT}$ - focal length of Fourier transform lens
- $f_m$ - focal length of microlens

3.7 Conclusions

We have reviewed many types of microlens, ranging from graded-index thorough surface relief, to micro-machining of individual fibres. Graded-index microlenses are mainly fabricated by the ion-exchange diffusion process, which takes a long time but they can have very long focal lengths.

The surface relief microlenses can be made by many different techniques on many different substrates. However, the main theme of the process involves lithography or e-beam direct write for monolithic fabrication. In most methods, reflow is required for the spherical lens shape to form. Photoresist reflow microlenses will probably have more impact in the field because of the versatility and easy accessibility of the technique, especially in small NA microlenses ($f/1$) which have excellent qualities. A noteworthy method from the remaining techniques is the Halftone technology. This technique should revolutionise the fabrication of both diffractive and binary optics.

In both the image multiplexer and the shared microlens system, the system size is related to the system length because of the restrictions due to the off-axis aberrations and the diffraction limit spot size. The hybrid system uses Fourier transforms to overcome these limitations, although more elements are required in the system. The growth of the system
size is limited by the aperture of the FT lens or that of the SELFOC lens in a compact system.
CHAPTER
FOUR

CORRELATOR SYSTEM:
PRELIMINARY SYSTEM DESIGN

• Introduction
• Review of correlators
• Correlation operation of the optical interconnect systems
• Imaging systems comparison - Extension of Sakano’s Theory
• Optical interconnection elements comparison

4.1 Introduction

After having reviewed three different optical interconnection elements in the previous chapters (ch. 2 and ch. 3), we will now carry out a critical comparison of these elements for use in an incoherent correlator under practical constraints. First, the concept of correlation and different types of correlators are reviewed in section 4.2. The new work begins in section 4.3 and continues to the end of the chapter. In section 4.3, we describe two new designs of incoherent correlator for the second of which we have obtained a patent and this latter system is selected for further designs and experimentation in the rest of this thesis. In section 4.5, we discuss the relative merits of various interconnect elements in the context of the proposed system design and commercial constraints. The summary and conclusions are given in section 4.5.4.
4.2 Review of correlators

4.2.1 Definition

The correlation operation gives a measure of similarity for two functions. The correlation function of the two signals \( f_1(x, y) \) and \( f_2(x, y) \) is given by [Yu83]:

\[
R_{12}(\alpha, \beta) = f_1(x, y) \otimes f_2(x, y) = \int \int f_1(x + \alpha, y + \beta) f_2^*(x, y) dxdy \tag{4.1}
\]

or by change of variable:

\[
R_{12}(\alpha, \beta) = \int \int f_1^*(x - \alpha, y - \beta) f_2(x, y) dxdy \tag{4.2}
\]

where \( \otimes \) denotes the correlation operation and \( f^*(x, y) \) is the complex conjugate of \( f(x, y) \). If this is compared to the convolution function which is defined as:

\[
f_1(x, y) \ast f_2(x, y) = \int \int f_1(\alpha - x, \beta - y) f_2(x, y) dxdy = \int \int f_1(x, y) f_2(\alpha - x, \beta - y) dxdy \tag{4.3}
\]

where \( \ast \) is the convolution operator, correlation can be expressed as convolution in the following way,

\[
f_1(x, y) \otimes f_2(x, y) = f_1(x, y) \ast f_2^*(-x, -y) \tag{4.4}
\]

If \( f_1(x, y) = f_2(x, y) = f(x, y) \), then eqn (4.1) becomes

\[
R(\alpha, \beta) = \int \int f(x, y) f^*(x + \alpha, y + \beta) dxdy \tag{4.5}
\]

This is known as the autocorrelation of a function and eqns (4.1) and (4.2) are called the cross-correlation of \( f_1 \) and \( f_2 \).

The Fourier transform (FT) of eqn (4.1) is known as the Correlation Theorem [Yu83]:

\[
\mathcal{F}\{R_{12}(\alpha, \beta)\} = F_1(\xi, \eta) F_2^*(\xi, \eta) \tag{4.6}
\]

where \( \mathcal{F} \) denotes the Fourier transform operator and \( F(\xi, \eta) = \mathcal{F}\{f(x, y)\} \).

From the two definitions, eqns. (4.1) & (4.6), above, optical correlation can be processed in two different domains: object-space and Fourier-space [Bar87]. The operation in object-space can be described by eqn (4.1): integrate spatially the product of \( f_1(x, y) \) with the variable shift \( (\alpha, \beta) \) of \( f_2^*(x, y) \). The Fourier-space implementation uses lenses to FT the
function \( f \) and it is multiplied it by the complex conjugate FT of \( f' \); the result is then inverse Fourier transformed, as in eqn (4.6).

### 4.2.2 Fourier-domain correlators

Fourier-domain correlation systems can be implemented by coherent and incoherent illumination and each will be considered below.

(a) Coherent correlators

1. **Vander Lugt Correlator**

A coherent correlator usually takes the form of a Fourier transform processor (fig 4.1), in which the desired correlation filter is synthesised in the Fourier plane, \( P_2 \). The input \( f(x, y) \) which is placed in plane \( P_1 \), is transformed by lens \( L_2 \) and multiplied by the filter with a complex amplitude transmittance of \( H \). The product is then Fourier transformed again by lens \( L_3 \) and the complex amplitude light field at \( P_3 \) is given by:

\[
g(\alpha, \beta) = K \iint F(\xi_1, \eta_1) H(\xi_2, \eta_2) \exp[i(\xi_1 \alpha + \eta_1 \beta)]d\xi_1 d\eta_1
\]

\[
= K \iint f(x, y) h(x - \alpha, y - \beta) dx dy
\]

\[
= K f(x, y)* h(x, y)
\]

where \( K \) is a constant. This is a convolution of \( f \) and \( h \) but if an inverted image \( h(-x, -y) \) is used, eqn (4.7) gives the correlation between \( f \) and \( h \). The fact that lens \( L_3 \) performs a forward transformation instead of the inverse FT merely causes a change of sign of the output co-ordinates, as shown in fig. 4.1.
A popular technique to synthesise complex spatial filters was proposed by Vander Lugt [Van64], using an interferometric technique (Fig. 4.2). The complex light field at the FT plane is the summation of the FT of the signal $s(x, y)$ and the reference beam $R$:

$$E(\xi, \eta) = S(\xi, \eta) + r_e e^{ia\xi}$$

where $a = f \sin \theta$, $f$ is the focal length of the transform lens, $\theta$ is the off-axis angle shown in fig. 4.2 and $r_e$ is the amplitude of the reference beam. The amplitude transmittance of the spatial filter corresponds to the intensity distribution:

$$H(\xi, \eta) = K |E(\xi, \eta)|^2$$

$$= K (r_e + |S(\xi, \eta)|^2 + 2r_e |S(\xi, \eta)| \cos(\alpha \xi + \phi(\xi, \eta)))$$

where $\phi$ is the phase function of $S(\xi, \eta)$.

Referring back to the system in fig 4.1, the complex light field at the output plane becomes

$$g(\alpha, \beta) = K \int \int F(\xi, \eta)H(\xi, \eta) \exp[i(\xi \alpha + \eta \beta)] d\xi d\eta$$

$$= K [r_e^2 f(x, y) + f(x, y)*h(x, y)*h^*(-x,-y) + r_e f(x, y)*h(x+a, y) + r_e f(x, y)*h^*(-x+a,-y)]$$

where the first two terms represent the zero-order diffraction at the origin of the output plane. The third and fourth terms are the convolution and cross-correlation results, which are diffracted to centre around $(a, 0)$ and $(-a, 0)$ respectively (Fig. 4.3). The reference beam, thus, acts as a carrier to diffract the correlation (and the convolution) to higher spatial frequencies. If the angle is sufficiently large, those terms will be separate from the d.c. term and can be detected directly. The spatial filter function is usually synthesised using an
absorbing material, such as photographic film. Note that this is a spatially invariant system as a shift in the input merely introduces a linear phase factor at the FT plane and the output is shifted correspondingly, i.e. if the input $f(x, y)$ is moved to $(-x_o, -y_o)$, the correlation term of eqn (4.10) can be shown to be $r_{K}Kf(x - x_o, y - y_o)*h^*(x - a, y)$.

Now if the impulse response of the filter $H$ is given by

$$h(x, y) = f^*(-x, -y),$$  \hspace{1cm} (4.11)

the filter is said to be matched to the particular signal $f(x, y)$. This correlation system is known as the Vander Lugt correlator or the classical matched filter system.

The early correlators used photographic films as the input and the filter. However, for real-time operations, the input and filter planes need to be implemented using spatial light modulators. Psaltis et. al. [Psa84] demonstrated the system with binary SLMs and the spatial filter is encoded as a Fourier computer generated hologram.

2. Phase-only filters

One of the disadvantages of using the above matched filter is that the optical efficiency is not very high [Hom82]. This is apparent from eqn (4.10) since only one of the four terms is useful for detection. A remedy is to use a phase-only filter instead of the original filter which is a complex function [Hom84]. As phase-only filters have no absorption, there is no energy loss. If an amplitude-only filter with the response of $1/|H|$ is placed in front of the classical matched filter, the resulting amplitude is a constant and gives rise to a phase-only filter. The $1/|H|$ filter effectively acts as a high pass filter since most of energy of a real object concentrates at low spatial frequencies. This is equivalent to edge-enhancement because edges contain high spatial frequency components. Since the high frequencies decorrelate very quickly, the correlation peak is very sharp. Consequently, the efficiency is many times higher than the conventional matched filter. A Phase-only filter can be synthesised optically but if it consists of only two phase levels [Horn85], it can be realised as a computer generated hologram or implemented with an SLM. The drawback is the introduction of quantisation noise which results in a slight reduction of efficiency. However, the ease of fabrication by photolithography and the possibility of implementing in real time are clearly more advantageous.
Pattern classification is a primary application of correlators. However, a significant drawback of using phase-only filters is the ambiguous result that can be produced. Since the high frequency components are enhanced, the tolerance regarding slight changes in scale or rotation is reduced. In fact, correlators with either matched filters or phase-only filters are often too sensitive to intraclass variations (e.g. between different typefaces of letter ‘A’) while being insensitive to interclass variations (e.g. ‘between letters ‘D’ and ‘O’) [Cau80]. In order to improve intraclass association, optimisation processes such as simulated annealing [Kim90] are employed. Circular harmonic phase-only filters with rotation-invariant properties have also been synthesised [Yau90].

3. Joint Transform correlators

Another disadvantage of using the matched filter system is the requirement of precise positioning of the filter when the space bandwidth product (SBWP) of the input is high and when the impulse response of the system is high, giving a high resolution in the system. This constraint can be removed using an alternative architecture called the optical joint transform correlator [Wea66] (Fig. 4.4). In that system, both the inverted reference image $h(-x, -y)$ and the input object $f(x,y)$ are presented at the input plane at $(-a, 0)$ and $(a, 0)$ respectively. By coherent illumination, the light field at the Fourier plane is:

$$U(\xi, \eta) = F(\xi, \eta)e^{i\phi} + H^*(\xi, \eta)e^{-i\phi}$$ (4.12)

The output intensity of this light field is detected by a square-law power detector and can be written as:
The convolution terms thus appear at (-2a, 0) and (2a, 0) in the output plane.

Other merits of this system are the avoidance of a complex filter synthesis, a lower spatial carrier frequency requirement and a generally higher output diffraction efficiency. The main disadvantage is the need to share the available SBWP in the input plane between the input object, the reference image and a separation band to avoid crosstalk. A square law converter can be a photographic plate, a LCLV or CCD camera [Jav88].

There are other coherent correlators that do not use the 4-f arrangement. Usually, these correlators are more compact, for example the 2-f correlators [Hom89]. Spatially variant correlators can also be obtained by using the Fresnel transform of the input [Dav92], as it introduces a quadratic phase factor.

(b) Incoherent correlators

A fundamental difference between coherent and incoherent correlators is that, whereas the former produces an amplitude distribution (although eventually it is the intensity which is detected), an intensity correlation is given by the latter. In other words, the need to synthesise complex amplitude filters is removed by using incoherent light. The loss of phase information means that only positive real functions can be processed. Also, devices with poorer phase uniformities such as LCTV can replace expensive SLMs [Yu86]. In addition, coherent optical processing usually suffers from optical artefact noise such as speckle generated by, for example, dust in the components. Furthermore, the use of non-coherent light provides merit in avoiding the need for a laser and subsequently the severe alignment requirements. Usually, illumination with only spatial incoherence is applied. Temporal incoherence implies polychromatic sources which can give some problems since the Fourier transform is wavelength dependent.

The incoherent matched filter system resembles that shown in Fig. 4.1 except that the collimated laser beam is replaced by an extended luminous source. Usually, the

\[
I(\xi, \eta) = \left| F(\xi, \eta) \right|^2 + \left| H(\xi, \eta) \right|^2 + F(\xi, \eta)H(\xi, \eta)e^{2\pi i(2a\xi)} + F^*(\xi, \eta)H^*(\xi, \eta)e^{-2\pi i(2a\xi)}
\] (4.13)

By coherent readout, the intensity distribution is FT to give the output light field as:

\[
g(x, y) = f(x, y) \otimes f(x, y) + h(-x, -y) \otimes h(-x, -y) + f(x, y) \ast h(x + 2a, y) + f(-x, -y) \ast h(-x - 2a, -y)
\] (4.14)
Correlator system: Preliminary system design

illumination results from a laser beam passing through a rotating diffuser. The output intensity response is now given by [Sherm83]:

\[ |g(x,y)|^2 = K|h(x,y)|^2 * |f(x,y)|^2 \]  \hspace{1cm} (4.15)

Because of the random phase distribution introduced by the diffuser, the filter does not need the crucial alignment as required in coherent systems. Low spatial frequency filtering in the incoherent correlators can be achieved by edge-enhancing images by computer prior to the correlation.

4.4.3 Real-space domain (Object-space) correlators

Correlations performed in object-space domain were achieved more than 30 years ago [Gre68]. Compared with Fourier-space correlators the designs are relatively simple and the reference masks are usually real and positive functions. This is because incoherent illumination is employed in the system. One of the most important systems is the shadow casting correlator which will be discussed in more detail later. Other processors rely on geometrical optics-based imaging such as the multichannel lenslet array processor proposed by Glaser [Gla82].

Whereas spatial filtering depends on diffraction, shadow casting systems operate best within geometrical limits [Rho81],[Gla87]. A shadow casting correlator takes direct advantage of the properties of spatially incoherent light because it requires diffusely scattered illumination to achieve the shifting of the pattern, an essential operation depicted in eqn (4.1). The light is then intensity modulated by the input function \( I_1(x_1, y_1) \) and attenuated by the function \( I_2(x_2, y_2) \) on the mask and arrives at the output plane (Fig. 4.5). Since light rays essentially travel in straight lines, those rays intersecting the output point, \( (x_3, y_3) \), must have crossed the mask plane in such a way as to form a shadow image of \( I_1 \). Using simple geometrical optics, this shadow image on the plane \( (x_2, y_2) \) is reduced by a factor of \( l_2/(l_1+l_2) \) and is displaced from the optical axis to the point \[ x_3l_1/(l_1+l_2), y_3l_1/(l_1+l_2) \]. This shadow is then multiplied by the transmittance function, \( I_2 \), and the net light distribution at the output \( (x_3, y_3) \) is given by [Mon77]:
\[ I_3(x_3, y_3) = \iiint I_1 \left( \frac{x_2 - x_1 l_1}{l_1 + l_2}, \frac{y_2 - y_1 l_1}{l_1 + l_2} \right) I_2(x_2, y_2) dx_2 dy_2 \]

\[ = \iiint I_1 \left( \frac{x_2 (l_1 + l_2) - x_1 l_1}{l_2}, \frac{y_2 (l_1 + l_2) - y_1 l_1}{l_2} \right) I_2(x_2, y_2) dx_2 dy_2 \]  

(4.16)

This represents a cross-correlation operation between \( I_1 \) and \( I_2 \).

An alternative implementation is to use collimated illumination [Rog77]. Collimated light from a single point source \((x_s, y_s)\) is modulated by the input function which is projected onto the mask and is collected by the second lens to be detected at the output point \((x_o, y_o)\) (Fig. 4.6). This system is known as "the collimated shadow casting correlator". In other words, the lenses form a image of the source plane onto the output plane. The intensity pattern at the output is, thus, a summation of all the output points:

\[ I_o(x_o, y_o) = \iiint_{\text{system aperture}} I_1(x - \frac{lx_o}{f}, y - \frac{ly_o}{f}) I_2(x, y) dx dy \]  

(4.17)

Compared with the simple shadow casting scheme which involves the demagnification of the input onto the mask, the collimated system offers the advantage of having both input and mask functions on the same scale. On the other hand, the latter system is more complex, has a larger system size and requires lenses. The lateral system dimension of the collimated correlator is also limited by the lenses’ aperture. Furthermore, the input and the source in the simple shadow casting correlator can be implemented with a single device such as an LED array [Raj93], CRT [Mat82], or liquid crystal television (LCTV) with built-in illumination [Li89],[Jut87], whereas the input in the collimated case must be
separate from the source. The mask is usually a transparency [Rog77] or LCTV panel [Tan90].

![Collimated shadow casting correlator](image)

**Fig. 4.6 Collimated shadow casting correlator [Rho81]**

Besides being correlators, shadow casting systems have many other applications. The most commonly published is their function as logic gates in digital optical computing [Ich84],[Muk93]. Other applications include fuzzy logic [Lin94], symbolic substitution logic [Lou89] and morphological transformation [Li89] for image processing.

The shadow casting systems rely on the principle of geometrical optics and diffraction effects become important when the input function has a large space bandwidth product [Rho81]. As the spatial frequency is a function of the smallest pixel size and the longitudinal distances between the planes, diffraction is important once the pixel becomes too small or the distance becomes too large, or both. Diffraction is always present where we operate our system.

In general, there are three configurations in which incoherent light can be used, depending upon the distance between the two objects to be correlated. They are the incoherent matched filtering, simple and collimated shadow casting correlators. Jutamulia [Jut85] showed that all three apparatus can be described by a general approach. The fact that the fringe visibility of each system is a function of the separation distance between the objects means that there is a certain degree of partial coherence in these experiments [Yu92].
4.3 Correlation operation of the optical interconnect systems

4.3.1 Introduction

In chapters 2 and 3, several systems implementing different interconnect elements are described. They can be placed under three categories:

(a) Spatial filtering system - which uses a Dammann or Fresnel Dammann grating
(b) Image multiplexer - which uses refractive or diffractive (FZP) microlenses
(c) Shared microlens system - which uses refractive or diffractive (FZP) microlenses

The first system is basically the configuration for the Vander Lugt correlator described in section 4.2.2. The disadvantages in system size and costs in such a system are mentioned earlier. Therefore this system is disregarded for our project for commercial reasons. The implementation of the other two systems is discussed in more details in the next two sub-sections.

4.3.2 Correlation operation of an image multiplexer

In this sub-section, we propose the use of an image multiplexer in a novel way to perform object space correlation as follow:

An image multiplexer involves multiple imaging of the input at the focal plane of the microlens array: there is one image of the input behind every microlens in the array. However, most of the microlenses are not aligned centrally with the input and are therefore in the off-axis imaging position. This gives rise to shifts of the image at the focal plane depending on the degree of off-axis. This is demonstrated in fig. 4.7a. Now, if a filter consisting of an array of the pattern for correlation is placed at the focal plane, intensity multiplication between one image of the input and one pattern at the filter takes place. Finally, all the intensity at each image position is integrated on a detector to give the correlation value at that position. The amount of shift is limited by the image extent which is a result of the aperture of the microlens. Consequently, part of the image will be cut off at the focal plane if the shift is too large compared to the aperture. The image of a point at
distance \( d_1 \) away from the lens is formed at a distance \( d_2 \) behind the lens with focal length \( f \), is given by the lens law (fig. 4.7b):

\[
\frac{1}{d_1} + \frac{1}{d_2} = \frac{1}{f} \quad \text{and} \quad \frac{x_1}{d_1} = \frac{x_2}{d_2} \quad (4.18)
\]

where \( x_1 \) and \( x_2 \) are the off-axis shift at the object and image. The shift of the image is, therefore, given by:

\[
x_2 = x_1 \cdot \frac{d_2}{d_1} = \frac{f}{d_1 - f} \quad (4.19)
\]

But \( x_1 = d_1 \tan \theta \), therefore

\[
x_2 = d_1 \tan \theta \cdot \frac{f}{d_1 - f} \quad (4.20)
\]

The shift is a function of the object distance \( d_1 \) and the displacement of the microlens from the central axis, \( x_1 \). The intensity in 1-D at each pixel \( I_p \) at a distance \( x_3 \) from the centre of the image is therefore given by:

\[
I_p(x_3) = I_1[-M(x_3 - x_2)] I_2(x_3) \quad (4.21)
\]

where \( M \) is the magnification factor given by \( d_2 / d_1 \). The correlation is a summation of the intensity at all the pixels within one image. Obviously this can be extended to 2-D:
Since the correlation is provided by the discrete translation of the images, the resolution of the correlator is then dependent upon the image shift \( x_2 \) and also the number of pixels that can be placed within an image of the microlens. This determines the space bandwidth product of the system.

### 4.3.3 Correlation operation of the shared microlens system

In this section we propose to use the shared microlens system in a novel way to perform correlations as follows:

Consider the operation of the shared microlens system (Fig. 4.8), the input intensity at \((m, n)\) is imaged by a microlens at \((m, n)\) to the output point \((m, n)\) and by the next microlens at \((m+1, n+1)\) to the corresponding output at \((m+1, n+1)\). At the same time, the output intensity at any point \((p, q)\) will be a sum of the contributions from each input pixel intensity \(O(m, n)\), for example, \(O(1, 1)\) is imaged by the lens at \((1+p, 1+q)\). If a mask, with a transmittance function, \(T(m, n)\), is inserted at the microlens plane, with each pixel having the same size as the pitch of the microlens, the output is given by:

\[
I(p, q) = O(0, 0)T(p, q) + \ldots + O(m, n)T(m+p, n+q) + \ldots + O(N, N)T(N+p, N+q)
\]

\[
= \sum_{m=1}^{N} \sum_{n=1}^{N} O(m, n)T(m+p, n+q)
\]

\[
(4.23)
\]

![Fig. 4.8 Correlation operation of shared microlens correlator](image)
Correlator system: Preliminary system design

Clearly, this is the discrete form of eqn (4.1) and is a cross-correlation of the input and the mask at the microlens plane. As the summation is from 1 to N only, it uses only a quadrant of the whole microlens plane which has \((2N-1)^2\) number of elements. This means that four cross-correlations can be achieved in parallel at the same time. In a 2f:2f system, the size of the microlens and, thus, the size of the pixel of the mask is half of that in the input plane. This resembles the simple shadow casting correlator except that each mask pixel has a corresponding microlens. We have patented this new configuration in collaboration with Sharp Laboratories of Europe.

4.4 Imaging systems comparison - Extension of Sakano’s Theory

After describing how the two imaging systems can be used as correlators, we now have to choose between these two systems remembering that the system chosen must meet commercial constraints, the most important (to our sponsor) being system size. In this section, Sakano’s theory [Sak90] introduced in chapter 3 will be extended for the analysis of the two systems with respect to the system size.

Sakano et. al. [Sak90] investigated both image multiplexer and shared microlens systems in ideal and optimum conditions. They argue that the system design is restricted by astigmatic aberrations and the diffraction limited spot size. We have re-expressed their equations in a useful format that involve parameters such as system length \((L)\) and pitch of the microlens \((d)\) (Appendix II). In this way, we can assess the size of a system by substituting practical values into the equations. First, if the input and output planes are close to the microlens array, then the off-axis angles for the furthest elements at the corners of the arrays are large which give rise to astigmatism (fig. 4.9). Astigmatism increases in a quadratic manner with the off-axis angle and cannot be easily compensated. This give the lower limit to the system length. The upper limit is determined by the size of the diffraction limited spot on the output element which increases with system length and depends on the focal length, \(f\), of the microlens. The limit is assumed to occur when the Airy disc is the same size as the pixel on the output plane giving crosstalk of 16\% of the energy in the focal sidelobes which falls on adjacent pixels. Whereas the distance ratio \(a\) in the image
Correlator system: Preliminary system design

multiplexer is clearly $N/f f$, that in the shared microlens system is variable as long as lens law is satisfied.

![Diagram](image)

**Astigmatism:** $x \propto \theta^2$

**Diffraction:** $x = \frac{1.22\lambda f}{d}$

Fig. 4.9 Two limits to the system size:

(a) Astigmatism: as the system size is reduced

(b) Diffraction: as the system size increases

For the shared microlens system, by the lens law,

$$\frac{1}{z_2} = \frac{1}{f} - \frac{1}{z_1} = \frac{z_1 - f}{fz_1}$$

(4.24)

and the system size is given by

$$L = z_1 + z_2 = \frac{z_1^2}{z_1 - f}$$

(4.25)

Differentiating (4.25) gives:

$$\frac{dL}{dz_1} = \frac{2z_1}{z_1 - f} - \frac{z_1^2}{(z_1 - f)^2} = \frac{z_1(z_1 - 2f)}{(z_1 - f)^2}$$

(4.26)

The minimum system length is given when $dL/dz_1 = 0$, i.e. when

$$z_1 = 2f = z_2$$

(4.27)

which is a unity magnification imaging system. We also know that this introduces less aberrations as well. By simplifying Sakano's Theory (Appendix I), the two inequalities
which \( L \) must satisfy are given for image multiplexer and shared microlens systems respectively, as follows:

Image multiplexer:

\[
\frac{d^2}{2.44\lambda} \geq L \geq 2 N^2 d I^1
\]  \tag{4.28}

Share microlens system:

\[
\frac{d^2}{2.44\lambda} \geq L \geq 2Nd I^1
\]  \tag{4.29}

where \( I \) is the dimensionless strength of the astigmatic aberration. If the microlenses are thin and unmounted on a substrate then \( I \) is given by \((3n+1)/4n\) [Sak90] where \( n \) is the refractive index of the photoresist.

---

The diffraction limit is the same for both systems, which gives a parabolic function of \( d \) in Fig. 4.10. The astigmatic limit is a series of straight lines in these plots, with slopes determined by the number of elements \( N \) linearly (\( N^2 \) in total) and the strength of
astigmatism. Clearly, the straight lines in the image multiplexer (Fig. 4.10a) have greater
gradients than those of the shared microlens system (Fig. 4.10b) because of the formers' N\(^{3/2}\) dependence on N. For example, if N = 32, then a practical system design would be possible in the shaded region in the plots to satisfy the inequalities. We can see that for a given microlens pitch (e.g. 250 \(\mu\)m), the system length allowed for the shared microlens system is smaller. Hence, we can conclude from this idealistic analysis, that more elements can be connected in the shared thin microlens system using a smaller microlens pitch and in an overall more compact system and, therefore, we chose to use the shared microlens system for further more detailed system analysis and experimentation for use in an incoherent correlator.

4.5 Optical interconnection elements comparison

4.5.1 Introduction

In the previous section, two different types of system architecture were analysed and compared with respect to system parameters. Now, the merits of the different interconnection elements which are used in these systems are discussed in more detail in the context of our chosen system design (the shared microlens system) and with respect to commercial constraints. Before the comparison, the characteristics of these elements are first summarised in the next sub-section.

4.5.2 Optical interconnection elements summary

The three classes of optical element most suitable for the implementation of optical routing and fanout are Dammann-like gratings, Fresnel zone plates and microlenses. Each class has many different structures and so we make an optimum choice of one example from each and compare them in the next section. The criteria for choosing these examples are as follows:

(a) large array generation - higher parallelism and SBWP which gives improved processing power
(b) Ease of manufacture - can either be fabricated in-house or cheap to purchase
(c) Design flexibility - to suit the system requirements
(d) High efficiency - in terms of power
(e) Good uniformity

The choices which satisfy the above conditions in each class are described as follows:

(a) Dammann-like phase gratings

Binary separable Dammann gratings are chosen because these gratings can generate a large array of spots (201 x 201 experimentally), they are easy to design (1001 x 1001 array generator has been designed) and they are simple to fabricate. The efficiency is about 65% and the uniformity is about ±8%.

(b) Fresnel zone plate arrays

Simple spatial multiplexing of Fresnel zone plates is the best alternative to a Dammann grating for diffractive array generation. The Fresnel zone plates can have multi-phase levels or can be blazed and an 8-level structure with 91% efficiency is the most appropriate. The fabrication of an array of 245 x 245 Fresnel zone plates has been reported.

(c) Refractive microlenses

Finally, photoresist reflow microlenses are the best candidates out of the various types of refractive optical elements. This is due to the versatility in design and a simple and easily accessible fabrication technique. Arrays of almost any size can be fabricated by this process. The limitations are the flexibility of the shape of the lens (which is formed by surface tension) and the size of the wafer that can be processed uniformly.
4.5.3 Comparison between the CGHs and microlenses

The three choices are now compared and contrasted in the following parameters to find out which is most suitable for our system:

(a) Design

The design of microlenses (both refractive and diffractive Fresnel zone plates) are simple although Fresnel zone plates require a small computer program to generate them. On the other hand, the computer generation of a Dammann grating is more complicated. First, the parameters such as array size, efficiency and uniformity of spots must be specified. Then, root-finding and optimisation algorithms are used to generate the desired grating. This involves more computational time and cost.

(b) Imaging properties

Both types of microlens behave like conventional lenses in terms of imaging. However, the Dammann grating is designed to generate an array of spots and in general, conventional imaging cannot be implemented by this element.

(c) Efficiency

Refractive microlenses have a collection efficiency of 100 % if there is no dead space between the lenses ignoring Fresnel reflections. In the case of the binary, separable, Dammann grating, the diffraction efficiency is only 65%. Higher efficiency is possible by making a multi-phase structure. An 8-level Fresnel zone plate has an experimental efficiency of 91%. There is also reflection loss at the air-photoresist interface for these elements. Application of anti-reflection coating on both computer generated holograms is straight forward because of the flat profiles. However, it is more difficult to deposit a uniform layer on the surface-relief microlenses.

(d) Focal length and system size

The photoresist microlenses cannot have very long focal lengths because of the inability of the surface tension to pull a thin layer into a spherical shape. On the other hand, the Fresnel zone plates are best designed to have longer focal lengths because short focal lengths would
require very fine structures in FZPs. Therefore, the system is more compact if a refractive microlens array is used. The Dammann grating requires external components in a system (two Fourier transform lenses), thus, making the system very bulky. The exceptional case is the Fresnel-Damman grating where the FT lens is incorporated in the design.

(e) Spot size
Both types of microlens can give diffraction limited spot sizes if there is little fabrication error. The size of the spots generated by a Damman grating is related to the overall size of the element assuming accurate fabrication and to the FT lenses' numerical aperture.

(f) Aberrations
The photoresist microlenses are mainly used for on-axis applications and off-axis uses often bring about large aberrations due primarily to field curvature. This is partly due to the relatively short focal lengths (low f-nos) inherent in these microlenses. Fresnel CGHs, on the contrary, can be designed to compensate for aberrations. Aberrations in a Damman system comes from the FT lenses and fabrication tolerances.

(g) Fabrication
The fabrication techniques for refractive microlenses and Damman gratings are the standard photolithographic processes. There are certain problems with the fabrication of Damman gratings. It is difficult to make accurate structures in glass whereas structures fabricated in photoresist may suffer from variations in refractive index and poor durability. The fabrication of multi-level Fresnel zone plates requires either the alignment of the masks or direct electron beam writing on the substrate. In all cases, once the master is made, copies can be reproduced very cheaply by moulding and embossing.

In addition, photoresist microlenses can be used in both coherent and incoherent systems whereas computer generated holograms are designed for a specific wavelength only. Refractive microlenses have, in general, less dispersion and higher numeral aperture which means smaller spot size. Fresnel zone plates can have 100% fill factor but the size is limited by the minimum feature size. In addition, the FZPs only work as ideal optical elements over small bandwidths. In a Damman grating, uniformity of spots is independent
of the profile of the incident beam. The advantages and disadvantages of the three optical elements are summarised in table 4.1.

<table>
<thead>
<tr>
<th>Optical elements</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microlenses</td>
<td>- Less Dispersion</td>
<td>- Mainly on axis</td>
</tr>
<tr>
<td></td>
<td>- Higher Numerical apertures</td>
<td>- Power lost through dead space</td>
</tr>
<tr>
<td></td>
<td>- Simple to fabricate</td>
<td>(circular aperture)</td>
</tr>
<tr>
<td></td>
<td>- Can work in coherent and incoherent systems</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Compact</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Efficient (100% if no dead space)</td>
<td></td>
</tr>
<tr>
<td>Fresnel zone plates</td>
<td>- Can design for off-axis (elliptical)</td>
<td>- Poor efficiency (binary) or complicated fabrication</td>
</tr>
<tr>
<td></td>
<td>- Can design to compensate for aberrations</td>
<td>- Size is limited by smallest feature size</td>
</tr>
<tr>
<td></td>
<td>- 100% fill factor</td>
<td>- Works as an ideal optical element only over small bandwidth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Use efficiently in coherent system</td>
</tr>
<tr>
<td>Dammann-like gratings</td>
<td>- Good for array generation</td>
<td>- Require Fourier lenses</td>
</tr>
<tr>
<td></td>
<td>- Uniformity of spots is independent of profile of the incident beam</td>
<td>- Need lengthy computation for simulated annealing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- In glass - difficult to make accurate structures</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- In photoresist - lacks of uniformity in refractive index and poor durability</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Use in coherent system only</td>
</tr>
</tbody>
</table>

Table 4.1 Comparison between computer generated holograms and microlenses

4.5.4 Conclusions

We conclude that if someway can be formed to reduce the aberration of refractive microlenses for off-axis operation, they would make ideal interconnection elements for our system. In fact we have found a way to do this by lengthening the focal length sufficiently as described in chapter 7. However, it is worth noting that perhaps the most overriding
argument in favour of them was that our commercial sponsor wanted to gain knowledge of how to fabricate them for other in-house applications and they had not been sufficiently impressed after testing several Dammann gratings. Cost will be cheap if manufacture in-house.
5.1 Introduction

The performance of a shared microlens correlator is determined by the space-bandwidth product (SBWP), which in turn depends upon the minimum pixel size at the input / output. In a lens system, the resolution at the output plane determines the pixel size. The smaller the pixel size, the larger the SBWP. In this chapter, the theoretical limits of the systems design are examined. First, a diffraction theory is chosen in section 5.2 and is applied to an aperture. This aperture can be either a microlens in our system (section 5.3) or a pinhole in an equivalent pinhole system without microlenses which was used in some earlier work (section 5.4). Both systems are analysed and are compared in detail in terms of resolution, efficiency and size (section 5.5). In addition, the effect of off-axis aberrations such as field curvature and astigmatism on the systems are examined (section 5.6). Finally, the chapter is summarised and concluded in section 5.7.
5.2 Diffraction theory

5.2.1 Justification for the use of Rayleigh-Sommerfeld diffraction theory

The shared microlens system which we will analyse in this chapter has apertures of sizes about 165 µm and focal lengths about 60 times that size. The size is determined by the pitch of our SLM and the focal length must be 60 times since such a long distance from the apertures, the foci already lie in the far-field (or Fraunhofer) region and the spot size is consequently diffraction limited. This implies that ray tracing is not the correct technique to solve the problem. Although the angular spectrum of plane waves can also be used, diffraction theories provide a more analytical approach whereas the former technique involves repetition of tasks which can best be carried out using a computer. Although the electromagnetic fields are vectors themselves, calculations using the full vector representations is not essential since the objects of interest are usually substantially larger than a few wavelengths and the distances involved are many times the wavelength. In this case, a scalar theory is usually sufficient to produce a reasonably good answer. The Rayleigh-Sommerfeld theory [Goo68] is chosen from the several scalar theories which exist because it does not have the inconsistencies of the Fresnel-Kirchoff Theory and is rigorous and general enough to include diffraction effects in both near and far fields.

5.2.2 Diffraction at an aperture

Using the Rayleigh-Sommerfeld scalar diffraction theory, the complex amplitude, \( U \), at the output plane \((x_3, y_3)\) due to diffraction at the aperture \( S \) at plane 2, is given by (Fig 5.1) [Goo68]:

\[
U(x_3, y_3) = \frac{A}{j\lambda} \int_S \int U(x_2, y_2) \frac{e^{jkr_3}}{r_3} \left(1 - \frac{1}{jk r_3}\right) \cos(n \cdot r_{23}) dS
\]

(5.1)

where \( A \) is a constant, \( S \) is the area of the aperture and \( dS = dx_2 dy_2 \) is the area element in the aperture, \( r_{23} \) is the distance between a point at the aperture \((x_2, y_2)\) and the output point. This aperture can be the microlens in the shared microlens system or a pinhole in a pinhole system.
Correlator system: theoretical Analysis

Fig. 5.1 System layout for the Rayleigh-Sommerfeld diffraction theory [Goo68]

$U(x_2, y_2)$ is the complex amplitude of the wave arriving at the aperture. If we make the approximations that $r_{23} \gg \lambda$ and that the angles are small, the equation simplifies to:

$$U(x_3, y_3) = A \int \int U(x_2, y_2) e^{i k r_{23}} dS$$  \hspace{1cm} (5.2)

In reality, $r_{23}$ is of the order of cm and the angle between the normal and $r_{23}$ will be a fraction of a degree (from $\sin^{-1} \left( \frac{100 \mu m}{1 \text{cm}} \right)$ on axis) since $r_{23}$ is much larger than the size of the aperture. Therefore, these approximations are justified.

A further approximation comes from the binomial expansion of $r_{23}$ and is based on the fact that $z_2$ is usually much larger than dimensions in either $x$- or $y$- directions.

$$r_{23} = \sqrt{z_2^2 + (x_2 - x_3)^2 + (y_2 - y_3)^2} = z_2 \left[ 1 + \frac{1}{2} \left( \frac{x_2 - x_3}{z_2} \right)^2 + \frac{1}{2} \left( \frac{y_2 - y_3}{z_2} \right)^2 \right]$$  \hspace{1cm} (5.3)

Substituting this into (5.2) gives:

$$U(x_3, y_3) = \frac{A e^{i k c}}{\sqrt{\lambda z_1}} \int \int U(x_2, y_2) \exp \left( \frac{jk}{2z_2} \left[ x_2^2 + y_2^2 - 2(x_2 x_3 + y_2 y_3) \right] \right) dS$$  \hspace{1cm} (5.4)

where $c = z_2 + \frac{x_3^2 + y_3^2}{2z_2}$ \hspace{1cm} (5.5)

This equation can be used to analyse a general circular aperture and can be applied to a microlens or a pinhole.
5.3 Diffraction theory for the shared microlens system

5.3.1 Introduction

In this section, the equation derived from the Rayleigh-Sommerfeld theory in the above section is used to analyse a microlens in a shared microlens system. The resolution of the system is also determined for real, practical values.

5.3.2 Derivation of the resolution for a microlens

Since the amplitude at the microlens is given by the product of the complex amplitude of the spherical wave from the input point \((x_i, y_i)\) and the phase change through the thin microlens:

\[
U(x_2, y_2) = \frac{1}{r_{21}} \exp(j k r_{21}) \exp\left(\frac{j k}{2 f} (x_1^2 + y_1^2)\right).
\]  

(5.6)

Substituting eqn (5.6) into (5.5), the resultant complex amplitude at the output plane is:

\[
U(x_3, y_3) = \frac{A e^{jkb}}{j \lambda z_1 z_2} \int_S \exp\left(\frac{j k}{2} \left(\frac{x_3^2 + y_3^2}{z_1^2} + \frac{1}{z_2} + \frac{1}{z_1} - \frac{1}{f}\right) - 2x_3 \left(\frac{x_1}{z_1} + \frac{x_2}{z_2}\right) - 2y_3 \left(\frac{y_1}{z_1} + \frac{y_2}{z_2}\right)\right) dS
\]

(5.7)

\[b = \left(\frac{1}{z_1} + \frac{1}{z_2} + \frac{x_1^2 + y_1^2}{2z_1} + \frac{x_2^2 + y_2^2}{2z_2}\right)
\]

(5.8)

In an imaging system, where the lens focuses the light onto the output \((x_3, y_3)\), the lens law must be satisfied, i.e. \(\frac{1}{z_1} + \frac{1}{z_2} = \frac{1}{f}\) to give the smallest spot size [Goo68]. The quadratic phase term in the integral disappears and eqn (4.27) is now:

\[
U(x_3, y_3) = \frac{A e^{jkb}}{j \lambda z_1 z_2} \int_S \exp\left[-j k \left(\frac{x_2}{z_1} + \frac{x_1}{z_2}\right) + y_2 \left(\frac{y_1}{z_1} + \frac{y_2}{z_2}\right)\right] dS
\]

(5.9)

by defining \(M\) as the magnification \((M = z_2/z_1)\). This equation is, in fact, the Fraunhofer diffraction formula and the pattern observed at the output will be of a far-field nature.
assuming that it is imaged. By solving the equation (5.9) (in Appendix II), the intensity profile takes the form of a Bessel function, giving rise to the usual Airy pattern (Fig. 5.2):

\[
I(x_1, y_1) = UU^* = \left( \frac{2\pi a^2 A}{\lambda z_1 z_2} \right)^2 \frac{J_1^2(kaq / z_2)}{(kaq / z_2)^2}
\]

(5.10)

where \(a\) - radius of the microlens
\[q = \sqrt{(Mx_1 + x_1)^2 + (My_1 + y_1)^2}\]

\(J_1\) is the Bessel function of the first kind and is of order one

The distance from the centre of the focal point to the first null is defined as the radius of the Airy disc [Hec87]:

\[q_o = \sqrt{(x_1 + Mx_1)^2 + (y_1 + My_1)^2} = \frac{1.22z_2\lambda}{2a}\]

(5.11)

The diffraction pattern is now centred at \((x_3 = -Mx_1, y_3 = -My_1)\) as expected in an imaging system. The resolution, according to Rayleigh’s Criterion, is the diameter of the Airy disc, i.e. \(1.22z_2\lambda/\lambda\), although sometimes it may be more appropriate to quote the Full width half maximum (FWHM) or the -3dB point since a conventional detector is usually a thresholding device.
5.3.3 Resolution of the shared microlens correlator under LED illumination

In this section, we substitute real, practical parameters into the theory derived above to obtain some implications for our system. In the calculations below, the LED has a central wavelength of 660 nm and the microlens diameter matches the pitch of the spatial light modulator (2a = 165 µm). The separations between the three planes (z₁, z₂) are equal to minimise system length and aberrations and are set to 2.06 cm and the focal length can then be calculated to be 1.03 cm. These values were chosen to match the focal lengths of the new long focal length microlenses we fabricate in chapter 7.

Monochromatic light has been assumed in the theory so far and the effect of an LED will now be investigated. An LED has a typical linewidth of about 50 nm and by summation of the effect of every wavelength within this bandwidth, the resultant graph is shown in fig. 5.3. The resulting profile is virtually the same as that in fig. 5.2. As the LED linewidth is relatively narrow, it has little effect on the intensity profile. The diameter at Rayleigh’s resolution limit is found to be 200 µm and the -3dB point is 85 µm.

![Intensity profiles of output spot using LED source in linear and logarithmic scales](attachment:image.jpg)

The efficiency of the shared microlens system is defined to be the total incident energy contained within the central lobe of the intensity profile transmitted by a single microlens and is given by (Appendix II):
\[ \eta(q_o) = 1 - J_0^1 \left( \frac{kaq_o}{z_2} \right) - J_1^2 \left( \frac{kaq_o}{z_2} \right) \]
\[ = 1 - J_0^1 (1.22\pi) \]
\[ = 0.838 \] (5.12)

where \( q_o \) is the radial distance from the system axis on the output plane. This means that 83.8% of the energy is contained within the output pixel if its size is the same as the Airy disc (fig. 5.4) and 16.2% contributes to the cross-talk. Please note that the Fresnel reflection and absorption losses are neglected in this discussion. In addition, the output pixels are assumed to be circular although in general they are square.

![Image of encircled energy plot](image)

Fig. 5.4 The encircled energy plot (energy enclosed within a circle of \( q \)) versus radius of the spot

### 5.4 Diffraction theory for the pinhole system

#### 5.4.1 Introduction

The apertures of a shadow casting correlator are usually square or rectangular shape pixels. If we approximate this pixel with a pinhole of diameter the same as the pixel, we neglect 22% of the area and assume this to be opaque. Circular symmetry is used in this case for the convenience of comparison with the shared microlens system.
5.4.2 Diffraction theory for the pinhole system

The complex amplitude at the pinhole is simply that of a spherical point source from the plane \((x_1, y_1)\):

\[
U(x_2, y_2) = \frac{1}{r_{21}} e^{j\kappa_{21}}
\]

and substituting this into eqn (5.5), we get:

\[
U(x_3, y_3) = \frac{A e^{j\delta_3}}{j\lambda z_1 z_2} \int \exp \left[ \frac{j\pi}{\lambda} \left( x_1^2 + y_1^2 \right) \left( \frac{1}{z_1} + \frac{1}{z_2} \right) - 2x_2 \left( \frac{x_3}{z_1} + \frac{x_3}{z_2} \right) - 2y_2 \left( \frac{y_3}{z_1} + \frac{y_3}{z_2} \right) \right] ds
\]

\[
= \frac{A e^{j\delta_3}}{j\lambda z_1 z_2} \int \exp \left[ \frac{jk}{2z_o} \left( x_1^2 + y_1^2 \right) \right] \exp \left[ -\frac{jk}{z_o} \left( x_2 x_3 + y_2 y_3 \right) \right] ds
\]

where \(x_o = z_o \left( \frac{x_1}{z_1} + \frac{x_3}{z_2} \right), y_o = z_o \left( \frac{y_1}{z_1} + \frac{y_3}{z_2} \right), z_o = \left( \frac{1}{z_1} + \frac{1}{z_2} \right)^{-1}
\]

and \(\delta_3\) is defined in eqn (5.8). Note that \(z_o\) is defined in a similar way to \(\delta\) in the lens law for the shared microlens system and this means that \(z_o\) defines the geometry condition of the system. We call \(z_o\) the effective focal length in the pinhole system. This final form, equation (5.14), is a Fresnel diffraction formula because of the presence of a quadratic phase factor. Solving (5.14) gives the following integral:

\[
U(x_3, y_3) = \frac{A e^{j\delta_3}}{j\lambda z_1 z_2} \int_0^\infty r \exp \left( \frac{jkr^2}{2z_o} \right) J_\nu \left( \frac{kpr}{z_o} \right) dr
\]

where

\[
p = \sqrt{x_o^2 + y_o^2} = z_o \sqrt{\left( \frac{x_1}{z_1} + \frac{x_3}{z_2} \right)^2 + \left( \frac{y_1}{z_1} + \frac{y_3}{z_2} \right)^2} = \frac{z_o}{z_2} \sqrt{(Mx_1 + x_3)^2 + (My_1 + y_3)^2} = \frac{z_o}{z_2} q
\]

using the same definitions of \(M\) and \(q\) as before. \(p\) is therefore related to the size of the output pixel \(q\) with a scaling factor depending on the geometry of the system \(z_o/z_2\).
5.4.3 Variation of intensity of diffraction with distance away from the pinhole

In the following calculations, the same parameters used in the microlens case are now applied to the pinhole system, i.e. the LED has a central wavelength ($\lambda$) of 660 nm, the pinhole diameter ($2a$) is 165 µm, for the sake of comparison.

The intensity at the output is simply $|U(x_1, y_1)|^2$ and the on-axis intensity (when $x_1 = y_1 = 0, x_3 = y_3 = 0$) is given by:

$$I(0,0) = \frac{2A^2}{(z_1 + z_2)^2} \left[ 1 - \cos \left( \frac{ka^2}{2z_o} \right) \right]$$

(5.18)

Now the intensity can be expressed as a function of $z_2$, since it is related to $z_o$ by eqn (5.15) for constant $z_1$. In most cases, we will assume that the distances $z_1$ and $z_2$ are equal to twice the distance of $z_o$ as in the shared microlens system it was $2f$. This means that $I(0,0)$ will be maximum at:

$$\cos \left( \frac{ka^2}{2z_o} \right) = -1 \quad \Rightarrow \quad z_o = \frac{a^2}{(2m+1)\lambda}$$

(5.19)

$$\Rightarrow \quad z_2 = \frac{2a^2}{(2m+1)\lambda} \quad \text{for} \quad z_1 = z_2$$

Fig 5.5a shows the variation of the on-axis intensity with distance $z_2$. The maxima (or focal points) closer to the aperture are less efficient than those further away since only those rays at steep angles can reach the former. However, this has not been shown in the graph because the obliquity factor, $\cos(n, \varphi)$ in eqn (5.1), was removed from earlier calculations.

In the presence of LED illumination, having a linewidth of 50 nm, the closer foci have reduced intensity whereas the furthest one largely remains unchanged (Fig. 5.5b). This is due to the beating of the intensity patterns calculated at different wavelengths.

In order to determine the intensity profile at a particular distance $z_2$ away from the pinhole, the equation (5.16) needs to be solved by a numerical technique. We use Simpson's rule for calculating the integral because it is more accurate than the trapezium rule. The intensity profile at various distances of $z_2$ are examined in figs. 5.4c-f as a function of radial distance $q$ in the output plane. We observe that the foci (maxima) closer to the pinhole are not very efficient. This is due to the presence of large sidelobes in the closer foci, which is a characteristic of Fresnel diffraction patterns (fig.5.5c). The furthest focus (or maximum) is the most efficient as the sidelobes become insignificant (fig. 5.5e).
Beyond this, the efficiency gradually fades away with distance (fig. 5.4f). Then, the radial spot size, which is defined as the radial distance to the first null in the profile, increases with the distance $z_2$.

![Intensity profiles of circular pinhole diffraction](image)

**Fig. 5.5** Intensity profiles of circular pinhole diffraction (pinhole diameter = 165 μm):

- On-axis intensity profile with $z_2$ in (a) monochromatic and (b) LED illumination.
- Intensity profiles at (c) second furthest focus, (d) half power point between second furthest and furthest foci, (e) furthest focus and (f) half power point beyond the furthest focus.

The furthest focus (when $z_2 = 2a^2 / \lambda$), which is the most efficient, is chosen to be the working point and will be used in all the discussions later. A closer examination shows
that this point is the onset of Fraunhofer diffraction since the far-field region is defined to be [Goo68]:

\[ z_2 \gg \frac{k(x^2 + y^2)_{\text{max}}}{\pi} \]
\[ \gg \frac{2a^2}{\lambda} \quad (\text{= 2.06 cm in this case}) \]

At distances greater than that point, the profile simply broadens with distance \( z_2 \). There is no change in the overall shape, unlike the Fresnel diffraction which has many foci.

### 5.4.4 Resolution of the pinhole system

The intensity profile of the furthest focus can be found by substituting the value of \( z_o \) (\( = a^2 / \lambda \)) at the appropriate peak into (5.16) and taking the modulus:

\[
I(x, y) = \left( \frac{Ak}{z_1z_2} \right)^2 \left[ \int_0^a r \exp\left( \frac{j\pi r^2}{a^2} \right) J_0\left( \frac{2\pi rp}{a^2} \right) dr \right]^2
\]
\[
= \left( \frac{Ak}{z_1z_2} \right)^2 \left[ J_0^2\left( \frac{2\pi p}{a} \right) + 1 - \frac{2\pi p}{a} \int_0^a \exp\left( \frac{j\pi r^2}{a^2} \right) J_1\left( \frac{2\pi rp}{a^2} \right) dr \right]^2
\]  
\[ (5.21) \]

If \( p \ll a \), this can be further approximated by dropping the integral involving \( J_1 \) and the resulting intensity profile becomes:

\[
I(x, y) = \left( \frac{A}{z_1 + z_2} \right)^2 \left[ J_0^2\left( \frac{2\pi p}{a} \right) + 2J_1\left( \frac{2\pi p}{a} \right) + 1 \right] \quad \text{(for } p \ll a \)  
\[ (5.22) \]

As can be seen in fig. 5.6, the approximation (5.22) to (5.21) is valid provided \( q \ll a \). \( q \) is the radius of the spot on the output plane and is related to \( p \) by the ratio \( z_2/z_o \) in eqn (5.17). It should be noted that in eqn (5.21) the shape of the intensity profile (fig. 5.6) is independent of the wavelength and \( z_o \). By comparing the two curves, the first minimum of the approximated profile coincides with the point of inflexion at the flat region next to the main lobe in the actual profile. If this point is defined to be the resolution limit, it is given by:

\[
\frac{2z_oq_o\pi}{z_2a} = 3.83 \quad \text{or} \quad q_o = 0.61a(1 + M)
\]
\[ (5.23) \]

or \( \sqrt{(Mx_1 + x_2)^2 + (My_1 + y_2)^2} = 0.61a(1 + M) \)
This is found to be about 100 \( \mu \text{m} \) for \( 2a = 165 \, \mu \text{m} \) and \( M = 1 \). The output diffraction pattern is centred at the \( (x_3 = -Mx_1, y_3 = -My_1) \) as before and the resolution is the diameter to the first turning point (the point of inflexion in this case), i.e. \( 1.22a(1+M) \). In this case when \( M = 1 \), the resolution is \( 2.44a \) (\( = 200 \, \mu \text{m} \)) which means that this effective focal spot is a little larger than the pinhole.

![Graph](image)

**Fig. 5.6** Intensity profiles of the pinhole diffraction at \( z_o = a^2 / \lambda \) by numerical integration and the approximation when \( q \ll a \)

![Graph](image)

**Fig. 5.7** Intensity profiles of the pinhole system at the output in (a) linear and (b) log scales
The effect of LED linewidth is minimal and the intensity profile is shown in fig. 5.7. The half width at the half power point is found to be 45 μm.

The efficiency of a pinhole system can be found in a similar way to the shared microlens system:

$$
\eta(q_o) = \frac{2k^2q_o}{a^2q_2^2} \int_0^{q_o} qI_1(q) dq
$$

where

$$
I_1(q) = \left| \int_0^a r \exp \left( \frac{j\pi r^2}{a^2} \right) J_0 \left( \frac{2\pi rq}{a^2(1 + M)} \right) dr \right|^2
$$

This integral can be evaluated numerically by Simpson's Rule and the results are shown in fig. 5.8. The fraction of the energy encircled within the resolution limit defined above and the half power point are 0.48 and 0.23 respectively. This implies either a larger pixel size is required at the output to receive more energy or a higher energy source is needed at the input.

![Graph](image)

Fig. 5.8 The percentage of focal energy enclosed with the radial distance, q, of the spot

### 5.4.5 Resolution of the simple shadow casting system

Here, we investigate the resolution of the simple shadow casting correlator and compare it with that of the pinhole system. The shadow casting correlator is based on the principles of geometrical optics. For the law of geometrical optics to hold [Tan83]:

- $\lambda = 660 \text{ nm}$
- $2a = 165 \mu\text{m}$
- $z_2 = 2z_1 = 2.06 \text{ cm}$
Correlator system: theoretical Analysis

\[
z_2 \ll \frac{4a^2}{\lambda}
\]  \hspace{1cm} (5.25)

This equation bears resemblance to the Fresnel diffraction limit. In this case, the spread of the light due to the pinhole is (fig. 5.9):

\[
q = 2a \left( \frac{z_1 + z_2}{z_1} \right) = 2a(1 + M)
\]  \hspace{1cm} (5.26)

If the plane separations are assumed to be equal, the magnification \( M \) is 1 and the spot size is given by \( 4a \) or \( 330 \) \( \mu \)m in this case. This is significantly larger than that for either a microlens or a pinhole. However, all of the energy is assumed to be detected at the output within this spot size \( q \). This assumption is true only when the distance is so small that even Fresnel diffraction can be neglected.

5.5 Comparison and discussions

The assumptions made in the last two sections are:

1. A red LED is used for the illumination, with central wavelength, \( \lambda \), of 660 nm and linewidth (\( \Delta \lambda \)) of 50 nm.
2. The diameters of the microlens and the pinhole (2a) are both set to match the pixel pitch of the spatial light modulator, 165 \( \mu \)m.
3. The distance between the input (\( z_1 \)) and filter plane and that between the filter plane and the output plane (\( z_2 \)) are equal and are given by:
   \[
z_1 = z_2 = \frac{2a}{\lambda} = 2.06 \text{ cm}
\]
The magnification \( M = z_2/z_1 \) is therefore 1 and both systems have the same length.

4. It follows that the focal length for the shared microlens system \( f \) and the effective focal length for the pinhole system \( z_o \) is:

\[
f = z_o = z_2/2 = 1.03 \text{ cm}
\]

The results obtained are given in table 5.1. The working point \( z_2 \) chosen in the pinhole system is in the Fraunhofer diffraction region and since Fraunhofer diffraction always occurs shared microlens system, the resolution limit of the pinhole system is same as that in the shared microlens system (Fig. 5.10). However, the lens system is about twice as efficient as the pinhole system since more light is collected within the central lobe. For the shadow casting correlator, the encircled energy within the 'shadow' spot increases linearly until the full spot size is reached and is then 100% efficient. At \( q = 100 \mu \text{m} \), the efficiency is 60% which is lower than the shared microlens system but larger than the pinhole system (fig. 5.10b). This value is chosen to maximise the number of elements and yet still maintain a relatively high efficiency. If fewer elements could be tolerated than the highest efficiency could be achieved using the shadow casting system at \( q = 165 \mu \text{m} \). So why not use the shadow casting correlator instead? The benefit of going through the trouble of using long focal length microlenses is that the number of pixels in 1-D that are allowed for the shared microlens system is \( N = 2.11 \text{ cm} + 100 \mu \text{m} = 211 \) whereas that in the shadow casting system is \( N = 2.11 \text{ cm} + 165 \mu \text{m} = 128 \); so we get 2.7 times improvement in \( N^2 \) total number of pixels.

<table>
<thead>
<tr>
<th>System</th>
<th>Resolution limit (diameter)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>formula</td>
</tr>
<tr>
<td>Microlens</td>
<td>1.22 ( z_2 \lambda/a )</td>
</tr>
<tr>
<td>Pinhole</td>
<td>1.22 ( a (1 + M) )</td>
</tr>
<tr>
<td>Shadow casting</td>
<td>2( a (1 + M) )</td>
</tr>
</tbody>
</table>

Table 5.1 Comparison of resolution and efficiency for the different systems

The distance \( z_2 \) chosen for the pinhole system is a realistic value as a smaller size would imply large off-axis angles since the size of the input and mask planes are both equal.
to 2.11 cm (128 × 165 μm). This gives an off-axis angle of about (\(\tan^{-1}\left(\frac{2.11}{2 \times 2.06}\right)\)) = 27°.

The illumination system cannot provide larger off-axis angles than this (see chapter 6).

\[\lambda = 660 \text{ nm} \quad 2a = 165 \mu\text{m} \quad z_2 = 2.06 \text{ cm}\]

If the focal length of the microlens system is allowed to be variable so that the system plane separation ratio is always kept at 1:1, the spot size (diameter) varies with the distance in the following relationship and is depicted in fig. 5.11:

\[q(\text{diameter}) = \frac{1.22z_2\lambda}{2a}\] (5.27)

\[\text{Microlens} \quad \text{Pinhole} \quad \text{Shadow casting}\]

\[\text{Shared lens system} \quad \text{Pinhole system} \quad \text{Shadow casting}\]

\[\text{Geometric limit}\]

Distance between the filter and output plane \(z_1\), (cm)

Spot size (μm)

Fig. 5.10 Comparison of shared microlens and pinhole systems in (a) resolution and (b) efficiency

Fig. 5.11 Focal spot size variation with distances for each system
The result is a linear relationship. Similarly, if the pinhole system is placed under the same constraints, the spot size varies in a very similar way to the shared microlens system (shown by the '+' in the graph which are calculated values). The data points correspond to the distances shown in fig. 5.5c-f. The reason for discrete points is that each spot size have to be numerically determined. The data point at \( z_2 = 1.4 \, \mu m \) lies outside the line because it is not at the one of the maxima shown in fig. 5.5b in the Fresnel diffraction region. Although it may suggest a smaller spot size can be obtained in the Fresnel region if a maxima is not chosen, this usually accompanied with the penalty of low efficiency and high crosstalk as the spot has large sidelobes. For the shadow casting system, the spot size remains the same with distance unless the geometry is changed, i.e. \( M < 1 \) for smaller spot size. We conclude that for medium size (3.5-4 cm) systems the shadow casting arrangement is best as it gives high efficiency and smaller spot size. For small systems (< 3.5 cm), the shared microlens system is best as it gives smaller spot size (higher SBWP) and higher efficiency than the pinhole system. The spot size and SBWP improves as the system shrinks but the maximum off-axis angle increases. The minimum size of the system is limited by the range of angles that is possible to generate by the illumination system. This is the reason why devote chapter 6 to this issue.

5.6 Off-axis diffraction

5.6.1 Introduction

So far, only the diffraction of paraxial rays have been discussed. However, in a correlation system where rays are travelling at a wide range of angles to reach an output pixel, the above theory must be modified to allow for off-axis aberrations. The off-axis effects on both the shared microlens correlator and the pinhole system will be investigated in this section.
5.6.2 Shared microlens system

A quick but approximate method to compute off-axis diffraction is to use the available paraxial theory with a rotation of the axes to account for the off-axis effect. For example, in fig. 5.12, the input is rotated so that it subtends an angle of $\theta$ at the centre of the lens. By changing axes, $x_1' = x_1 \cos \theta$ and $x_3' = x_3 \cos \theta$, the diffraction theory derived above can be used with the modification of $\alpha$ to $\alpha \cos \theta$ and the separation distances $z_1, z_2$ to $z_1 \cos \theta$ and $z_2 \cos \theta$, respectively. In addition, the phase change through the lens at off-axis incidence has to be calculated. It is shown (in Appendix II) that for a thin lens approximation, only a shift of the output focus results and this has already been accommodated. The thin lens approximation is also valid because the $f/# (=f/D)$ is very large ($1.03 \text{ cm} / 165 \mu \text{m} = 60$).

![Fig. 5.12 Schematic showing parameters for off-axis diffraction calculation](image)

In this calculation, we assume that all of the off-axis spots are in focus at all of the angles and investigate the change of spot size with off-axis incidence. All of the system parameters used here, e.g. $z_2$, are the same as in the on-axis case. The radial spot size in the $x_3$ direction (projected into the output plane) becomes larger as the off-axis angle increases but that in the $y_3$ axis remains unchanged, giving rise to an elliptical spot (Fig. 5.13).
5.6.3 Pinhole system

The same procedure is repeated for the pinhole system, in which case the lens at the aperture is removed this time. It has a similar spread of the spot size at large off-axis angles (Fig. 5.14).

By comparing the variations of spot size with angles for both systems, we can observe that they have virtually the same increase at a particular angle (Fig. 5.15). In fact, both systems
are in the Fraunhofer region where the spot size is dominated by diffraction. This means that the spread is independent of the presence of the lens and that the effect of the aberration cannot be seen. Consequently, the spot size is merely due to the field curvature at off-axis angles.

5.6.4 Optimisation for off-axis applications

Previously, the values for $f$ in the case of the lens system and $z_o$ for the pinhole system were chosen for each off-axis angle so that they both satisfied the lens law: $1/f = 1/z_1 + 1/z_2$ in the microlens system and $1/z_o = 1/z_1 + 1/z_2$ in the pinhole system. In this way, all the spots in the output plane are in focus, i.e. the smallest spot with the highest efficiency. However, this is not strictly true in reality where only one angle is in focus and the spot becomes blurred at other angles.

We shall use $30^\circ$ as the maximum off-axis angle that can be allowed in this section because this is a typical value measured by us for the illumination system (chapter 6). Within this angular variation, the focussed spot size can vary from 100 $\mu$m @ $0^\circ$ to about 150 $\mu$m @ $30^\circ$ (fig. 5.15). In order to minimise the maximum deviation of the spot size, the system is designed so that the focussed spot occurs at an off-axis angle of incidence, rather than at the usual on-axis case. For example, the following procedure can be used to determine the focal length: The relationship shown in fig. 5.15 is approximated by a best fit
A horizontal line is cut across the parabola and by the exact location of the line is determined such that the areas above and below that line are equal. Where the line meets the parabola is the angle $\theta_0$, at which the spot is focussed so that minimum deviation of spot size is obtained. This point is found to be at $\theta_0 = 17.3^\circ$ (fig. 5.16). Now, the system parameters can be chosen, i.e. $r_I(\theta_0) = r_2(\theta_0) = 2.06$ cm and $f(\theta_0)$ is half of that distance. At another off-axis angle, $\theta$, $r_I(\theta)$ is given by $r_I(\theta) = r_2(\theta) \cdot \frac{\cos \theta_0}{\cos \theta}$, the respective distance $r_2(\theta)$ can be calculated similarly. These distances will not be the same as the distance at the angle of $\theta_0$ from the centre of the aperture to the output plane at $(x_3, y_3)$. In general, the focal plane does not coincide with the output plane except at $\theta_0$. In order to obtain the actual spot sizes at the output plane, one must propagate back both plane the spot size, $w(z)$, at a distance, $z$, from the focussed spot using the Guassian beam relationship:

$$w^2(z) = w_o^2 \left[ 1 + \left( \frac{2z}{nkw_o^2} \right)^2 \right]$$

(5.27)

where $w_o$ is the focussed spot size and $n$ is the refractive index of the medium.

![Fig. 5.16 Variation of radial spot size at ± 30°](image)

![Fig. 5.17 Schematic showing calculation of astigmatism](image)

So far, only one dimension has been discussed. Clearly, the on-axis analysis can be applied in the other dimension assuming $y = 0$. As a result, there are two focal planes and this effect is known as astigmatism. If the input is off-axis in both the x and the y directions, and at a
different angle in each, then, there are two distinct foci (sagittal and meridional), corresponding to the different off-axis shifts.

The effect of astigmatism in this system is less significant since the numerical aperture is very small which implies a long depth of focus. This is evident from fig. 5.18a which shows the a 1% variation of spot size with off-axis angle in the range -30° to 30° when the focussed spot is designed at 17.3°. If on-axis beams are chosen to be focussed instead, the variation is about two times larger (fig. 5.18b). However, the spot size in the first case is generally larger than that in the second. Therefore, if a small spot size is crucial to give a large number of pixels, then this off-axis compensation is not required but if the variation of the spot size is the important issue, this compensation technique provides a solution.

![Graphs showing spot size variation](a) ![Graphs showing spot size variation](b)

**Fig. 5.18** Actual variation of spot size when the focussed spot is at (a) 17.3° and (b) 0°

### 5.7 Conclusions

In this chapter, we used a diffraction theory to analyse the shared microlens system to calculate the resolution and encircled energy within a particular output focal spot size. These parameters have great implications for the system length. The pinhole system is also analysed by the same theory for comparison with our system. The properties of a shadow casting system are also briefly examined.
Assuming that both the microlens and the pinhole have the same diameter, the shared microlens system has the same spot size and, thus, resolution as the pinhole system for the same system size but has almost twice the efficiency within the focused spot. In a particular shared microlens system with an LED of wavelength 660 nm with a diameter of 165 \( \mu \text{m} \), focal length of 1.03 cm and a plane separation ratio (magnification) of 1, the resolution is determined to be 200 \( \mu \text{m} \) in diameter for both systems. The 'shadow' or spot size of the shadow casting system under same conditions is found to be 330 \( \mu \text{m} \).

The shared microlens system has about 86\% of the focal spot energy contained within the mainlobe of the diffraction pattern. The rest of the energy may be lost or give rise to cross-talk if the pixel size is made to be the same size as the resolution limit. On the other hand, the pinhole system is only 46\% efficient. Although shadow casting is supposed to be 100\% efficient, it has to operate with the geometric limit (i.e. without diffraction) and the spot size is larger than both microlens and pinhole systems.

Another drawback of the pinhole system is that it is most efficient only at one system size for a specific mask pixel size at a particular wavelength. At other working distances, the efficiency will be lower. Whereas for the shared microlens system, the efficiency remains the same for different distances as long as lens law is satisfied and microlenses of various focal lengths can be fabricated.

Off-axis aberrations such as field curvature and astigmatism are also considered. Both systems give the same results with the above system parameters. Since the focal length of the microlens is long compared to its diameter, the spot size of the focus, even at off-axis angles, is mainly diffraction limited and aberrations are less significant. By setting the system to be focused at an off-axis angle of 17.3° for a system with maximum off-axis of 30°, the variation of the spot size is half of that in an on-axis focused system. However, the latter system has an overall smaller spot size. There is a compromise between the variation of spot size and the actual spot size.
6.1 Introduction

Incoherent illumination is a crucial and integral part of our correlation system. It provides the operations of shift, multiply and enables correct summation of intensities at the output. The strengths or weights of the individual pixel to pixel interconnections is partly determined by the illumination uniformity. Therefore, we discuss in this chapter the characteristics of our illumination system in detail. First, the specific illumination requirements of our system are outlined (section 6.2). The illumination system consists of two elements: a light source and a diffuser. Different types of light source and diffuser are reviewed and discussed in section 6.3 and section 6.4 respectively. Other related system parameters such as the contrast ratio of the SLM and sensitivity of the CCD camera are taken into account in making the choice of a suitable light source (section 6.5). We propose the novel use of a microlens array as a diffuser and we compare its characteristics with that of a plastic diffuser in section 6.6. The design of novel non-regular microlens arrays for use as diffusers are discussed in the penultimate section with suggested applications (section 6.7).
6.2 Illumination requirements for the shared microlens correlator

In the shared microlens system, diffusely scattered illumination from every input point is employed to achieve the pattern shift operation. The angular spread of the illumination from each input pixel varies with its location in the input plane (fig. 6.1). In the simple case when the illumination at all pixels is normal to the input plane, the pixels at the edge have a larger half-angular spread than those nearer to the centre of the array. This is simply because half of the light emanating from those at the edge is lost. Therefore, the angular spread for such a system is determined by the half-angle required from the pixels at the boundary of the input plane to the middle of the second plane ($\theta_{\text{max}}$).

The second requirement for the illumination system is to have incoherent addition of intensities at the output. This can be achieved by either using spatially incoherent light, i.e. the phase of the light from different input pixels is completely randomised so that no interference occurs, or using multiple wavelengths from each input pixel.

Finally, the uniformity of the illumination is important so that the strength of each connection is the same. This includes uniformity with both angle and spatial position. Non-uniformities will degrade the signal-to-noise ratio and may result in errors in the correlation operation.

![Diagram](image)

Fig. 6.1 The ideal illumination for the shared microlens correlator: full angular coverage and uniform strength at angular and spatial positions.
6.3 Source for incoherent illumination

6.3.1 Definition of photometry quantities

Illumination is quantified with several photometry terms [Lon73],[Lev68] and are briefly described:

(a) Luminous flux (Φ) is the energy flux (in Watts) weighted according to its efficiency in producing a visual response and has the units of lumens

(b) Luminous intensity (I) is the luminous flux per steradian, and has the units of candelas

(c) Luminance (L) is the luminous intensity per unit area

(d) Illuminance (E) is the luminous flux incident per unit area

Illuminance measures the flux received by the surface and luminance measures the flux emitted in a given direction.

6.3.2 Incoherent illumination source

Following the discussions on the requirements described in the section 6.2, the light source suitable for the system must be spatially extended and uniform, have spatial incoherence and enough angular spread. There are four types of sources that are appropriate for this purpose:

(a) Laser with rotating diffuser [Ath95]

An unexpanded laser beam gives a very narrow beam waist and the beam needs to be expanded to increase the spatial extent. The light in a expanded laser beam is in phase and can cause interference. A rotating diffuser is used to introduce incoherence to the source: the scattering provides the spatially extended source and the rotation removes the temporal coherence. Since a laser is used in the setup, a higher luminance is delivered compared to a normal incoherent light source. Another advantage is the ease of alignment in the system. The drawback is that a mechanical rotation stage is required.
which implies a bulky system. At some distance from the diffuser, the interference of the light scattered causes a large local variation of intensity which is perceived as speckles. The average size of the speckles is inversely proportional to that of the area illuminated at the surface of the diffuser. Speckles are also created by the usual scattering from dust and small particles on the surface of optical elements but the effect is negligible compared to that from the diffuser.

(b) Vertical cavity Surface-emitting laser diode arrays with different wavelengths [Pae91]

Recently, vertical cavity surface-emitting laser diode arrays have been developed [Jew90]. The individual lasers can be very small (a few microns) so that high resolution and large integration within a device is attainable [Pae92]. More importantly, each laser has an unique wavelength so that they are not temporally coherent with each other. Furthermore, the phases of the light from each laser are not locked with each other and, therefore, they are not spatially coherent. Consequently, incoherent processing is possible without interference [Pae91]. This replaces the need for a diffuser and the signal-to-noise levels are much enhanced. However, the device is not suitable for our system since it is more costly and large arrays are not yet available.

(c) Light-emitting diodes [Raj93]

LEDs have a larger spectral bandwidth (which is typically about 50nm) than the laser (typically about a few nm). Radiation from an LED results from the spontaneous emission when electron-hole pairs recombine in a p-n junction and, therefore, the illumination is incoherent in phase. Usually a poor quality lens has been built into the device to obtain some degree of collimation. In order to increase the illuminating angles and to obtain better spatial uniformity, a diffuser is often required in the system. Using an array of LEDs instead of a single LED gives higher luminance and a larger spatially extended light source.

The LEDs used are the ultra-bright red LEDs that are available commercially. The maximum luminous intensity is 14 cd (made by Toshiba) although 1 cd LEDs (made
Incoherent illumination

by Stanley) are also used in some experiments. Given the area of the active region where light is emitted, the luminance can be calculated.

(d) White light source [Lev68][Eal94]

White light illumination can be obtained from natural sunlight or artificial lamps. One of the most common forms of artificial lamp is the tungsten lamp. The tungsten filament inside the bulb is heated by the current to incandesce so that it becomes a light source. Usually the bulb is filled with inert gas to reduce the evaporation of the tungsten but iodine is used in tungsten halogen lamps to increase the brightness. Iodine reacts and combines the tungsten deposited on the bulb wall and the resulting tungsten iodide diffuses to the hot filament where the tungsten is redeposited. As a result, the bulb can be operated at a considerably higher temperature with an increase in efficiency and luminance. Tungsten lamps have luminance typically about 50 cd/mm². Higher luminance can be obtained from gas discharge lamps such as the carbon arc, the high pressure mercury and xenon arcs. Some can exceed the solar luminance (1500 cd/mm²).

On the other hand, fluorescent lamps which have a low pressure and low current discharge, have a much lower luminance, e.g. 0.4 cd/mm² for a 8W tube (12" long, dia. 1½”). The bandwidth of these lamps usually cover the whole visual spectrum (300 nm) and, hence, is much wider than both that of the laser and the LED.

These artificial lamps are mainly designed for illumination in surroundings with scale much larger than an optical system, so if the light source is implemented directly in our system, a significant amount of illumination radiates away in other directions and is wasted. Modifications of the light source are required to give an efficient illumination system. Usually, the artificial lamp is enclosed within a box with inner reflective walls (e.g. dichroic reflectors) and an aperture on one wall for the output. This ensures that most of the light is captured. These type of light boxes or lamphouses can be found in back-lit liquid crystal display screens and fibre optic light box. The former uses a serpentine-shaped fluorescent tube with reflectors at the back. An 4” active-matrix LCD display back lit unit (made by Sharp) has been investigated but the luminance of a fluorescent tube is low compared to other light sources (the figures are compared in the last paragraph). A fibre optic light source consists of an enclosed (75W) tungsten
halogen lamp and the output light is guided by a optical fibre bundle (dia. 7mm). This light source will be investigated in the suitability for our system later in the chapter. The most efficient 'light box' is an integrating sphere because of its geometry. It provides an additional advantage of rendering the illumination more uniform. The drawback is the size of the sphere which is usually large to give a very uniform illumination.

Although most white light sources emit light at large angles, the use of a light box for higher efficiency means that some angular spread and spatial uniformity is compromised. In the case of a fibre-optic light source, the illumination has a spatial extent of 7 mm, and is partially collimated. In order to achieve a larger spatial extent and angular spread, an additional diffuser is needed.

6.3.3 Conclusions

The illumination system incorporating a laser with a rotating diffuser is too bulky for our constraints on system size. The VCSEL arrays are not yet widely available and will not be considered further. Therefore, the superbright LEDs and the white light fibre optic source are chosen for the reasons of high luminance, cost and availability. They will be investigated further in section 6.5.

6.4 Diffusers

6.4.1 Introduction

From the discussion in the previous section, a diffuser can be used to remove the coherence in a laser beam, or to provide spatial uniformity and angular spread as in the case of incoherent sources of LEDs and white light lamps. In this section, the mechanism and characteristics of the different diffusers are discussed and compared.
6.4.2 Operation of the diffuser

Most diffusers, one way or another, rely on the same physical principle of scattering. This is caused by either the small optical non-uniformities within the material or non-uniform roughness on the surface. The properties of the diffuser are defined by the size and spatial distribution of the scattering centres. Each scattering centre can be viewed as an obstacle causing the diffraction of light. From either the diffraction theory for flat obstacles and generalised Mie theory for volume scatters [Hec87], the diffraction angle increases as the size of the obstacle decreases. On the other hand, the intensity of the scattered light is proportional to the size of the scatterer. Consequently, either efficient scattering of light in narrow angles or inefficient spreading of light over large angles may be achieved. The only way to increase both efficiency and angular spread of the scattered light is to use multiple scattering. This usually means volume diffusers with multi-layered structures of small scattering centres [Paw94]. However, the resolution of the diffuser can be low, giving poor uniformity. A compromise between the size of the scatterers, thickness of the scattering layer and efficiency is necessary for the design of diffuser.

6.4.3 Angular distribution of the scattered light

The angular spread of the diffuser depends on the illumination conditions (e.g. angle of incidence) and properties of the diffuser. A perfectly uniform diffuser (Fig. 6.2a) obeys Lambert's cosine law of emission [Lon73] with the luminous intensity $I$ at a particular angle $\theta$:

$$ I = I_o \cos \theta $$

(6.1)

where $I_o$ is the luminous intensity in a direction normal to the surface of the diffuser. This is depicted as a smooth spherical lobe in polar coordinates (fig. 6.2a). In reality, most diffusers have angular properties far from this ideal condition and have an ellipsoidal-like lobe elongated in the direction of propagation of incident beam. For a random diffuser under normal incident illumination, symmetry is observed with the maximum intensity along the propagation direction whereas asymmetry occurs for inclined incidence. There is a little shift of the maximum of the scattered light away from the normal direction [Paw94] and the shift increases with the angle of incidence (fig.6.2c). The amount of
shift of the direction of the lobe from the angle of incidence depends on the effectiveness of scattering in the diffuser although in most cases, the lobe roughly follows the propagation direction of the incident beam. Therefore, illuminating traditional diffusers with a divergent beam gives scattered light in divergent fashion (fig. 6.2e).

Fig. 6.2 The polar diagrams of the angular distribution of scattered light from a diffuser
(a) comparison of actual profile with the ideal case 
(b) at normal incidence and (c) at oblique incidence [Paw94]  
(d) from a parallel beam and (e) from a divergent beam [Paw94]

For volume diffusers with thick material of small scatterers, the angular distribution of the scattered light may become independent of the direction of the incident beam.

Different kinds of diffusers are described in the following sections with reference to the requirements of the shared microlens correlator.

6.4.4 Traditional diffusers

There are two main types of diffusers: forward (transmissive diffuser) and backward (reflective diffuser) scattering diffusers. In the geometry of our correlator, transmissive diffusers are more suitable than reflecting ones although it should be noted that reflective diffusers coated with magnesium oxide are almost perfectly Lambertian at 45° incidence [Lev68].
Ground (and sand-blasted) glass [Kno49],[Kno60] is one of the most popular media for transmissive diffusers. It does not have a very large angular spread (typically about 10°).

More perfect scattering can be obtained by opal glass which consists of neutral diffusing particles suspended in glass. It provides almost uniform diffusion of light although the efficiency is much lower than that of the ground glass.

Plastic materials, like bulk acrylic, can have a variable degree of scattering. It is usually used as diffuser screens for liquid crystal display backlight illumination box. As it is a volume diffuser, some of the incident light is reflected backward. This reflection loss becomes more significant as the material becomes thicker, e.g. white plastic diffuser diffusely reflects half of the light.

6.4.5 Holographic diffusers

Recently, a new type of diffuser has emerged to offer better efficiency and lower power consumption compared to traditional diffusers in applications which are usually difficult to implement with the latter. These are holographic diffusers which can offer the following advantages [Paw94]:

(a) concentration of significant amount of light in one or many areas
(b) control of angular distribution of light
(c) better luminance at the prescribed direction or lower overall power consumption

and the drawbacks are:

(a) complex production technology
(b) difficult to produce large size elements
(c) higher cost

Holographic diffusers can be thin or volume holographic optical elements or computer generated holograms. Thin holograms that use silver halide films [Wad94] have low efficiencies and narrow diffusion angles. Some holographic diffusers (those available from Ealing [Eal94]) are surface relief holograms made by etching a 2-D interference pattern into a deep relief structure (with 0 or π phase change) for higher
efficiency. Master holograms are recorded and etched in photoresist, and can then be copied at low cost by embossing onto a plastic substrate. Typical diffusing angles are up to 60°.

Volume holograms are made by optical recording in materials such as dichromated gelatin [Ted94] and Du Pont photopolymers [Wad94]. They have high efficiencies and angular selectivity for use off-axis. The volume holographic diffusers exhibit large angular scattering (up to about 60°) and higher efficiency than conventional diffusers with similar angular characteristics.

We propose the use of computer generated holograms as a diffuser. The advantages over the optically recorded holograms are the design flexibility and simulation using a computer and the avoidance of complex production technology as in the case of the volume holograms. The disadvantages are that both the efficiency and the resolution (about 0.6 μm) are compromised (see chapter 2 for more detail on CGHs). A relatively low resolution diffuser can be implemented using the Fresnel-Dammann grating [God93] which can be suitable for the illumination of the correlator. This will be described in more detail in section 6.8.

6.4.6 Refractive diffusers

Besides the scattering of light either at the surface or within the material and the diffraction of light by fringes, there is a third mechanism: refraction using lensing surface profiles to redirect light to a more uniform distribution. This could either be transmission lenses or reflective curved mirrors. This method has been implemented by embossing grooves in 1-D (which act like lenticular lenses) or small square lenses in 2-D on plastic sheets. These diffusers can be commonly found in scattering screens for blurring the objects behind for security or as a diffuser surrounding a fluorescent tube for better uniformity.

Fibre arrays [Web92], which act as high numerical aperture cylindrical lenses, have been used to provide uniform illumination with half-angle of 35°. This diffuser is implemented in a line-scan camera [Web92]. In 2-D, Lenslet arrays (φ 2 cm, f = 20m) have been used as a diffuser for providing more uniform illumination in laser fusion and laser heat processing [Den86]. Microlens array diffusers can also be made using a
particular type of DuPont photopolymer material for low frequency holograms [Phi91]. When the photopolymer undergoes contact lithography, the U-V irradiated regions will be polymerised and monomer diffuses from the unexposed surrounding. This causes a change in refractive index and a structure which behaves like a GRIN lens is formed. However, the monomer diffusion is usually accompanied by a surface relief profile. Finally, total polymerisation completes the process. The resulting screen with local index variations can be used as a volume and surface diffuser. The quality of the diffuser relies on the quality of the material and there is little control on the characteristics of the diffuser.

We propose the use of the microlens array, in particular using photoresist reflow microlenses, as a diffuser independently of the papers reported on the photopolymer microlens. Fabrication using photoresist reflow technique yields consistently good quality microlenses. These microlenses can then be etched into the glass substrate to give robust stable and non-absorptive diffusers.

6.4.7 Other diffusers

Most material can, in fact, be used as diffusers providing that it scatters uniformly. For example, the scattering properties of dielectric material such as silicone rubber, liquid crystals etc. are investigated. Since these materials have elongated scatterers, the polarisation and the angular spread will be different in each direction. Usually an elliptical angular spread is expected. White paper can also be used as a reflective diffuser.

6.4.8 Summary of the different diffusers

Figure 6.3 gives a comparison of the most common diffusers. Ground and sand-blasted glass have very small angular spread. Opal glass and white plastic diffusers, on the other hand, approaches the perfect diffuser, with angular spread up to 80°. The drawback is that the intensity is very low (about a couple order of magnitudes lower). Other bulk acrylic diffusers fall somewhere between the ground glass and the white plastic in terms
of angular response and diffuse transmittance. Some holographic diffusers can achieve about 60° spread without compromising too much scattering intensity.

![Diagram of diffuse transmittance](image)

**Fig. 6.3 Summary of the typical shape of diffusion and angular spreads of the different diffusers described [Ted94]**

Although the traditional scattering diffusers have little flexibility, they are cheap and commercially available. Volume holographic diffusers involve complex fabrication processes and will not be considered further for our applications. In terms of microlens array, the photopolymer material is still not advanced enough for repeatable results. Consistently good quality microlenses can be obtained by the photoresist reflow process. Moreover, NSG planar microlens arrays are readily available.

### 6.5 Effect of illumination on other system components

#### 6.5.1 Introduction

Other system parameters which are affected by the illumination, include the contrast ratio of the spatial light modulator and the sensitivity of the CCD camera. These will be addressed in the following sub-sections.
6.5.2. Effect of illumination on contrast ratio of the SLM

The Ferroelectric liquid crystal spatial light modulator consists of 128 × 128 pixels with pitch 165 μm and dead space of 5 μm between pixels. The contrast ratio within individual pixel is quoted to be about $10^3 : 1$ [CRL]. However, since two dimensional images are processed in the system, the overall contrast ratio (CR) is more important. The overall CR is defined to be the ratio of the intensity of the transmitted light when the pixels are switched to maximum transmission ('white' state) to that when the pixels are turned to minimum transmission ('black' state). The contrast ratio is measured with the experimental setup shown in fig. 6.4 with a different light source and the results are given in table 6.1.

The degradation of the contrast ratio with the broadening of the linewidth / bandwidth of the illumination is due to the optimisation of the liquid crystal cell as a half wave plate at a single wavelength. More details will be given in section 8.3.2. Using the experimental results, we can rule out the use of the white light source because of its poor contrast. Therefore, we use LED as the illumination source for our incoherent system.

![Fig. 6.4 Experiment setup for measuring the contrast ratio](image-url)
<table>
<thead>
<tr>
<th>Illumination source</th>
<th>Overall contrast ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>HeNe laser</td>
<td>16:1</td>
</tr>
<tr>
<td>red LED</td>
<td>10:1</td>
</tr>
<tr>
<td>White light fibre optic source</td>
<td>2:1</td>
</tr>
</tbody>
</table>

Table 6.1 Overall Contrast ratio for different illumination source

6.5.3 Sensitivity of the CCD camera

The sensitivity of the CCD is important to ensure enough illuminance is provided at each angle. The sensitivity quoted for the Sony XC-77RR CCD camera is 400-lux. The CCD is sensitive to the visible spectrum and especially to infra-red wavelengths since it is a silicon-based detector. An IR blocking filter (Schott glass) may be required if the illumination has a significant proportion in the IR region. This is particularly important if a wide spectrum source such as a white light source is used.

6.6 Uniformity comparison between plastic and microlens array diffusers

6.6.1 Introduction

Uniformity of the illumination is important to ensure successful operation of the system. Ideally, we would like the illumination across the output to have a flat spatial extent and a large enough angular spread. In reality, it is always non-uniform and insufficient angular coverage. In this section, a plastic diffuser and a microlens array diffuser will be characterised using a red super-luminescent LED as illumination for our system.
6.6.2 Choice of diffusers

Many different diffusers were evaluated for their suitability in providing uniform incoherent illumination in our system. These included a ground glass, a flashed opal glass, plastic diffuser from a Sharp LCTV, perspex diffusers with transmissivity of 40%, 55% and 78%, and a microlens array. After some initial examination, we found that the angular spread of ground glass is too small (~10°) and the opal glass, whilst giving a Lambertian response, is extremely wasteful of energy. We chose the plastic diffuser from Sharp because it has a relatively large diffusing angle and was more efficient than the opal glass. In addition, the sponsor feels that it is better if the in-house product is used to develop this system because it is already available. The potential of using a microlens array is also explored in this application. In this first instance, a NSG planar microlens array is used since microlenses made in photoresist have absorption loss. Eventually, the benefits of photoresist microlenses as a diffuser can be realised by etching the surface relief microlenses into the glass substrate. The NSG array is available commercially and it consists of a square packing of 32 x 32 microlenses, each has a diameter of 250 \( \mu \text{m} \) and a focal length of 560 \( \mu \text{m} \).

6.6.3 Properties of the light emitting diode used in the experiment

The LED is a Stanley No. H1000-L and has a centre wavelength of 660 nm with a spread of about 40 nm and the typical luminance is specified to be 1000 mcd @ 20 mA. It is one of the superbright LEDs available commercially. The LED has an integral plastic lens (\( \phi 4.45\text{mm} \)) moulded on to the housing which appeared to poorly collimate the light. We observed that the direct illumination was not radially uniform and resulted in an irregular pattern of a roughly square central spot surrounded by a circular bright ring and other incomplete rings of larger diameter on a plane placed more than a few millimetres away. The diameters of the beam waists of the LED illumination were measured at several distances away from the LED and the angle at which the light from the LED is spreading was calculated to be about 5.7°.
Incoherent illumination

We decided to use a single LED for the measurement of the diffuser characteristics although this was extended into a 2x2 array in the final system as this improved the spatial uniformity into the diffuser.

6.6.4 Experimental apparatus

One of the important requirements in this experiment is that the source and detector must be properly aligned. This proves to be difficult using the LED for the alignment since it appears to poorly collimate the light. Therefore, a HeNe laser is used for precise alignment of the incoherent system. The laser beam is arranged in a ring configuration so that different components in the system can be aligned separately by the two beams. The laser is turned off after the alignment is completed.

The LED was placed a distance, x, from the diffuser with the LED aligned so that the brightest illumination fell at the centre of the diffuser (fig. 6.5). A circular pinhole of diameter 200 μm was made in the first opaque plane and a circular pinhole of 1mm diameter was made in the second opaque plane placed a distance, y, away. The second pinhole was made larger to increase the signal falling on the detector to obtain a better signal to noise ratio. The detector is placed to intersect the entire beam through the second pinhole. A Newport No. 835 picowatt power meter was used to record the intensity. The first pinhole selects the lateral position on the plane while the second selects rays of light travelling at different angles from the diffuser. In the case of the microlens diffuser, the first pinhole is centrally aligned with a microlens.

Fig. 6.5 Experimental apparatus for measuring the uniformity of the illumination
6.6.5 Spatial uniformity measurements, results and discussion

The distance between the LED and the diffuser (x) and that between the diffuser and the detector (y) determines the system length and the intensity of light detected. Fig. 6.6 shows graphs of intensity recorded with varying distances of x and y when the system components are centrally aligned. As expected, the intensity falls off roughly \(1/(\text{distance})^2\) in both cases. These curves gives a maximum x and y for a particular sensitivity of the detector. If the LED is very close to the diffuser, the luminance is high but only part of the diffuser is illuminated since the light emanating from the LED is only spreading at a small angle. This spatial extent of the illumination on the diffuser is important because the entire input plane which is placed next to the diffuser in our system must be illuminated. Similarly, the distance between the diffuser and detector (output plane) determines the size of the output plane for a given angular spread of the diffuser. A compromise is required as a result.

In the experiment, we chose the configuration with x = 10 mm and y = 10 mm. Keeping these values for the separations we moved the pinhole nearest the diffuser laterally in its plane along one dimension (z) and at each position we scanned the pinhole in the second plane laterally along the same dimension and plotted the light intensity.
Fig. 6.7 (a) Spatial uniformity of the LED illumination with the plastic diffuser
(b) Intensity plots cut at the different microlens plane, z

detected in fig. 6.7a as a function of these two lateral displacements. This plot is normalised to the peak intensity of 13.8 nW/mm² and the contour lines show the
fractional variation relative to this value. Fig. 6.7b shows the intensity plots at the output plane along the different lateral position, $z$, of the pinhole on the diffuser. For example, the solid line curve is the intensity variation at the central position (when $z = 0$). Similarly, the spatial uniformity profile of the microlens diffuser with the intensity contour plot normalised to the peak intensity, $480 \text{nW/mm}^2$, is shown in fig. 6.8.

Comparing the two plots 6.7a and 6.8a, it is clear that the microlens array tends to produce a more laterally restricted distribution. The highest intensity was observed approximately on axis as expected. The distributions shown appear to be stretched along a line of slope of about 1.3 independent of the diffuser used. This is to be expected since the ratio of the distance moved by the second pinhole to that moved by the first should be a constant, given by the ratio of the distances $(x+y)/x$ if the LED is assumed to be a point source. In this particular system configuration ($x = y = 10 \text{ mm}$), the ratio would be 2. In reality, the LED is not a point source and has a certain spatial extent. However, this cannot be used to explain the discrepancy since the ratio is independent of this factor. There are no components between the pinholes to cause any apparent change in distance, so the measured value of $y$ should be correct (10 mm). Assuming the distance $x$ is affected by the different media present, the new $x'$ can be calculated from the ratio measured on the uniformity maps:

$$\frac{10 + x'}{x'} = 1.3 \quad \text{or} \quad x' = 33 \text{ mm} \quad (6.1)$$

The difference between the values of $x$ and $x'$ is 23 mm! This cannot be explained by the presence of the diffuser which can introduce an apparent depth since the refractive index of plastic is about 1.3-1.5. Furthermore, the microlens array and the plastic diffuser have different indices but both plots give the same slope. Therefore, this ratio is also independent of the diffuser. The LED has a collimating lens (of dia. 4.45 mm) at the front of the housing to give a narrow angular spread of $5.7^\circ$. The apparent point source giving rise to this spread can be calculated as $\frac{4.45}{2 \tan(5.7^\circ)} = 22.3 \text{ mm}$ behind the LED (fig. 6.9). This apparent depth of the point source corresponds to the difference in the distances and is the cause for the discrepancy.

If the spatial plots are now used in the design of the system, they indicate the largest area in the output plane under uniform illumination is $2 \text{ mm} \times 2 \text{ mm}$ for the
plastic diffuser and only 0.5 mm x 0.5 mm for the microlenses (if the illumination is not to vary by more than 5% from the peak intensity).

Fig. 6.9 Geometry of the system showing the apparent location of the source

If the uniformity map is cut along the slope of 1.3, a plot of variation of the maximum intensities of the rays at all the different angles is given. This means the detector moves according to the distance moved by the diffuser multiplied by the ratio \((x'+y/x')\). If both pinholes are moved laterally by the same amount, the intensity detected corresponds to rays travelling at the same angle but from different lateral locations on the

Fig. 6.10 Representations of the different line cut in (a) the spatial uniformity map as (b) different intensity profiles measured.
diffuser and, thus, is a measure of the uniformity of spatial spread across the diffuser. On fig. 6.10, this corresponds to travelling along a line of constant unity slope. A line with unity slope offset from the origin of the map gives a lateral intensity distribution at an angle. As discussed previously, the cut at $z = 0$ shows an intensity variation across the detector plane at centre of the diffuser. This plot will give an angular spread diagram if the different positions on the detector plane are calculated as the angles of the rays emanating from the diffuser. This will be discussed in greater detail in the next section.

### 6.6.6 Angular spread measurements and discussion

The angular uniformity can be assessed at one point, say the centre, of the diffuser if the second pinhole alone is moved laterally. As the distance between the plastic diffuser and the detector becomes larger, the intensity profiles begin to spread out as expected. In the case of the microlens array diffuser, the spatial extent of the illumination on the detector plane roughly stays the same. These plots are shown in fig. 6.11. If the distance $y$ is fixed now, the spatial spread on the detector plane from the plastic diffuser remains the same for all values of $x$ whereas that from the microlens array increases as the distance with increasing distance $x$. These correspond well with the results in fig. 6.11 and will not be

![Intensity profiles](image)

**Fig. 6.11** Plots showing the intensity profiles for fixed $x$ as $y$ is varied in the case of
(a) the plastic diffuser and (b) the microlens array
Fig. 6.12 Polar plot of angular uniformity as measured at the centre of the diffuser for LED illumination for (a) a plastic scatterer and (b) a microlens array diffuser shown here. Polar plots can be derived from these measurements and the results shown in fig. 6.11(a) and (b) are translated into polar diagrams in fig. 6.12(a) and (b) respectively. From these plots, we can see that the plastic distributed scatterer diffuser gives an angular spread of up to about 25° either side of the axis but weighted in strength being stronger nearer the axis whereas the microlens diffuser has a much narrower angular range up to about 13° either side of the axis but the intensity is very uniform within this. This figure is in good agreement with the numerical aperture for the microlenses which is 0.223 giving a half angular spread of 12.6°. Note that the microlens diffuser is more efficient than the plastic diffuser and there is an inverse relationship between the angular uniformity and the luminance of a diffused source.
6.7 Design of a microlens array diffuser with variable diameters

One of the important properties of the microlens array diffuser is that the angular spread is determined by the numerical aperture. This relationship can be shown as follows:

If a lens is illuminated by a collimated beam, quadratic phase is introduced and the beam converges to a focus. The diffraction limited focal spot of a lens can be determined by the Fourier transform of the light wave and is given by a Bessel function (see chapter 5). The spot size is then given by the Airy disc at the focal plane:

\[ q = \frac{1.22 \lambda f}{D} \]  \hfill (6.2)

where \( D \) is the diameter and \( f \) is the focal length. As the light begins to diverge from the focus, the flat phase at the focal plane changes into quadratic phase again of the opposite sense. At a long distance away (far field) from the focus, the diffraction can be determined by the Fourier transform of the focus. Now if the focus is approximated with just a circular spot of the size of the Airy disc without any outer rings, the far field diffraction intensity at a distance \( R \) is given by the Bessel function:

\[ I(\theta) = I(0) \left[ \frac{2J_1(kq \sin \theta)}{kq \sin \theta} \right] \]  \hfill (6.3)

where \( k \) is the propagation constant, \( q \) is the size of the Airy disc given by (6.2) and \( \theta \) is the angle of divergence:

\[ \sin \theta = \frac{R}{p} \]  \hfill (6.4)

where \( p \) is the spot size at the distance \( R \) and can be expressed as the Airy disc diameter:

\[ p = \frac{1.22 \lambda R}{2q} \]  \hfill (6.5)

Therefore the angular spread is given by:

\[ \sin \theta = \frac{R}{p} = \frac{1.22 \lambda}{q} = \frac{1.22 \lambda D}{2 \times 1.22 \lambda f} = \frac{D}{2f} \]  \hfill (6.6)

which is the numerical aperture.

For short systems, large angular spreads will be necessary. As the numerical aperture is given by \( \phi/2f = \tan \theta \), the half-angular spread \( (\theta) \) can be increased by either
Increasing the diameter or shortening the focal length, or both. f/1 lenses, which give a half angular spread of about 26°, can be fabricated with the photoresist reflow method but faster lenses (f/# <1) are not easily available. The system length is dependent upon how close the output plane can be placed, which in turns depends on the angular spread of the microlens at the edge for a given size of the output plane. We propose the use of different diameters of microlenses in an array to give a bigger spread at the edge. Obviously some light is lost in such a wide angle spread. Therefore, the aperture of the microlenses in the middle are smaller to give similar intensity at the output to that coming from a bigger lens at the edge (Fig. 6.13). The array will have the same focal lengths but with progressively larger diameter from the centre to the edge. This can be easily fabricated with the use of a customised e-beam mask and a uniform layer of photoresist. As discussed in chapter 7, the lenses will not be at good quality for all the different diameters but image resolution is not critical in this application.

A more efficient system is to incorporate a macro lens on top of a microlens array. The lens acts as a continuous prism to diffract the beams from the outer corners to the output plane and let those from the centre pass through more or less unchanged. Compared with just a fixed pitch microlens array, the macro lens helps to bend the directors of the lobes inwards so that the light normally lost is now contributed to the system (Fig. 6.14). Another advantage is that the maximum angular spread is the angle...
required for the light from the pixel in the centre of the input plane to reach the output plane, which is smaller than a system without a macro lens. If the lens pitch coincides with the size of the input pixel, each input pixel is imaged to the whole output plane. Another way to achieve this is to have an array generator which gives an array of discrete spots since the output is usually pixelated. This is essentially a Dammann grating situated next to a FT-lens, which can be implemented with Fresnel-Dammann grating [God93].

Fig. 6.14 The effect of the microlens illumination system (a) without a macro lens and (b) with the macro lens
7.1 Introduction

The diffraction theory of chapter 4 showed that the use of microlenses rather than pinholes can increase the power efficiency of the shared microlens interconnect while in chapter 6, we proposed the use of microlenses for diffusers. In this chapter, we describe the fabrication and report on the performance of photoresist microlenses and also the novel long focal length photoresist / index-matching fluid microlenses. The design of the photoresist microlenses will be outlined in section 7.2. The fabrication procedure of the photoresist microlens arrays is described in section 7.3 and the effect of the different fabrication conditions and parameters on the final quality of the microlenses is discussed. The imaging quality of these microlenses are examined and compared with graded-index microlenses purchased from Nippon Sheet Glass in section 7.4. Further quantitative comparison is provided in section 7.5 where the profile of the microlenses were measured in an interferometer. We also measured spot size and quantified aberrations. In section 7.6,
we describe the design, fabrication and optical quality of a new type of long focal length microlens. We compare their performances with that of the photoresist microlenses. We carried out an imaging system experiment to investigate the imaging quality of the long focal length microlenses. Finally, the conclusions are given in section 7.7.

7.2 Microlens specifications

We purchased several GRIN planar microlens arrays from NSG, which have the standard specification shown in Fig. 7.1. These microlens arrays are one of the cheapest available commercially and the lenses are good quality. In addition, these microlenses have the unique planar geometry since refraction is provided by the graded-index profile inside the substrate. From the specifications, the pitch and the diameter (250 µm) are the same, which suggests very little dead space. Also, the focus falls outside the substrate, which is useful for integration with other components.

![Fig. 7.1 Specifications of an NSG planar microlens array](image)

Photoresist reflow microlenses can easily be made as arrays in our clean rooms (UCL & SHARP). A mask is required to pattern the photoresist using standard uv-lithography techniques. This can take two forms: either a photo-reduced high contrast slide from a laser printer printout, or a chrome on-glass mask written by electron beam. The latter has higher resolution and yields better results.
The electron beam chrome-on-glass mask with array of opaque circles was designed using the “Wavemaker” program on a Sun Sparc Workstation. Then the data was sent to the Rutherford Appleton Laboratory, Didcot, where the mask was fabricated by electron beam writing. The designs on the mask are shown in fig. 7.2. Five arrays of 32 x 32 microlenses with a standard 250 μm pitch were designed. Each array has microlenses with slightly different diameters. These are used to match the specifications of the NSG microlenses (in terms of diameters and array size) for comparison. By designing different diameters in each of these arrays, the smallest gap between the microlenses before the molten resist joins during reflow can be determined. The smallest gap used is 2 μm. We also designed two other arrays which match the pitch and array size of the pixels in the Ferroelectric liquid crystal spatial light modulator and one array for matching a SHARP 7” liquid crystal TV screen. Finally, an additional array has 64 x 64 microlenses with half the pitch (125 μm) of the NSG microlenses.

Fig. 7.2 Schematic of mask layout for fabrication of microlenses
7.3 Fabrication of photoresist reflow microlens arrays

7.3.1 Description of the Fabrication process

The fabrication process of the photoresist microlenses involves four major steps:
(a) Preparation of the substrate
(b) Deposition of the photoresist
(c) U-V lithography and development
(d) Thermal reflow of the photoresist cylinders

Two types of substrates are used for making the microlenses: normal microscope glass slides and Pyrex discs. The 1" square microscope slide substrates are prepared by cutting normal microscope slides into three pieces. The 1" Pyrex discs (1.15 mm thick) purchased from Jencons have a fine ground finish on the edges, unlike the slides which have rough edges. The substrates are cleaned using isopropanol, 1,1,1-trichloroethane and methanol and blown dry. They are then placed inside an oven and dehydrated at 170°C.

After the preparation, the substrate is spun with a few drops of a primer, hexamethyl disilazane (HMDS) to promote photoresist adhesion. A thick, positive photoresist (Hoechst AZ 4620-A) is then applied onto the substrate using a syringe. The photoresist is then spun at a certain speed to give the desired thickness. This is followed by an oven prebake of the photoresist at 90°C for about 30 minutes which hardens the photoresist so that it will not stick to the mask later in the u-v lithography process.

Next, the photoresist is irradiated by ultra-violet light through an e-beam mask. The patterned substrate is developed in a 351 developer (diluted to ratio of 3.5:1 with de-ionised (DI) water) and washed with DI water. The substrate is checked under a microscope to ensure all the u-v exposed photoresist is developed away. The cylindrical islands of photoresist that remain on the substrate are put inside an oven or on top of a hot plate. Upon melting, the surface tension of the photoresist draws it into a spherical shape.
7.3.2 Factors which control the quality of the photoresist microlenses

The final quality of the microlenses depends on the raw materials (substrate and photoresist), the cleanliness of the substrate and every step that is involved in the fabrication process. These important factors will be discussed in more detail as follows:

(a) Substrate

Glass slides and polished pyrex discs are used as the substrates in the fabrication. Although microlenses form after reflow on both types of substrates, the quality of the lenses is quite different. This is because the shape of the lens is governed by the surface tension at the boundaries between air and melted photoresist, and the substrate [Sheri93]. These forces solely depend on the materials in close contact and give rise to a certain rim angle ($\theta$) at the photoresist-substrate interface with a particular photoresist and substrate. The microlenses fabricated on the glass slide are better than those on pyrex in terms of spherical aberration (interferometric measurements are presented in the next section). This is because the angle at the photoresist-pyrex interface is larger (Fig. 7.3a) As a result the central region of the lens has a larger radius of curvature than the side, giving rise to an imperfect hemisphere. Other factors such as rough edges of the substrates contribute to excessive photoresist build-up at the edge which can affect the uniformity of the microlens array.

\[ \theta = \text{function(surface tension, adhesion, cohesion)} \]

![Fig. 7.3](a) Dependence of the surface Profile of microlens on contact angle with the substrate
(b) Variation of thickness of the photoresist with spin speed
(b) Photoresist

<table>
<thead>
<tr>
<th>Photoresist type</th>
<th>Viscosity (cSt)</th>
<th>Thickness range (μm) at high speeds</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZ4330</td>
<td>100</td>
<td>2.5 - 3.8</td>
</tr>
<tr>
<td>AZ4620</td>
<td>360</td>
<td>4.7 - 7.1</td>
</tr>
<tr>
<td>AZ4903</td>
<td>1350</td>
<td>7 - 30</td>
</tr>
</tbody>
</table>

Table 7.1 Properties of AZ 4000 series photoresist

In order to fabricate fast microlenses (with large N.A.), the radius of curvature must be small. Thick microlenses require a thick, uniform layer of photoresist to be deposited. The type of photoresist must be viscous enough so that it gives the desired initial thickness for the design of microlens. The viscosity and thickness range of the different resist types are listed in table 7.1 [Jay92]. In our fabrication process, the Hoescht AZ4620-A photoresist is used.

(c) Spin speed

During the deposition of the photoresist, different spin speeds result in different thickness of the photoresist and this relationship has been determined by Daly et. al. [Dal91] (Fig. 7.3b). We noticed that although a slow spin speed such as 1200 rpm for 10 secs give 20 μm of photoresist, the thickness is not uniform across the whole substrate. At high spin speeds, uniform thickness can be obtained but the photoresist may not be thick enough. This is one of the trade-offs in the fabrication process. This problem can be overcome by using a more viscous photoresist. Another solution is to spin several thin layers of photoresist. In our fabrication, most of the substrates are spun at 4500 rpm for 10 secs to give an uniform thickness of about 6 μm. At this high spin speed, the resultant thickness is usually independent of the spinning duration.
Fig. 7.4 Surface profile plots of:

(a) Photoresist layer before melting

(b) Photoresist microlenses of \( \phi 120 \mu m \), pitch \( 125 \mu m \);
    on sample A5 (\( T = 210^\circ C \) for 3mins on hot plate)

(c) Photoresist microlenses of \( \phi 240 \mu m \), pitch \( 250 \mu m \);
    on sample A5 (\( T = 210^\circ C \) for 3mins on hot plate)
After the u-v lithography, development and melting processes, we find that microlenses on sample A5\(^7\) with diameter 120 \(\mu\)m only are successfully made using a substrate with an initial photoresist thickness of 6 \(\mu\)m (Fig. 7.4a). The surface profile of these microlenses was measured by a Tencor Alpha-step stylus profilometer and is shown in fig. 7.4b. Note that the horizontal axis is significantly compressed. The other microlens designs, i.e. those with diameters greater than 120 \(\mu\)m, have a dip in the middle of the lens. For example, a 240 \(\mu\)m diameter microlens profile is shown in fig. 7.4c. The dip is due to insufficient photoresist thickness and, hence, inadequate surface tension to pull it into a spherical shape. In addition, the contact angle discussed earlier also plays a part in the shaping of the microlenses during reflow. Indeed, this phenomenon which is all too often observed can be simulated by modelling using a set of approximate equations. By solving these equations with the known initial conditions, such as the contact angle, the lens shape and dip can be simulated \[Sch93\].

Microlenses with 240 \(\mu\)m diameter can be fabricated with a thicker layer of photoresist (12.5 \(\mu\)m). (Fig. 7.5) However, this requires slower spin speed (spin speed = 2000 rpm for 10 secs) and, hence, the microlens array is less uniform. The spikes on the profile graph are caused by the residual photoresist which is difficult to develop away due to the large thickness.

Fig. 7.5 Surface profile plot of photoresist microlenses of \(\phi\) 240 \(\mu\)m, pitch 250 \(\mu\)m; on B4
(T = 170\(^\circ\) C for 3 mins on hot plate)

\(^1\) 'A' stands for the first batch, 'B' means the second batch etc.; the number which follows stands for the sample number in that batch
(d) Melting parameters: time and temperature

The effect of various melting temperatures on the fabrication of microlenses on the pyrex substrate from 150°C to 210°C for 10 minutes is shown in the fig. 7.6. The reflow is conducted in a hot air convection oven. Below 190°C, a dip in the centre of the microlenses are formed. As the temperature is increased, the dip is becoming less significant. This means that the melting temperature is also one of the boundary conditions in the modelling of the reflow process. At 190°C, the lenses appear to have a flat top in the centre. This can give rise to serious spherical aberrations. The spherical shapes of the microlenses are formed at 200°C. However, at higher temperatures, the photoresist of the cylinders starts to join up. We conclude that reflow at 200°C for 10 minutes is the optimum condition for the resist thickness of 6 μm and microlens diameter of 120 μm. In fact, Jay et al. [Jay92] published independently a study on the effect of melting temperature and time and also determined that the ideal conditions are exactly the same as our findings. We also observed that the microlens with less than 5 μm gap between the lenses tend to join up after melting.

![Surface profile of microlenses at different temperatures](image)

Fig. 7.6 Surface profile of microlenses at different temperatures for 10 mins

We have attempted re-melting the unsuccessful microlens arrays at the same temperature as previously applied (for temperatures less than 190°C) but found that the dip was only slightly reduced. Therefore, the time of melting is not important. In addition, the photoresist islands were melted upside down but it seemed that gravity was not a major factor causing the dip of the lens.
(e) Melting environment

Melting can take place either in a hot air convection oven or on a hot plate. In an oven, the melting is slower than on a hot plate and the temperature is more stable. In contrast, melting is fast on a hot plate but in an uncontrollable fashion unless there is a temperature control for the hot plate. We have fabricated the best thin lenses using the hot plate. However, bubbles of nitrogen introduced in the u-v lithographic process are released quickly in a thick lens so that most microlenses are destroyed when the bubbles burst using the hot plate. Better results with thick lenses can be obtained in a hot air oven.

(f) Design of the microlenses: f-number and aspect ratio

The diameter of the aperture of the microlens is also an important parameter. In theory, different initial thickness of the photoresist can give the desired radius of curvature. From this relationship of thickness and radius of curvature, the focal length can be determined [Jay92, Dal91]. In practice, the limited range of photoresist thickness and the formation of undesirable sag on microlenses give rise to restrictions in the aspect ratio (ratio of focal length to Diameter) or the f-number (f/#). The range of f/# that has been achieved is f/1 to f/4. Higher f/# requires thinner lenses but the surface tension layers is then not sufficient to pull the lens into a spherical shape. As a result, long focal length microlenses are not readily available. We have made f/2.5 (A5) and f/0.5 (B4) photoresist microlenses on glass substrates, and f/3 (C5) microlenses on pyrex substrates (table 7.2).

Other factors such as exposure time, developer concentration and development time must be carefully controlled so as to produce good quality microlenses. Array uniformity can be affected by poor contact between mask and substrate, overexposure and inadequate adhesion of the developed photoresist cylinder to the substrate [Jay92].

The scanning electron microscope photograph of the A5 photoresist microlens array is shown in fig. 7.7.
<table>
<thead>
<tr>
<th>Type</th>
<th>Diameter (µm)</th>
<th>Initial thickness (µm)</th>
<th>Sagittal height of microlens (µm)</th>
<th>Melting temperature (°C)</th>
<th>Melting time (min)</th>
<th>Melting process</th>
<th>Focal length (µm)</th>
<th>f/#</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5</td>
<td>120</td>
<td>6</td>
<td>8</td>
<td>200</td>
<td>10</td>
<td>oven</td>
<td>360</td>
<td>3</td>
</tr>
<tr>
<td>A5</td>
<td>120</td>
<td>6</td>
<td>8.3</td>
<td>210</td>
<td>3</td>
<td>Hot plate</td>
<td>300</td>
<td>2.5</td>
</tr>
<tr>
<td>B4</td>
<td>240</td>
<td>12.7</td>
<td>16</td>
<td>170</td>
<td>3</td>
<td></td>
<td>112</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Table 7.2 Summary of the fabrication conditions of the microlens arrays made

Fig. 7.7 SEM photograph of the A5 photoresist array

### 7.4 Imaging quality of the microlenses

The imaging quality of the microlenses fabricated were compared to the NSG microlenses using a simple microscope arrangement (Fig. 7.8). The plan microscopic views of the photoresist microlens array (A5) and the NSG lens array are given in figs. 7.9 (a) and (b) respectively. Although the photoresist appears in a pale green colour, it is reddish brown in reality. To perform a multiple imaging experiment, a resolution target is placed just on top of a white light source and the microlens is placed at the sample holder about 15 cm away. The microscope objective is then focussed onto the image plane of the microlenses (about a focal length away from the microlens array) where multiple images form, one from each microlenses (fig. 7.9). When one of these images is magnified to fill the aperture of the
objective lens, the resolution of the microlens can be determined by finding the finest resolvable line in the image of the resolution target. It is important to note that the resolution of all the equipment used in the system have a higher resolution than the microlenses so that the resolution is limited solely by the microlens under test. We observe that the NSG microlens can resolve no. 36 in the resolution target by eye, which corresponds to a resolution of about 35 μm (14.3 lp/mm). For the photoresist microlens, severe degradation of image quality is observed due to the absorption of visible

![Fig. 7.8 Microscope setup for imaging experiment to test resolution of microlenses](image)

![Fig. 7.9 Microscopic plan view of the (a) photoresist A5 (b) NSG microlenses](image)
wavelengths giving it a reddish appearance. This observation is not an accurate measurement of the quality of the microlens and interferometric measurements are required to determine its resolution precisely (which are described in the next section).

Fig. 7.10 Multiple images of the resolution target observed at a focal length away from the microlens array for (a) photoresist microlens array (b) NSG microlens array

Fig. 7.11 Magnified version of one of the images for (a) photoresist and (b) NSG microlens array
7.5 Interferometric measurements of the microlenses

Although there are a number of different techniques to assess the imaging qualities of microlenses [Hut91a, Dal93a, Lin93, Hal92], measurements by an interferometric method can give wavefront aberrations, point spread functions and modulation transfer functions all at the same time. One of the most convenient configurations to implement an interferometer is that of the Mach Zehnder [Hut91a, Dal93a].

7.5.1 Mach-Zehnder interferometer

The experimental configuration of the Mach Zehnder interferometer is shown in fig. 7.12. All the measurements were conducted at National Physical Laboratory with generous help from Dan Daly and Prof. Mike Hutley.

![Schematic of the Mach Zehnder interferometer](image-url)

Fig. 7.12 Schematic of the Mach Zehnder interferometer used for measurements

An unexpanded HeNe laser beam is divided by a beam splitter into two paths. One beam passes through the working laser arm in which the microlens array under test is situated while the other passes through the reference optics. Because the beam is unexpanded, it is designed to illuminate one microlens so that only one lens is tested at a time. The microlens array is mounted on a rotation stage so that off-axis measurements can be made. The light from the microlens is focused to the same point as the beam coming through the reference arm and combined by another beam splitter. The interference between these two beams is
then imaged to a CCD camera which sends the signal to an 8-bit frame grabber. If the reference lens (or lenses) is assumed to be an ideal lens, the difference of the two waves from the fringes should give the wavefront aberrations of the microlens.

The mirror in the working arm is mounted on a piezo-electric stage which is under computer control to provide phase stepping. The phase stepped fringes are analysed by WYKO Phase II software to give the optical path difference (OPD) between the two beams. An 8-term Zernike polynomial is then calculated to fit the phase profile and the Seidel coefficients of the aberrations are calculated from the polynomials to give the spherical aberration, coma and astigmatism can be determined. From these data, the point spread function (PSF), encircled energy (which is the normalised encircled energy at a radial distance away from the focus) and modulation transfer function can also be calculated and plotted.

7.5.2 On-axis aberrations measurements

(a) Measurements on photoresist microlenses

The measurements of a microlens from the array A5 are shown in Fig. 7.13. The fringe pattern (fig. 7.13a) is calculated to give the optical path difference plot (fig. 7.13b) The surface roughness on the OPD plot is due to air disturbances and the accuracy of the system is about 0.1\( \lambda \). From the PSF plot (Fig. 7.13c), we can see that it is a good quality lens with a Strehl ratio of 0.88. The Strehl ratio is defined as the ratio of the maximum intensity of a point spread function to that when it is diffraction limited. The ratio is 1 when diffraction limited. The resolution of a microlens can be obtained from the spot size of the PSF plots. The MTF is the Fourier transform of the PSF and gives an indication of the resolution of the microlens. The MTF plot (Fig. 7.13d) represents the modulation transfer function of the four cross-sections of the 2-D point spread functions. The encircled energy at a certain radius from the maximum intensity gives efficiency of the light falls within a circular spot of that radius. This is a useful design curve for determining the pixel size and the loss of light outside the pixel. The Seidel coefficients of the aberrations are given in table 7.3.
To compare with the A5 microlens, we have also tested microlenses fabricated on a pyrex discs (C5) and the results are presented in fig. 7.14. This microlens has much higher aberrations, especially the spherical aberration. This may be due to the steeper contact angle as explained in section 7.3. The Strehl ratio is also very poor (about 0.1).

The measurement results (Fig. 7.15) of B4 microlens show that the microlens made from a thick layer of photoresist is of bad quality due to the slow spin speed. The comparison of the aberrations for the different arrays are given in table 7.3.

(b) Measurements of NSG GRIN planar microlenses

We have also measured the quality of the NSG microlenses (Fig. 7.16). There is a dip of the optical phase profile which gives rise to poor results on the PSF and the MTF plots. We have also examined the surface profile of these microlenses by a reflectometer at NPL and found that a small bulge exists on the surface of the microlens which gives rise to a phase difference of one wavelength (fig. 7.17). This is confirmed by an NSG representative [Wong93b] and subsequent papers on convex GRIN microlens [Sas92]. Although these microlenses do not have large numerical apertures, nevertheless, swelling takes place on the surface from the fabrication process. (refer to chapter 3). For optimum use, these microlenses need to be apertured or stopped down.

<table>
<thead>
<tr>
<th>Type</th>
<th>f/#</th>
<th>Dia. (μm)</th>
<th>Theoretical spot size (μm)</th>
<th>Strehl ratio</th>
<th>phase variation (λ)</th>
<th>Measured spot size (μm)</th>
<th>S.A. (μ)</th>
<th>Coma (μ)</th>
<th>Astigmatism (μ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A5</td>
<td>2.5</td>
<td>120</td>
<td>3.9</td>
<td>0.88</td>
<td>0.68</td>
<td>4.1</td>
<td>0.39</td>
<td>0.20</td>
<td>0.13</td>
</tr>
<tr>
<td>C5</td>
<td>3</td>
<td>120</td>
<td>4.6</td>
<td>0.10</td>
<td>1.4</td>
<td>11.5</td>
<td>3.8</td>
<td>1.3</td>
<td>0.17</td>
</tr>
<tr>
<td>B4</td>
<td>0.5</td>
<td>240</td>
<td>0.8</td>
<td>0.08</td>
<td>1.65</td>
<td>3.1</td>
<td>1.7</td>
<td>0.44</td>
<td>1.3</td>
</tr>
<tr>
<td>NSG</td>
<td>2.2</td>
<td>250</td>
<td>3.4</td>
<td>0.05</td>
<td>2.22</td>
<td>13</td>
<td>3.4</td>
<td>0.84</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Table 7.3 Comparison of the quality and the aberrations of the different microlenses

(S.A. stands for spherical aberrations)
(c) Comparison of imaging quality

The ideal diffraction limited spot size (diameter) is given by $2.44 \lambda f/\#$ and is included in table 7.3 for comparison. The A5 microlens has almost diffraction limited spot size, indicated by the good Strehl ratio of 0.88. This correspond to a variation of 0.68 $\lambda$ in the wavefront which is mainly spherical. The NSG microlenses, which has similar $f/\#$, has a Strehl ratio of 0.05 (i.e. a variation of 2.22 $\lambda$ in the wavefront). Therefore, the measured spot size of the NSG microlenses are nearly four times as big as that of the A5 microlenses. In addition, the primary Seidel astigmatic aberration coefficient is 1.9 times larger, the comatic aberration coefficient is 4.2 times larger and the spherical aberration coefficient is 8.7 times larger for the NSG microlenses than for the photoresist microlenses.

Fast microlenses (with small $f/\#$) have smaller spot sizes and therefore B4 microlens has a smaller spot size than the A5 microlens. However, the measured spot size for the B4 microlens is almost four times bigger than when diffraction limited (or a variation of 1.65 $\lambda$ in the wavefront) and this gives poorer resolution in imaging. In addition, all the aberrations of the B4 microlens is worse than those of the A5 microlens.

Both A5 and C5 microlenses have the same diameter but the C5 microlens has longer focal length and, thus, larger $f/\#$. The measured spot size is 2.5 times bigger than the ideal case which corresponds to $1.4 \lambda$ variation in the wavefront. With the exception of astigmatism, the aberrations are worse than those of the A5 microlenses. We conclude that the A5 microlenses ($\phi$ 120 $\mu$m) have the optimum $f/\#$ to give best optical quality in terms of aberrations and resolution.

7.5.3 Off-axis aberration measurements

The microlens array is mounted on a rotation stage in the Mach Zehnder interferometer and by positioning the array at an angle to the laser beam, off-axis aberration measurements can be made. Two off-axis aberrations, namely, coma and astigmatism are measured. Field curvature, another important off-axis parameter, is not included in the Zernike fit. Thus, some of these results cannot be directly compared to other results obtained from ray-tracing
Fig. 7.13 Interferometric measurements on A5 microlens
Microlenses: Design, fabrication and performances

(a) Wavefront aberrations plot  
(b) Point spread function plot  
(c) Modulation transfer function plot

Fig. 7.14 Measurements on B4 microlens

Fig. 7.15 Measurements on C5 microlens
Fig. 7.16 Interferometric measurements on the NSG microlens (φ 250 μm, f = 560 μm)

Fig. 7.17 Reflectometer result showing the evidence of surface profile on a NSG microlens
Fig. 7.18a shows the comparison of the two aberrations for a A5 microlens at different off-axis angles. As expected, astigmatism is more important at large angles. Spherical aberrations are not significant off-axis.

Fig. 7.18  
(a) Dependence of coma and astigmatism on the off-axis angle for A5 microlens 
(b) Comparison of astigmatism for A5 and C5 microlenses

Astigmatism between the microlenses made on the glass (A5) and pyrex (C5) substrates (Fig. 7.18b) is also compared. Both astigmatism curves resemble a quadratic relationship with the off-axis angle. Indeed, it has been shown that astigmatism varies with the square of the off-axis angle [Ham90]. We made quadratic polynomial fits to both data curves (Fig. 7.19) and the C5 microlens has a slightly larger quadratic coefficient (0.0014) than that of A5 microlens (0.001). This quadratic coefficient correspond to astigmatism.
7.6 Long focal length microlenses

7.6.1 Introduction

For the photoresist microlenses, the range of f-numbers achieved are typically from f/1 to f/4, which means that the longest focal length possible is only four times the diameter of the microlens. For example, the longest focal length for a 120 μm diameter microlens is 480 μm and the focus falls inside the pyrex substrate (thickness is 1.15 mm). In order to overcome this limitation and give flexibility in the design of focal lengths and, hence, the design of the f-numbers for microlenses, we can immerse the array in fluids having different indices and enclose the photoresist / index-matching fluid hybrid lens with a cover glass [Dal91a] (Fig. 7.20). This method cannot be applied to NSG graded-index microlenses because of the lack of surface relief structures. In general, long focal length
Microlenses also offer less spherical aberrations and reduced astigmatism, as will be shown later in this section.

![Fig. 7.20 Structure of long focal length hybrid microlens](image)

### 7.6.2 Design and fabrication

In order to design long focal length microlenses using index-matching fluid, the relationship between the original and overall focal lengths must be determined. From Fermat’s Principle and thin lenses [Hec87]:

\[
\frac{n_2}{f} = (n_1 - n_2)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{7.1}
\]

where

- \( n_1 \) - refractive index of the lens medium
- \( n_2 \) - refractive index of the medium surrounding the lens
- \( R_1, R_2 \) - radii of curvature of the lens
- \( f \) - focal length of the lens

Since \( R_2 \) is infinite for a plano-convex lens, the equation is simplified to:

\[
\frac{1}{f} = \left(\frac{n_1 - n_2}{n_2}\right)\frac{1}{R} \tag{7.2}
\]

By referring to fig. 7.21 and using equation (7.2), the focal length in air is then given by:

\[
f = \frac{R}{n_1 - 1} \tag{7.3}
\]

If the lens is surrounded by a medium with index \( n_2 \), the focal length \( (f_{n_2}) \) is now:

\[
f_{n_2} = \frac{n_2R}{n_1 - n_2} \tag{7.4}
\]
In the case of a hybrid microlens and assuming an air-lens interface on the plane surface of the lens, the light rays leaving the microlens travel faster compared to the previous case. Therefore, they come to focus faster and the focal length is shorter by the fraction \( \frac{n_1}{n_2} \) or simply \( \frac{1}{n_2} \):

\[
 f' = \frac{R}{n_1 - n_2}
\]  

(7.5)

By combining equations (7.3) and (7.5), we get the relationship between the overall \( f' \) and original focal length \( f \):

\[
 f' = \frac{f(n_1 - 1)}{n_1 - n_2}
\]  

(7.6)

Fig. 7.21 Diagram for the derivation of the relationship

The possible focal lengths range from the original focal length when there is no fluid to infinity when the fluid matches the index of the lens exactly. In our assumptions, we have ignored the substrate and this has no effect on the design of the focal length.

The relationship from this simple approach can be verified by a more rigorous method that is based on the fact that the phases of the different rays going through the lens and then coming to the focus must be the same.

In order to design different focal lengths, the refractive index of the photoresist \( n_1 \) must be known. The measurement of the refractive index of the photoresist is difficult as it changes during the fabrication and gradually in time as the solvent in the resist evaporates. Measurements of \( n_1 \) using an ellipsometer give an inaccurate reading of 1.6. Daly uses a refractometer which measures the refracting angle and determines the index to be 1.618 ±
0.001 after spinning, baking, exposure and development. However, after melting at 160°C for 25 minutes, the index changes to 1.642 ± 0.001. These melting conditions are not exactly the same as those we used to fabricate the microlenses but they give a good indication of the extent of the increase in refractive index after melting.

7.6.3 Focal length measurements

We conducted a series of focal length measurements with hybrid index matching fluid/photoresist microlenses of different focal lengths using the experimental apparatus shown in fig. 7.22. This is a vertical setup and the beam of a HeNe laser is directed upwards by a beam splitter. The laser beam is unexpanded so the speckle problems inherent in a coherent experiment are reduced. Only a few microlenses are illuminated by the narrow beam and they are imaged by a magnifying lens to a CCD camera. The signal from the CCD camera is sent to a monitor and a laser beam analyser with real time line-scanning on both x, y directions. First, the position of the translation stage is recorded when the rim of the microlens is imaged (Fig. 7.23c). Then, the microlens is moved until the focus is found (Fig. 7.23d) and the distance travelled is taken to be the effective focal length. As the focal length increases, the focus is no longer a point and we obtain the same size circle of least confusion over a range of focal length values. So we locate the entire range and take the
midpoint of it and put the appropriate error bars in the plot. We plot values of $1/f'$ against refractive index $n_2$ (fig. 7.24) and from equation (7.6), the x-intercept of the curve will give the index of the resist ($n_1$). If we fit a best straight line to the points, we find that the photoresist refractive index is about $1.64 \pm 0.02$. This agrees well with the result from the refractometer within experimental errors.

\[ f' = \frac{m_1 - m_0}{m_1} \]

Fig. 7.23 Procedure for measuring focal lengths:
(a) & (c) The rim of the microlens is imaged
(b) & (d) The focus is imaged and is also determined using a laser beam analyser
As the focal length increases, the spot size also increases until its size is of the order of that of the microlens. Diffraction becomes more pronounced as the index of the fluid nearly matches that of the lens. This causes the inaccuracy of the model derived from geometric optics. Therefore, in practice, the focal length achieved is less than expected from the equation (7.6) and the straight line fit is found to be less accurate at longer focal lengths.

\[
\frac{1}{f'} = 6.6864 - 4.082n
\]

![Graph showing the relationship between focal length and refractive index of index-matching fluid.](image)

Fig. 7.24 Focal length plot with index of fluid

A diverging photoresist microlens was also demonstrated by using an index-matching fluid with refractive index larger than that of the photoresist.

### 7.6.4 Interferometric measurements

We fabricated several long focal length microlens arrays with a fluid of index 1.60 ± 0.002. The standard array for testing was that of C5 which gives an overall focal length of about 4.7 mm and the microlenses are f/39 lenses. The testing of the hybrid microlenses are made on the Mach Zehnder interferometer.
(a) Quality of the microlens: point spread functions

The point spread function (Fig. 7.25a) of the hybrid lens on-axis is almost diffraction limited (Strehl ratio = 0.92). This is a significant improvement on the quality of the original microlens without index-matching fluid (Strehl ratio = 0.10). At 40° off-axis, the maximum angular position in the experimental configuration, there is an emergence of an asymmetric side lobe. However, the Strehl ratio is still acceptable (0.57), compared to 0.01 for the microlens without index matching fluid. The comparison is made between different microlenses in the array because of the difficulty in locating the same lens. However, the uniformity within the array is good and, therefore, the comparison is valid. The spot sizes of both foci are of the order of the diameter of the microlens and the microlens behaves rather like a light waveguide.

(b) Angular dependence of astigmatism

We have also investigated the variation of astigmatism with the off-axis angle. The plot of astigmatism at various angles is shown in Fig. 7.26. If we now compare these results with those for a microlens without index-matching fluid, we can see that the long focal length hybrid microlenses are better at angles greater than 10°, otherwise they are comparable within the experimental error. The measurements for the hybrid lens cover only one off-axis side because we expect the results for the negative angles is very similar.
Fig. 7.26 Comparison of the off-axis dependence of astigmatism between the microlens with and without index-matching fluid (ϕ 120 μm)

We also made an array of f/13.7 microlenses from the array B4 with focal length of 3.28 mm. The astigmatic aberrations of these microlenses vary with the angle in a quadratic manner (Fig. 7.27).

Fig. 7.27 Comparison of the off-axis dependence of astigmatism between hybrid microlenses made on (a) C5 microlens (f/39) (b) B4 microlens (f/13.7)
(c) Dependence of aberrations on focal length

The previous comparison between the hybrid microlenses from the C5 and B4 arrays suggest there is a dependence of astigmatism on focal length. Indeed, the astigmatism (in $\lambda$) can be shown to be (Appendix III):

$$A = \frac{d^2 \theta^2}{8f\lambda}$$  \hspace{1cm} (7.7)

where $d$ - diameter of the microlens
$q$ - off-axis angle
$f$ - focal length

Astigmatism is proportional to $1/f$ at a given angle, so the larger the focal length, the smaller the aberration. A plot of astigmatism against the focal length is shown in fig. 7.28a with the aberration normalised to that of the conventional microlenses. The worst case, at 40° off-axis, is taken and we find that the experimental result is in good agreement with the theoretical curve.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig7.28}
\caption{(a) Dependence of Astigmatism on focal length
(b) Dependence of spherical aberrations on focal length}
\end{figure}

In addition, we have measured the on-axis spherical aberrations (SA) of the lens and find that the large aberrations seen in the uncoated case (SA = 3.9 $\lambda$) become
negligible, i.e. within the experimental error of the measurement, for the longer focal length lenses. This is again expected as the SA coefficient varies as $1/f^3$ (Appendix IV) (Fig. 7.28b).

7.6.5 Imaging experiment using an apertured long focal length microlens

In our correlator system, each long focal length microlens images the entire input plane to the output, resulting in overlapping of the images at the output but this cannot be easily observed. Therefore, an opaque chrome mask was made to conceal the entire microlens array except for one microlens. This microlens, which is shown in fig. 7.29a, has a diameter of 240 μm and focal length of 1.1 mm ($f/# = 4.6$). The index-matching fluid of index 1.60 was applied and the microlens was enclosed with a cover glass. The resulting focal length is about 15.4 mm and the new $f/#$ is 64. This long focal length microlens is shown in fig. 7.29b. This is the longest focal length microlenses that we made.

An experiment (fig. 7.30) was set up to investigate the imaging properties of this apertured long focal length microlens. The microlens was situated at $2f = 3$ cm away from the input plane on which is placed a resolution target. The CCD camera is placed at 3 cm

![Figure 7.29](image.png)

Fig. 7.29 Photograph showing (a) an apertured photoresist microlens and (b) after the index matching fluid has been applied.
away from the microlens and so the system is set up in a unity magnification imaging configuration. The illumination is provided by a LED and a diffuser. Only part of the resolution target can be viewed because of the mismatch in size between the image of the resolution target and the CCD chip. The finest line that can be resolved is no. 31 on the image of the resolution target. This correspond to 62.5 μm. The diffraction limited resolution was calculated to be 51.7 μm. This shows that an almost diffraction limited spot size can be obtained with long focal length microlenses illustrating their lack of aberrations when used in a symmetric imaging configuration.

![Fig. 7.30 Experimental apparatus for an unity magnification imaging experiment](image)

The microlens was then replaced by a single pinhole with the same. The image was dimmer than before but the resolution is the same (resolvable line is 31). This agrees well with the predictions of the theory in chapter 5. Therefore, we conclude that the efficiency of the system can be enhanced by using a long focal length microlens array.

### 7.7 Conclusions

We fabricated several arrays of photoresist microlenses by deposition of the resist, u-v lithography and reflow. The array designs are on a electron beam mask made by RAL. There are many factors which can affect the final quality of the microlenses. We find that the microscope glass slide is a better substrate and the uniformity of the array can be best achieved by high spin speeds of the photoresist. However, this often results in a thin layer of photoresist which does not have enough surface tension to draw into a convex shape
when melted. Consequently, a sag appears in the centre of the lens. The effect of melting time and temperature has been investigated and we conclude that reflow at 200° C for 10 mins is the best condition. The melting environment is another factor in fabrication of good microlenses. Furthermore, only short focal length microlenses (f/1 - f/4) can be made.

We have also measured the optical quality and aberrations of the microlenses using a Mach Zehnder interferometer. We can conclude that the photoresist microlens has better optical qualities than the latter. This is due to the presence of the surface profile on the NSG microlenses caused by swelling. Our best photoresist microlenses have a Strehl ratio of 0.88. The off-axis astigmatic aberrations are also investigated and we make a good quadratic fit to the relationship between astigmatism and off-axis angle.

To obtain long focal length microlenses from conventional photoresist microlenses, a layer of index-matching fluid is enclosed on top of the array. The difference between the indices of the fluid and the photoresist gives an overall longer focal length. The smaller the difference, the longer the focal length. We also have to determine the refractive index of the photoresist, which is found to be 1.64 ± 0.02 by focal length measurements. The quality of the long focal length microlenses is good in both on-axis and off-axis measurements. The point spread function is almost diffraction limited and the aberrations are minimal. These are expected in long focal length lenses as astigmatism and spherical aberrations vary with 1/f and 1/f^3 respectively. The only drawbacks are that the spot size is approaching the size of the lens and the range of focal length becomes very long.
8.1 Introduction

This chapter describes the final choice of design parameters for our incoherent correlator system. First, it starts with the computer simulation of the system under real constraints (Section 8.2). These simulation results not only model the performance of the system in practice, they also helped us to choose the test patterns for the actual experiment. In section 8.3, the effect of the SLM in the system performance is investigated. In section 8.3.2, we analyse the reason for the lack of contrast. Simple correlation experiments without the use of the SLM are then demonstrated and compared with the simulation results (section 8.4). In section 8.5, the final experiment with the SLM in place is described.
8.2 Computer simulations of the correlator performance

8.2.1 Introduction

In this section, the correlator system is simulated using a computer. Since the shared microlens correlator behaves in a very similar way to the shadow casting correlator, the simulation results will apply to both systems. The different techniques for the simulation are first described and results are shown for correlation of binary patterns with 10:1 contrast ratio similar to that of our FLC-SLM. It will be followed by how the resulting signal-to-noise ratio can be improved using bipolar and randomised patterns. These results also helped us to generate masks for the input and filter planes to give more easily observable results.

8.2.2 Simulation techniques

In chapter 4 we showed that correlation can be achieved in two domains: - real-space and Fourier-space. Similarly, computer simulation of the system can also be achieved digitally in both domains. In the Fourier-space domain, Digital Fourier Transforms are involved and an algorithm called the Fast Fourier Transform (FFT) is commonly used. A number of steps is required to prevent the aliasing of the data due to sampling. For a symmetric signal \( f(p) \) in 1-D which is represented by \( N \) pixels, the signal is repeated every \( N \) pixels in real space due to the sampling in the Fourier space (fig. 8.1a). Usually, only the positive part of the signal is taken to avoid negative indexing in a computer. Since the pattern repeats, \( f(N - p) = f(-p) \) and the negative part can be taken into account. Then this rearranged signal undergoes a FFT which gives the frequency spectrum in a similar manner, i.e., the negative half is shifted by \( N \) pixels. After shifting back the negative half of the signal, the correlation signal in Fourier domain \( G(\omega) \), is given by:

\[
G(\omega) = F(\omega)H^*(\omega) = F(\omega)H(\omega)
\]

(8.1)
since the filter \( h(p) \) is a real and symmetric function. Then the pattern is split once again in a similar manner before the inverse FFT is operated and the two halves are recombined afterwards.
For a 2-D pattern, the splitting and shifting operation is done similarly and is described in fig. 8.1(b).

In the real-space simulation, the digital correlation is achieved by the discrete form of the correlation equation, which involves the shifting and multiplication operations. For a 2-D pattern, the intensity at the output pixel \( O(p,q) \) is given by:

\[
O(p,q) = \sum_{m=1}^{N} \sum_{n=1}^{N} I(m,n)T(m+p,n+q)
\]  

where \( I(m,n) \) is the intensity at the input plane and \( T(m,n) \) is the transmittance at the filter plane. Since the shared microlens correlator operates in real-space, we use the latter simulation although both give the same results.

Fig. 8.1 The split and shift operation before and after the FFT in (a) one dimension and (b) two dimensions
8.2.3 Simulation results of binary patterns

The correlator system is simulated by using Matlab\textsuperscript{†} programming on a Sun SPARC Unix workstation (Appendix IV). Since the FLC spatial light modulator has binary amplitude, the input and mask (filter) patterns can take values of either 1 or 0. When the contrast ratio of the SLM under LED illumination (10:1) is taken into account (described in chapter 6), the values of a pixel will be either 10 (if turned on) or 1 (if it is off). The effect is that the resultant correlation pattern will be scaled by 10 times. A pattern which shows an inverted letter ‘A’ is used as the filter (fig. 8.2a). The inversion is due to the software representation of matrices. The input pattern depicts two displaced letter ‘A’s and a letter V (fig. 8.2b). The patterns are assumed to be displayed on a SLM. If the SLM has infinite contrast, the correlation pattern of these is given in fig. 8.2c In reality, the contrast ratio is only 10:1 and the resulting correlation is shown in fig. 8.2e in which it is harder to distinguish the correlation peaks from the background.

Each correlation pattern between the letters is symmetrical as expected. The autocorrelation patterns (the correlation of the letter ‘A’ with itself) feature a cross with a horizontal stroke through the centre bright spot. The diagonal of the cross signifies the correlations of the diagonals of the letter ‘A’ and the horizontal line is the result of the correlation of the corresponding horizontal stroke in the letter. These are the sidelobes which reduces the signal-to-noise ratio. The cross correlation between the letter ‘A’ (inverted) and ‘V’ just gives a cross with a lower correlation peak.

One of the useful parameters in a correlation is the signal-to-largest sidelobe ratio (SLSR) in 2-D and it is found to be 3.2 in the infinite contrast case (from fig. 8.2d). For the correlation from patterns with CR of 10:1, the SSLR is reduced to 1.6 (fig. 8.2f). If a linescan is run across the correlation peaks as shown in fig. 8.3a, the signal-to-sidelobe ratio (SSR) along that scan is 15:1 for infinite contrast and 2.2:1 for 10:1 CR. This line does not necessary include the largest sidelobe but will serve as a convenient point for comparison.

In the case of the patterns with CR of 10 : 1, the low signal-to-noise ratio observed in the correlation is due to in the leakage of a tenth of the intensity of the bright pixels (those turned on) through the supposedly dark pixels (those which are off).

\textsuperscript{†} Matlab v.4.2 (by Mathworks Co.) is mathematical software for scientific calculations and graph plotting
Fig. 8.2 The correlation between patterns displayed at the (a) input plane and (b) the filter plane is given in (c) for infinite contrast ratio and (e) for 10:1 contrast ratio. (d) and (f) show the corresponding 3-D mesh diagrams of the correlation results.
Consequently, there is a larger degree of correlation between the dark pixels which gives the high background sidelobe (noise).

Fig. 8.3 (a) The contour plot showing the correlation when the contrast is infinite
(b) Line scan through the correlation peaks along the dotted line in (a)

Fig. 8.4 (a) The contour plot showing the correlation when the contrast is 10:1
(b) Line scan thorough the correlation peaks along the dotted line in (a).
8.2.4 Simulation of bipolar patterns

There is a lot of background noise in fig. 8.2e which is a d.c. term in the spatial frequency domain. The SSLR can be improved if this d.c. term is filtered off. This can be achieved by using bipolar representations of the patterns. For a CR of 10:1, the bipolar values (+10, -10) are represented by two binary patterns. The first input pattern holds the usual pattern and the second input pattern is the inverse of the first with pixel values of -10 if previously 10, and -1 for 1 (fig. 8.5). The two input patterns are correlated with the filter pattern in turn. The positive correlation between the first input pattern and the filter gives the same result as before (fig. 8.2e) The correlation between the inverted pattern

Fig. 8.5 The input and filter patterns for (a) Positive correlation and (b) negative correlation
(Positive: Black = 1, White = 10; Negative: Black = -1, White = -10)

Fig. 8.6 The negative correlation (c) between patterns (a) and (b)
and the filter is shown in fig. 8.6. As expected in this negative correlation, the features of peaks previously, have become troughs and the whole pattern is inverted with a large d.c. correlation term in the middle. Since the number of white pixels in the inverted input pattern is greater than that in the original input, simple subtraction of the two patterns does not give a meaningful result. The correlation peaks must be scaled to the same size before the subtraction operation. The number of white pixels in the correlation pattern is given by the size of the correlation peak. The peak values of the positive and negative correlations are found to be 3978 and 12076 respectively. This means the negative correlation pattern must be scaled down to 33% of its original value. Practically this could be done by reducing the illumination intensity to allow easier subtraction. The resultant correlation after the subtraction of the two correlation patterns is given in fig. 8.7.

![Fig. 8.7](a) The correlation pattern after subtraction and (b) its mesh diagram

The centre of the pattern is largely in the negative region and the rim is about zero. The rim can be removed by placing an aperture at the output plane but this limits the field of view for the recognition. If the aperture is placed as shown in fig. 8.8a, then the SSLR is found to be 2.6:1 while the line scan SSR is 7.1:1 (fig. 8.8b) In each case, there is improvement of the original figures.
This bipolar representation can be implemented spatially or temporally. If the space-bandwidth product is large and speed is paramount, then the pattern and its inverted version are placed side by side in the SLM so that the two correlations proceed in parallel. Otherwise, the patterns can be displayed one after another. The final result is obtained by subtraction in the computer.

Fig. 8.8 (a) The contour plot of the correlation pattern after subtraction  
(b) Line scan crosses the correlation peaks along the dotted line within the aperture  
(c) Correlation pattern and (d) 3-D mesh pattern within the aperture
8.2.5 Simulation of randomised patterns

The disadvantage of the bipolar technique is that when there is an imbalance of 'bright' and 'dark' pixels, the d.c. terms in the middle will not simply cancel each other. This problem can be solved by using either orthogonal or randomised patterns which have the same number of bright and dark pixels. Correlations of orthogonal patterns give better discrimination and lower sidelobes are obtained when using randomised patterns. For example, the pattern which we use in the experiment described later in section 8.4.3 is a resolution target which contains features of many spatial frequencies at various spatially separated locations and can be considered as a special type of 'random' pattern i.e. they have high space bandwidth product.

In this simulation, we use a pattern which is generated by a random number generator in the Matlab program and is then normalised to unity. If the pixel value is above 0.5, it is considered having a value of '1' (white pixels), otherwise it takes the value of '0' (dark pixels). The subsequent correlation of these patterns after subtraction of the negative from the positive patterns will have a very high SLSR. For example, a 10:1 CR randomised pattern is used for the simulation (fig. 8.9a) and the correlation result after subtraction is given in fig. 8.9b. The SLSR is found to be 10:1 which is same as the contrast ratio of the original patterns.

Fig. 8.9 (a) A randomised pattern with 10:1 contrast ratio
(b) The autocorrelation pattern of the randomised pattern in mesh form
8.3 Ferroelectric liquid crystal spatial light modulator

8.3.1 Introduction

In our system, both the input and the mask patterns can be displayed with an electrically addressable ferroelectric liquid crystal spatial light modulator (FLC-SLM). In this section, the properties of the spatial light modulator such as contrast ratio are investigated in greater detail. These will be used in the practical design of the final system.

The FLC-SLM is made by CRL, formerly STC. The SLM is an array of 128 x 128 pixels of pitch 165 μm with gaps of 5 μm in between pixels. The SLM is connected to a power supply and is interfaced to the parallel port of a computer. Ferroelectric liquid crystals are bistable, so only binary patterns can be displayed. Grey scale patterns can be achieved by either spatial multiplexing (using more than one FLC pixel to represent a data pixel) or by time multiplexing. The SLM is optimised for use at 633 nm, ie. with red HeNe illumination. In the next section, the factors which affect the contrast ratio are discussed with theoretical and experimental results.

8.3.2 Contrast ratio

The contrast ratio of the SLM depends on the transmission through the pixels in the 'bright' state and more importantly on the transmission through the pixel in the 'dark' state (the extinction efficiency). These are the main factors which the degrade the CR from that measured within the pixel (1000:1):

(a) Angular and wavelength responses

1. Operation of the FLC cell

In the chiral smectic C phase, the orientation of the ferroelectric liquid crystal molecules tilts away from the layer normal (fig. 8.10a) [Sta93]. In bulk material (thickness > 10 μm), they are packed in helical form and the molecules rotate around the cone and the
azimuth angles are changed cyclically under no bias voltage. However, in the SLM cell, where the thickness is small, about 2 μm, the molecules are allowed to occupy only two orientations in directions parallel to the substrate plane. This is known as surface-stabilised FLC. The molecules switch from one stable state to another under the application of electric pulses. The relationship between the optic axes and the polarization directions of light for a FLC molecule is shown in fig. 8.10b [Kur92]. Light polarised in the horizontal direction passes straight through the cell if the molecules are switched to the negative optic axis. A vertical polariser (or analyser) is placed after the cell which blocks the light and this is the ‘dark’ state. If the molecule is now switched to the positive optic axis, the horizontal polarised light can be resolved into the components along and perpendicular to the positive optic axis. Because of the birefringence of the FLC, the component along the axis travels slower (higher refractive index) than that perpendicular to the axis. If the thickness is set to give a half-wave plate condition, the difference in phase will be \( \kappa \). This causes the polarisation to be rotated of 90°. After the recombination of the components vertically polarised light is produced. This will be allowed through the analyser and the FLC cell is in the ‘bright’ state (Fig. 8.10c).

\[\text{Layer normal} \quad \text{azimuth angle} \quad \text{FLC molecule} \]

**Fig. 8.10** (a) Cone of angular positions occupied by a FLC molecule tilted from the layer normal
(b) diagram showing FLC optic axes and rubbing orientation
The optical properties of the FLC can be described by a Jones Matrix [Day91]:

\[
M = \begin{pmatrix}
\cos^2 \theta e^{i \delta/2} + \sin^2 \theta e^{-i \delta/2} & i \sin 2 \theta \sin(\delta/2) \\
\sin 2 \theta \sin(\delta/2) & \cos^2 \theta e^{-i \delta/2} + \sin^2 \theta e^{i \delta/2}
\end{pmatrix}
\] (8.3)

where \( \theta \) is the switching angle of the FLC from one state to another and \( \delta \) is optical path difference given by:

\[
\delta = \frac{2 \pi \Delta n d}{\lambda} \quad \text{(8.4)}
\]

where \( d \) is the thickness of the cell and \( \Delta n \) is the birefringence of the ferroelectric liquid crystals and is the difference between the two refractive indices \( n_o \) and \( n_e \). The Jones matrices for the crossed polarisers are \( \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \) in the \( x \)-direction and \( \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \) in the \( y \)-direction. The amplitude of the light after passing through the FLC cell is given by:

\[
\begin{pmatrix} U_x \\ U_y \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} M \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 \\ i \sin 2 \theta \sin(\delta/2) \end{pmatrix}
\] (8.5)

By using the notation of the angles in fig. 8.10a, the relative transmission is therefore:

\[
T = \left( \sin[2(\theta_{rub} \pm \theta_{tilt})] \sin \left( \frac{\pi \Delta n}{\lambda} \right) \right)^2
\] (8.6)

where \( \theta_{rub} \) and \( \theta_{tilt} \) are the rubbing and tilt angles respectively. In our SLM (thickness of the cell is 1.7 \( \mu \)m), the typical birefringence quoted from the manual is 0.18 [CRL]. This is unusually high as the typical value for BDH SCE13 is 0.15. The positive and negative optics axes correspond to the aligned molecule axes of the FLC under the application of positive and negative electric pulses respectively. The readout efficiency is maximised when:

\[
\frac{\pi \Delta n}{\lambda} = \frac{m \pi}{2} \quad \Rightarrow \quad \Delta n = \frac{m \lambda}{2}
\] (8.7)

where \( m \) is an odd integer. This is the half-wave plate condition. From the above data quoted, the optimised wavelength can be calculated to be 612 nm. However, the SLM is supposedly optimised at 633 nm and this would imply a birefringence of 0.186 (3.3 % error) for the given thickness or 1.76 \( \mu \)m (3.5 % error) for the given \( \Delta n \). The contrast is maximised when:
which means that the switching angle or the switching angle must be \( \pi/4 \) (45°).

2. Angular response

Assuming that the switching angle is at 45°, the relationship of the transmission efficiency \( T \) with the angle of incidence is (fig. 8.11a):

\[
T = \sin^2 \left( \frac{\pi \Delta n d}{\lambda \cos \phi_{LC}} \right)
\]

(8.9)

where \( \phi_{LC} \), the angle of incidence in the liquid crystal is related to the angle of incidence in air by Snell’s Law:

\[
\sin \phi_{LC} = \frac{\sin \phi_{air}}{n_o}
\]

(8.10)

In this calculations, we have ignored the reflection losses at the air-glass and glass- LC interfaces. This changes with the angle and the polarisation (TE or TM mode). The variation of \( T \) with \( \phi_{air} \) is plotted in fig. 8.11b where \( n_o \) is assumed to be 1.5. A shift of the maximum transmission efficiency from the on-axis position to about 34° is observed. This is because the cell is not optimised at 660 nm which is the wavelength of the LED used. For a 1.5% of degradation allowed in \( T \), the angle of incidence can vary from 0 to 50°. Therefore, we conclude that the angle of incidence has negligible effect on the

\[
d' = d / \cos \phi_{LC}
\]

\[
d = 1.7 \mu m
\]

\[
\Delta n = 0.18
\]

\[
\lambda = 660 \text{ nm}
\]
transmission efficiency of the SLM. In fact, the SLM was rotated at an angle up to 30° and the pattern could still be clearly viewed with little degradation.

3. Wavelength response

The relationship between the transmission efficiency and wavelength at normal incidence can also be evaluated from eqn (8.9) by putting $\phi_{LC} = 0$ and it is plotted in fig. 8.12. If the system is illuminated by an LED of central wavelength of 660 nm, its linewidth (about 50 nm) causes about 3% of variation of transmission efficiency. However, if a white light source is employed, the 300 nm bandwidth degrades the transmission by 55%. Therefore the white light source is ruled out from providing the illumination for the system. This explains the low contrast ratio obtained with a white light source in an experiment described in chapter 6.
Correlator system: Optimum system design, analysis, simulations and experiments

The pitch of the SLM pixels is 165 μm and the dead space is about 5 μm. In the worst case all of the dead space lets light pass when all of the pixels are in the dark state. Then the total leakage of intensity is proportional to the area of the dead space. When all the pixels are in the bright state, we assume that all of the dead space is switched to the opposite dark state. The intensity is then proportional to the area in the active area. The dead space area is given by twice the area A plus the area B (fig. 8.13).

The ratio of the active area to the dead space (d.s.) area gives the worst possible contrast ratio:

$$\text{Worst CR} = \frac{\text{Total active area}}{\text{Total dead space area}} = \frac{128^2 \times 160^2}{(2 \times 5 \times 160 + 5^2) \times 127^2} = 15.8 : 1 \quad (8.12)$$

In the case that all the dead space is not switched and lets the light pass through regardless of the state, the CR becomes:

$$\text{Contrast Ratio} = \frac{\text{Total area}}{\text{Total dead space area}} = \frac{128^2 \times 165^2}{(2 \times 5 \times 160 + 5^3) \times 127^3} = 17 : 1 \quad (8.13)$$

In reality, some pixels will switch to the required state. Suppose now, if half of them switch to the correct state and the rest remain at the wrong state, the CR is:

$$\text{Contrast Ratio} = \frac{128^2 \times 160^2 + [(2 \times 5 \times 160 + 5^2) \times 127^2]/2}{[(2 \times 5 \times 160 + 5^2) \times 127^2]/2} = 33 : 1 \quad (8.14)$$

From the experimental results obtained in chapter 6, of which the contrast ratio under coherent illumination gave the best result of 16:1. This indicate that the dead space that switches to the correct state is minimal assumed that the switching angle is 45°. In reality, the switching angle would be less and the contrast ratio of 16:1 is more likely a combinations of all the factors discussed in this entire section.

(c) Addressing scheme for the FLC SLM and the switching angle

There is a choice of two addressing cycles for our SLM. In the usual mode, the two scan cycle, the relevant pixels are first written in one state (positive scan) and then those in the
opposite state are written in the next scan (negative scan). Hence, two complete scans are required to change an image. At the end of the cycle, there is a pause period in which no pixels are written and a high frequency voltage $2f$ is applied where $f$ is the maximum operating frequency during writing (fig. 8.14). The pause length can be varied by programming.

The ideal switching angle for achieving good extinction of light and, hence, maximum contrast is $45^\circ$ (from eqn. (8.8)). This can be achieved by an application of a DC pulse. However, in the addressing scheme used, where an a.c. voltage is applied, the switching angle of the FLC is lower than this value. During the scanning, the operating frequency can be varied by a control knob on the optical head of the SLM. The largest switching angle for the maximum operating frequency, $f$, is quoted to be about $29^\circ$ by the manual (fig. 8.15a) [CRL]. Furthermore, the switching angle decreases with rising temperature. The rise of temperature may be due to the infra-red spectrum of the illumination, e.g. from a halogen-tungsten lamp in the fibre-optic source. The relationship between the transmission efficiency and the switching angle, $\theta$, can be calculated if the half-wave plate condition is satisfied (fig. 8.15b):

$$T = \sin^2 2\theta$$  \hspace{1cm} (8.15)

At the maximum operating frequency $f$, the contrast is degraded to about 62% @ 40°C from 84% @ 10°C of the maximum contrast achieved by $45^\circ$ switching angle. During the pause period, the frequency is doubled ($2f$) and the contrast is increased from 84% to 91% @ 10°C and from 62% to 74% @ 40°C. At room temperature, the switching angles for the normal scanning (at frequency $f$) and the pause period (at frequency $2f$) are 26° and 30° respectively and the corresponding relative transmission are 78% and 88%.

---

**Fig. 8.14** The diagram of two scan addressing cycle within a whole frame [CRL]
Therefore, to achieve higher contrast, the SLM must be read during the pause period only. This can be implemented by using synchronization of the pause period with a pulsing source.

![Graphs showing the dependence of the switching angle on temperature and frequency of the scanning voltage.](a)

Fig. 8.15(a) Dependence of the switching angle on temperature and frequency of the scanning voltage [CRL]

(b) Effect of the change in switching angle with temperature on the transmission efficiency of the FLC cell

An experiment (fig. 8.16a) was performed to measure the contrast ratio during the pause period. The SLM was illuminated by diffused, collimated LED light and the whole pattern was imaged by a lens onto a sensitive detector. The detected signal was observed on a digital oscilloscope. The frequencies of the writing signal and the pause signal were found to be about 2.6 kHz and 5.8 kHz. The contrast ratio can be measured by taking the ratio of the differences between the signal in the bright state and the noise floor and that between the dark state and the noise floor of the CCD (Fig. 8.16b). The pause lengths are varied and the contrast ratio results are given in table 8.1.
Table 8.1 Table of results for contrast ratio during a pause period

<table>
<thead>
<tr>
<th>Pause length</th>
<th>Difference between noise floor and bright state (b) (average value)</th>
<th>Difference between noise floor and dark state (d) (average value)</th>
<th>Contrast ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>84 mV</td>
<td>0.44 V</td>
<td>5.3 : 1</td>
</tr>
<tr>
<td>5</td>
<td>54 mV</td>
<td>480 mV</td>
<td>8.9 : 1</td>
</tr>
<tr>
<td>10</td>
<td>54 mV</td>
<td>400 mV</td>
<td>7.4 : 1</td>
</tr>
</tbody>
</table>

Fig. 8.16 (a) Experimental apparatus for measuring the contrast ratio during the pause period
(b) Diagram showing how the measurements are made and how the CR is obtained

The maximum contrast ratio is about 8.9:1 and is similar to previous measurements in chapter 6 within the experimental errors.

(d) Contrast ratio of the polariser

Leakage of light through the crossed polarisers in the dark state also reduces the signal-to-noise ratio. Although polarisers in a camera lens mount are provided with the SLM, small sheet polarisers are required to make the system more compact. A sheet of linear polaroid polariser, HN38, is obtained and it is cut into the same size as the area of the
SLM (about 21 mm square). At 660 nm, the specified transmission is 10% when two polarisers are parallel. The transmission when two polarisers is crossed (extinction) is specified to be 0.005% at the same wavelength giving a specific contrast ratio of 5000:1. An experiment is carried out to test the contrast ratio of these polarisers in our system (fig. 8.17). The system is illuminated by an LED and detection is made with a CCD camera. The image of the CCD is captured and the intensity is digitised by an 8-bit frame grabber. The intensity can take values from 0 to 255. The average frame values with different brightness is given in table 8.2. The CR is reduced at the lower end due to noise in the CCD and by saturation at the upper end. The dynamic range for maximum contrast is for 3-6 V across the LED and so we set the voltage in this range for our experiment.

<table>
<thead>
<tr>
<th>Voltage across LED (V)</th>
<th>Average value for polarisers crossed (dark state)</th>
<th>Average value for polarisers parallel (bright state)</th>
<th>Contrast Ratio</th>
<th>CR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3.3 (noise)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1.7</td>
<td>3.3</td>
<td>13.1</td>
<td>4:1</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>5.1</td>
<td>64.7</td>
<td>12.7:1</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>5.3</td>
<td>189.1</td>
<td>36.7:1</td>
<td>15.5</td>
</tr>
<tr>
<td>4</td>
<td>5.1</td>
<td>189.8</td>
<td>37.2:1</td>
<td>15.7</td>
</tr>
<tr>
<td>5</td>
<td>6.0</td>
<td>189.6</td>
<td>31.6:1</td>
<td>15</td>
</tr>
<tr>
<td>5.3</td>
<td>7.2</td>
<td>189.6</td>
<td>26.3:1</td>
<td>14.2</td>
</tr>
</tbody>
</table>

Table 8.2 Contrast ratio of the polarisers with different brightness of the LED at d = 10mm
8.3.3 Programming the SLM

The SLM can be directly controlled by writing data to the parallel port of an IBM compatible PC. This interface uses a handshaking protocol to transfer one byte at a time and the speed of the data transfer depends on the speed of the computer and that of the SLM. The computer used for controlling the SLM is a Compaq 386SX-20 computer, with the clock speed of 20 MHz. There are altogether 32 frames that can be stored in the memory of the SLM buffer. The data are sent in such a way that the rows of data interleave, i.e. all the odd rows are sent first followed by the even ones. Within the rows, four bits (each bit, 1 or 0, represents a pixel) are sent at a time. Therefore, there is a need to process and interleave the data which could be in the form of a ASCII text file. The program is written in C and the listing is given in Appendix V. An example pattern displayed on the SLM is shown in fig. 8.18.

![Fig. 8.18 Photograph showing an example pattern displayed on the SLM by programming](image)

8.4 Correlation system results without SLM

8.4.1 Introduction

Due to the low contrast ratio and efficiency of the SLM, several systems were implemented without the presence of the SLM in the first instance. The correlation
experiments using transparencies and resolution targets as the input and the filter planes are described in this section. These give some preliminary results for the correlation systems.

8.4.2 Correlation with transparencies under LED illumination

A simple experiment can be implemented by using transparencies as the input and the filter planes (fig. 8.19). The contrast ratio of a transparency is measured to be 10 : 1. However, the contrast ratio is not sufficient to produce a good result and it is improved by masking off all the black area with black tape. Two different diffusers were used, namely the plastic diffuser and the microlens array diffuser (they are described in chapter 6). The architecture of the Cohu CCD camera poses a serious constraint on the design of the system. It has a small detection area 6.4 mm × 4.8 mm and the C-mount housing in front has a length of 11 mm which prevents the filter being placed closer to the CCD chip. In this system, the correct choice of geometry for the system is paramount to observe the correlation. Any mismatch in size and distance between the filter will either prevent the correlation being viewed or only part of the pattern can be observed. Therefore, a pattern ‘o’ (fig. 8.20a) is used in this case as this would give small side lobes in the autocorrelation peak. The autocorrelation simulation of the pattern ‘o’ is given in fig. 8.20b. This shows that the correlation pattern is mainly a high intensity spot with little side lobes. The optical autocorrelation results are shown in fig. 8.20c and 8.20d using the plastic and microlens array diffusers respectively. This shows that both diffusers provide enough angular spread but a closer examination reveals that the microlens diffuser gives a less uniform correlation spot.
Correlator system: Optimum system design, analysis, simulations and experiments

Fig. 8.19 Experimental apparatus of a shadow casting correlator using transparencies

Fig. 8.20 (a) The pattern 'o' used on the transparency and (b) its autocorrelation simulation.
Optical correlation result with (c) plastic and (d) microlens array diffusers
8.4.3 Correlation with resolution targets under LED illumination

In this experiment, chrome on glass resolution targets are used as both the input and the filter. The resolution target provides a very high contrast and also a near randomised pattern with a range of spatial frequencies for better correlation results. Because the pattern on the resolution target repeats itself in different sizes, it is necessary to mask off part of the resolution target so that the input pattern is bigger than the filter pattern. This compensates for the demagnification of the object with distance (fig. 8.21). The resolution of the largest stripe (no. 02) at the input is 1.12 lp/mm and the corresponding linewidth is $1/(2 \times \text{Resolution}) = 0.446$ mm. At the filter plane, the corresponding stripe (no. 22) has resolution of 4.49 lp/mm or a linewidth of 0.111 mm. Therefore, the input is four (4.02) times bigger than the filter and the ratio of the distance between the input and the filter to that between the filter and the output is 3.02:1. The experimental setup is shown in fig. 8.22.

![Fig. 8.21 The patterns displayed on the (a) input and (b) filter planes](attachment:image.png)
The strongest correlation peak is detected when the distance between the input and the filter is 70 mm (Fig. 8.23a). Other peaks are detected at $d = 60$ mm (fig. 8.24a), 55 mm, 45 mm and also 80 mm. The simulation results are also presented as a comparison. The series of correlation peaks arise from the repetition of the peaks, the strongest being the exact match of the two patterns. When the distance decreases, the projection of the input on the filter is bigger. The second correlation happens when the second row of the input correlates with the first row and so on (fig. 8.25). As a result, the peak is shifted off-axis.

![Fig. 8.22 Experimental setup for a shadow casting correlator with resolution targets](image)

![Fig. 8.23 (a) Correlation of the masked resolution targets and (b) the corresponding simulation](image)
and has weaker strength. As the input moves further closer, the peaks get weaker and eventually some do not have enough SNR to be detected. Similarly, as the distance $d$ increases beyond the main correlation point, there are also a series of correlation peaks.

Fig. 8.24 (a) The second correlation peak at $d = 60$ mm and (b) the corresponding simulation

Fig. 8.25 Geometries of the system showing (a) the autocorrelation of the whole pattern and (b) the correlation of part of the pattern which gives rise to a series of correlation peaks

### 8.5 System Experiment with the spatial light modulator

#### 8.5.1 Experiment setup

One of the most important rules when building the system is to put the different planes at the corresponding locations constrained by the geometry of the system. The FLC-SLM is
placed inside an housing and some modifications of the system is required to achieve the geometry. The first attempt is to use an SLM as the input and a chrome mask such as the resolution target as the filter as it would be very difficult to put two SLMs with tens of millimeters apart (fig. 8.26a). The first modification requires the removal of the back casing of the SLM and the SLM active area is exposed so that the filter can be held closely after it. The front of the SLM is about 40mm away from the front casing which cannot be removed because of the built-in circuit boards. If the source is placed at the entrance to the front, only rays with small angles can reach the SLM and beyond. Therefore, the illumination source must be situated next to the SLM and two source designs have been built for that purpose. First, a clear perspex tube is used to guide the light from the LED which is placed at a side aperture at the back of the tube (fig. 8.26b). The perspex tube is painted with metallic silver colour to maximize the reflections within the tube. A diffuser is held on the exit of the tube to give an even distribution. Finally, a sheet polariser is fitted on top of the diffuser so that the whole illumination unit can then be placed next to the SLM cells. In the experiment, two LEDs are used although many more can be inserted around the cylindrical wall. The drawback of this illumination

---

![Diagram of SLM system](image)

**Fig. 8.26 (a)** Cross section of the FLC-SLM.

Illumination guide using *(b)* a perspex cylindrical tube and *(c)* an aluminium rectangular tube
system is that a considerable amount of light is lost and reflected at the air-perpex interface. Another idea is to use an aluminium rectangular tube which can house up to 4 LEDs (fig. 8.26c). This system is brighter and is used in the experiment.

8.5.2 Result of the Experiment

Although we tried very hard by varying the geometry of the system and the different input / filter patterns, we could not observe the correlation peak which was expected from the simulations. We concluded that the most serious problem was the lack of contrast of the two ferroelectric liquid crystal spatial light modulators and that this was so serious that even if one was replaced by a high contrast transparency no correlations could be observed.

8.6 Summary and conclusions

Having concluded in chapter 5 that the shared microlens correlator offered the optimum system configuration and having developed the novel routing elements required in chapter 7, in this chapter we began by demonstrating the performance expected of the shared microlens correlator by computer simulation. The patterns used are binary letters and have a contrast ratio of 10:1 which matches that of the FLC-SLM under LED illumination. The signal-to-largest sidelobe ratio (SLSR) of the correlation between two patterns of 'A's is 1.6:1. The high noise level is due to the leakage of light through the supposedly dark pixels if the inter-pixel gaps are assumed to be opaque. The SLSR can be partially improved by subtracting the correlation of the two original patterns from the correlation obtained when one of the patterns is replaced by its negative. Before subtraction, appropriate scaling is required to take account of the difference in correlation peaks caused by unequal number of white and black pixels in the image and the resultant SLSR is improved by about 1.6 times. By using orthogonal or randomised patterns, the correlation gives better discrimination since the cross-correlations of the former are zero or the sidelobes are suppressed if latter patterns are used. In this case, the limit of SLSR which is the contrast ratio of the patterns (10:1) can be achieved.
From the simulation, we concluded that the contrast ratio of the patterns is the most important parameter for the correlator to succeed. We, therefore, investigated the different factors which affect the contrast ratio of the main element for inserting the patterns in our system - the FLC SLM. The fundamental and most important reason for the low contrast is the presence of 5 μm transparent gaps in between the pixels which lets light pass through. This will reduce the extinction efficiency (the transmission through the dark state) and the best overall contrast obtained is limited to 16:1. Other secondary factors which degrade the contrast by various percentages include the angular response (1.5% loss in transmission for angles up to 30°), wavelength response (3% for about 50nm) and the effect of temperature on the switching angle (62% at 40°C) and the contrast ratio of the polarisers. The practical contrast value of 10:1 for LED illumination is a combination of all these factors.

We carried out some prototype experiments without the use of the FLC-SLM. Transparencies with contrast of 10:1 were used and the black area has to be masked by a black tape in order to give any correlation result. We conclude that the input and the filter must both have a contrast ratio of more than 10:1. Another correlation experiment involved using two resolution targets. These patterns contain features of many spatial frequencies at various spatially separated locations and can be considered as a special type of 'random' pattern i.e. they have high space bandwidth product. The number of white and black pixels are closer than for the characters simulated but are not equal. Correlation results obtained experimentally agreed well with the simulation for these very high contrast chrome-on-glass patterns. Finally, we attempted to build the system with a FLC-SLM and a high contrast transparency since using two SLMs would have too low contrast. We designed special illumination systems using a perpex tube with silver walls or an aluminium tube to provide sufficient luminance to the system. The back casing of the SLM has also been removed to ensure the correct geometry of the system. The lack of any correlation result is a clear indication that the success of the system is severely dependent upon very high contrast ratio at both the input and the filter planes. In order to raise the contrast to 1000:1, say, the gaps (if transparent) would need to be reduced to about 0.09 μm. This is not feasible since the narrowest gap spacing likely to
be about 1-2 \( \mu \text{m} \). So to raise the contrast ratio further, masks must be fabricated in the gaps or they must be switched to remain black.

We finally conclude that our feasibility study has been a success. Not only we identified the optimum system and have developed the optimum interconnect element, but we have also established the feasibility and have settled many key design issues. A full operational system was not possible within the time scale allowed and the financial resources. However, our successful system demonstration using high-contrast, high space bandwidth images show that the system does indeed work in practice and will do so programmably once a suitable high contrast SLM has been developed.
9.1 Conclusions

Let us remind ourselves of the original aim first before going on to see how we achieved it. The aim of the research described in this PhD thesis was to identify the design constraints, design rules and limitations for designing an optical total interconnection which could be used in a correlator or neural network system. The optimum choice of components and systems were made by taking account of both technical and commercial considerations. In some of the chapters the discussions were somewhat involved and to clarify the immediate results these chapters have conclusions at the end of them. Here we do not simply repeat these but instead draw out the main threads of the research:

1. We compiled detailed reviews of computer generated holograms and refractive microlenses and used these, together with more cursory examinations of other elements in chapter 1 to conclude that photoresist surface relief microlenses were the optimum choice of interconnection element for our system.

2. We compiled detailed reviews of optical interconnection systems and correlators and by re-expressing Sakano’s theory we concluded that the optimum system was a shared microlens system operating using incoherent light since it gave a shorter system length
Conclusions and future work

than the image multiplexer for the same number of interconnected pixels. We published a paper at an international refereed conference on this.

3. We invented two novel incoherent correlator system arrangements using both the image multiplexer and shared microlens arrangements and obtained a joint patent with our sponsor for the latter one.

4. The most similar previous system was one constructed by Sakano but he used an array of pinholes instead of microlenses. Other similar systems used shadow casting arrangements. Therefore, we carried out a theoretical analysis of all three types of system using scalar diffraction theories to establish their relative benefits and drawbacks. We analysed them for both on-axis and off-axis illumination and plotted the results as a set of graphs which effectively constitute design curves for the systems.

5. After our theoretical diffraction analysis we concluded that a shared system using microlenses gave almost twice the power efficiency of a pinhole array system while having the same resolution and cross-talk.

6. After our theoretical diffraction analysis we also concluded that for more compact systems the benefit of using either microlenses or pinholes became greater compared to using simple shadow casting in that the resolution was improved leading to less crosstalk and blurring at the output plane. This benefit offset the slight loss of efficiency incurred by using a shared microlens system rather than shadow casting.

7. We concluded that the ultimate system design should be a shared microlens system using long focal length microlenses with programmable images inserted into the system on ferroelectric liquid crystal binary amplitude spatial light modulators for use as a correlator. However, this required the development of novel long focal length microlenses and an illumination scheme amongst other requirements.

8. We developed and characterised novel long focal length microlenses by immersing surface relief photoresist microlenses in fluids of various refractive indices. We
demonstrated focal lengths from diverging to converging. We had to develop the optimum fabrication technique and transferred this knowledge and ability to our sponsor. We showed that these microlenses were ideal for our system and indeed were almost diffraction limited and so that aberrations were insignificant. In particular, astigmatic aberrations at off-axis angles of incidence were virtually eliminated within experimental error over the angular range measured. We published a paper at an international conference on this.

9. We proposed that microlens arrays could also be used as incoherent diffusers to generate the correct illumination for our correlator. We characterised the angular scattering profiles of a bulk diffuser and compared it with that of a microlens array. The diffuser gave a larger angular spread of 50 degrees but the response was stronger towards the axis whereas the microlens diffuser gave a uniform response over a small angular range of 25 degrees and was far more power efficient. We concluded that for our incoherent correlator we would use the bulk plastic diffuser as we needed sufficient angular range. However, we proposed that in the future it may be possible to use a microlens array as a more efficient diffuser and we outlined how such an array could be designed. We concluded that it was important in designing such a system to match the numerical aperture and angular response of all components such as the diffuser and the SLM in order to obtain the highest power efficiency.

10. We wrote a computer simulation to predict the correlation patterns and this enabled us to model the effect of the presence of an SLM having a 10:1 contrast ratio. The main reason for this poor contrast was the presence of gaps between the pixels of the SLM which were often transparent. The program showed that the output signal to noise ratio expected was very poor also being 2:1.

11. We devised ways to improve the contrast ratio of the output correlation and successfully demonstrated it in a computer simulation. The correlation between the two original patterns was subtracted from the correlation obtained when one of the patterns was inverted to give its negative. After appropriate scaling high contrast was achieved.
Even better results could be obtained with random or orthogonal patterns having a wide range of spatial frequencies resulting in high processing gain.

12. We carried out some prototype experiments to establish any practical difficulties that might occur. We concluded that the most serious problem was the lack of contrast of the two ferroelectric liquid crystal spatial light modulators and that this was so serious that even if one was replaced by a high contrast transparency no correlations could be observed.

13. Despite the lack of contrast of the SLMs we wanted to confirm that our system design did perform correlations as predicted by the simulations so we used two high contrast transparencies instead. We chose these to have a wide range of spatial frequencies to increase the processing gain. Ideally an array of long focal length microlenses should have been used but the photoresist microlenses have a higher absorption at the LED wavelength and some of the microlenses etched into glass were not available. Therefore, we set up the system in a shadow casting configuration which had the benefit of a higher efficiency although lower resolution. We successfully demonstrated translation invariant pattern recognition by obtaining the predicted correlation patterns at the output.

9.2 Future research

I would have liked to have had more time to actually demonstrate incoherent correlation using the ultimate system design with two SLMs in a shared long focal length microlens total interconnect configuration. If the project were to continue the following should be done:

1) The SLM should be replaced by one with a higher contrast. We understand that companies are now developing such SLMs with masks to block out the transmitted light. Alternatively an electrically addressed SLM could be used to cast an image onto an optically addressed ferroelectric SLM which could be switched to store the image. Then the optically addressed SLM could be read and ought to have higher contrast due to the
absence of gaps. It is necessary to use a ferroelectric one to achieve the large angular range.

2) An array of highly luminous surfacing-emitting LEDs should be purchased and used for illumination to improve power input and uniformity.

3) A CCD device dismounted from its camera housing should be used to implement the correct geometry of the system.

4) The diffuser should be followed by a lens or blazed Fresnel zone plate to rotate the edge polar scattering distributions around to point more nearly towards the centre to improve uniformity of interconnect strengths and power efficiency.

5) The long focal length microlenses should ideally be immersed in a solid matrix rather than a liquid to make a practical system and some investigation of casting and different epoxy resins should be made.

During the course of this PhD numerous ideas were discussed with the various supervisors and below we list the more promising ones.

1) We described the design of a novel microlens diffuser using different sizes of microlens with the same focal length in the array.

2) During the course of the PhD the novel Fresnel-Dammann grating was invented in our department which operates over a range of wavelengths without change of either focal length or spacing in the array of focal spots produced [God93]. This would appear to make an ideal element for the total interconnect and this should be investigated.
REFERENCES


**Bir91a** K.D. Bird, D. Daly & T.J. Hall, "Computer generated microlens arrays and their application to optical free space switching networks", *Proc. of IEE Int. Conf. on Holography*, 57-61 (1991)


**Bor91** N.F. Borrelli, "The generation of lens arrays using photothermal techniques", *IOP Short Meetings on Microlens arrays*, Series No. 30, 1-16 (1991)


Buy93  L. Buyden, IMEC, University of Gent, private communication (1993)


CRL   CRL 128x128 Ferroelectric liquid crystal spatial light modulator 2DX128 Instruction manual

Dal93a  D. Daly, private communication (1993)

Dal93b  D. Daly and M.C. Hutley, "Microlens measurement at NPL", *EOS Topical Meeting on Microlens arrays*, Digest Series 2, 50-54 (1993)

Dam70  H. Dammann, "Blazed synthetic phase-only holograms", *Optik*, 31, 95-104 (1970)


Eal94  Ealing catalogue (1994)


References


References


Gla87  I. Glaser, "Information processing with spatially incoherent light", Progress in Optics, ed. by E. Wolf, 24, 391-509 (1987)


Goo68  J.W. Goodman, "Introduction to Fourier optics", pub. by McGall-Hill (1968)


References


Hec87 E. Hecht, "Optics", 2nd Ed., pub. by Addison Wesley (1987)


Hut90a M.C. Hutley, R.F. Stevens and D. Daly, "The manufacture of microlens arrays and fan-out gratings in photoresist", Technical Digest of IEE Colloq.
on Optical connection and switching networks for communication and computing, 11/1-11/3 (1990)

Hut90b M.C. Hutley, "The manufacture and testing of microlens arrays", SPIE Proc., 1319, 491-492 (1990)


Hut91a M.C. Hutley, D. Daly and R.F. Stevens, "The testing of microlens arrays", IOP Short Meetings on Microlens arrays, Series No. 30, 67-81 (1991)

Hut91b M.C. Hutley, R.F. Stevens and D. Daly, "Microlens arrays", Phy. World, 27-32 (July 1991)


Ima91 H. Imam, A.G. Kirk, K.D. Bird and T.J. Hall, "Design of multiphase level holograms", IEE Int. Conf. on Holography, 166-170 (1991)


References 226


Kim90  M.S. Kim and C.C. Guest, "Simulated annealing algorithm for binary
phase only filters in pattern classification", Appl. Opt., 29(8), 1203-1207
(1990)

Kir91  A.G. Kirk, H. Imam and T.J. Hall, "An efficient holographic interconnect in
0.01 cm^3", Proc. of 3rd Int. Conf. on Holographic systems, components and
applications, 161-165 (1991)

Kno49  W.E. Knowles Middleton and F.D. Smith, "The diffusion of light by
ground glass, with special reference to color changes", Can. J. Res., 27F,
151-166 (1949)

Kno60  W.E. Knowles Middleton, "Diffusion of ultraviolet and visible light by

Koi82  Y. Koike, Y. Kimoto and Y. Ohtsuka, "Studies on the light-focussing
plastic rod 12: The GRIN fibre lens of methyl methacrylate-vinyl

Kob79  N. Kobayashi, T. Ishikawa, K. Minemura and S. Sugimoto, Fibre
Integrated Opt., 2, 1 (1979)

Kra90  U. Krackhardt, M. Heissmeier, S. Sinzinger and N. Streibl, "New algorithm
for the design of binary phase gratings", SPIE Proc., 1319, 342 (1990)

Kra92  U. Krackhardt, F. Sauer, W. Stork and N. Streibl, "Concept for an optical

Kre93  B.C. Kress and S.H. Lee, "Iterative design of computer generated Fresnel
holograms for free-space optical interconnections", Tech. Digest of OSA
Topical Meeting on Optical Computing, 7, 22-25 (1993)

Kri92  A.V. Krishnamoorthy, G. Yayla and S.C. Esener,"A scalable optoelectronic
Net., 3(3), 404-413 (1992)

Kuf93  M. Kufner and S. Kufner, "Microlenses in PMMA by deep proton
irradiation - capacity of the process", EOS Topical Meeting on Microlens
arrays, Digest Series 2, 5-8 (1993)

Kur92  T. Kurokawa and S. Fukushima, “Spatial light modulators using
Lee90a  S.K. Lee, "Computer generated holograms for optical interconnection of wafer scale integrated circuits", Dept. of Elec. Eng., UCL, MSc project report (1990)


Leg91  J.S. Leggatt and M.C. Hutley, "Microlens arrays for interconnection of single mode fibre arrays", Elec. Lett. 27(3), 238-240 (1991)


References


Lon73 R.S. Longhurst, "Geometrical and physical optics", pub. by Longman, 3rd Edition (1973)


References


Newport  Newport's guide to GRIN-rod lenses

Nippon  Nippon Sheet Glass technical information on Selfoc lens arrays


Nis91  K. Nishizawa and M. Oikawa, "Planar microlens by ion-exchange", IOP Short Meetings on Microlens arrays, Series No. 30, 17-21 (1991)


NSG  NSG Planar microlens array (PML) preliminary specifications


Pan92 P. Pantelis and D.J. McCartney, "Polymer microlens arrays", 3rd Int. Conf. on the Electrical, Optical and Acoustic Properties of Polymers (EOA III), (Sep 1992)


Rog77  G.L. Rogers, "Noncoherent optical processing", pub. by Wiley (1977)


<table>
<thead>
<tr>
<th>Year</th>
<th>Reference</th>
</tr>
</thead>
</table>
References 235


Stu  R.J. Stuart, "An investigation into the use of zone plates as optical interconnects", Dept. of Elec. Eng., UCL, Third year project final report


References


References

Wada81 O. Wada, S. Yamakoshi, M. Abe, Y. Nishitani and T. Sakurai, "High radiance InGaP/InP lensed LED's for optical communication systems at 1.2-1.3 mm", IEEE J. Quan. Elec., QE-17(2), 174-178 (1981)


Wong93b H. Wong, private communication, 1993


References


Appendix I - Re-expression of Sakano's theory

(i) For the image multiplexer and can be derived from Sakano's theory:

(a) Astigmatic limit:

\[ \beta^2 \leq \frac{\alpha}{4I(1+\alpha)^3} \]  

where \( \beta = \frac{Nd}{L} \)

NxN - array size of microlenses

d - lens diameter

L - total system length

\( \alpha \) - ratio of distances between microlens and output planes

and that between input and microlens array

\( I = \frac{(3n+1)}{4n} \)

n - refractive index of the lens

In the case of the image multiplexer, the closest one get is when the ratio of the separation \( \alpha \) is \( Nf:f \), or \( \alpha = 1:N \). Substituting for both \( \alpha \) and \( \beta \):

\[ L \geq Nd \sqrt{4IN \left(1 + \frac{1}{N}\right)^3} \]  

For a large array, ie. \( N \gg 1 \), it can be simplified as:

\[ L \geq Nd \sqrt{4NI} \geq 2N^3dI \]  

(b) Diffraction spot size limit:

\[ \beta \geq \frac{2.44\alpha\lambda N^3}{L(1+\alpha)^2} \]  

Substituting for \( \alpha \) and \( \beta \) gives:
For large $N$, the inequality becomes:

(ii) For a shared microlens system:

\[ L < \frac{d^2}{2.44\lambda} \]

(a) The astigmatic limit is now given by:

\[ \beta^2 \leq \frac{4\alpha^2}{L(1+\alpha)^4} \]

This time the ratio $\alpha$ is 1:1 for a 2f:2f imaging system and therefore:

\[ \frac{N^2 d^2}{L^2} \leq \frac{4\alpha^2}{L(1+\alpha)^4} \]

\[ L^2 \geq 4N^2d^2I \]

\[ L \geq 2NdI^{\frac{1}{2}} \]

(b) The diffraction spot size limit is:

\[ \beta^2 \geq \frac{9.76\alpha N^2}{L(1+\alpha)^2} \]

which can be simplified to:

\[ L \leq \frac{(1+\alpha)^2 d^2}{9.76\alpha\lambda} \leq \frac{d^2}{2.44\lambda} \]

for $\alpha = 1$

Notice that the diffraction spot size limit in both cases are the same.
Appendix II - Aspects of diffraction theory
(not explained in Chapter 5)

(1) Diffraction limited spot size of shared microlens system

From eqn (4.27), the complex amplitude at the output plane \((x_3, y_3)\) for the shared microlens system is given by:

\[
U(x_3, y_3) = \frac{A e^{i\phi}}{j\lambda z_2} \iiint \exp \left[ \frac{-jk}{z_2} \left( x_2 (Mx_1 + x_3) + y_2 (M y_1 + y_3) \right) \right] dS
\]  

(II.1)

For a circular lens, if we define the plane \((x_2, y_2)\) in terms of circular coordinates, i.e.

\[
x_2 = r \cos \phi, \quad y_2 = r \sin \phi;
\]

\[
M x_1 + x_3 = q \cos \phi, \quad M y_1 + y_3 = q \sin \phi
\]

and \(dS = r dr d\phi\)

Equation (I.1) becomes

\[
U(x_3, y_3) = \frac{A e^{i\phi}}{j\lambda z_2} \int_0^{2\pi} \int_0^r \exp \left[ \frac{jk r q \cos(\phi - \theta)}{z_2} \right] r dr d\phi
\]  

(II.3)

Because of the complete axial symmetry, the solution must be independent of \(\phi\), so let us set \(\phi\) to be zero for convenience.

\[
U(x_3, y_3) = \frac{A e^{i\phi}}{j\lambda z_2} \int_0^{2\pi} \int_0^r \exp \left[ \frac{jk r q \cos \phi}{z_2} \right] d\phi dr
\]

\[
= \frac{A e^{i\phi}}{j\lambda z_2} \int_0^r \int_0^{2\pi} \left( \frac{krq}{z_2} \right) dr
\]

(II.4)

\[
= \frac{A e^{i\phi}}{j\lambda z_2} \left( \frac{2\pi a^2}{\lambda z_2} \right) \left[ J_1 \left( \frac{kaq}{z_2} \right) \right]
\]

This gives rise to the Airy disc profile.
Appendix II

(2) Efficiency

The energy reaches the output plane is the total energy incident on the aperture, ie.

\[ \iiint |U(x_1, y_1)|^2 dx_1 dy_1 = E \]  

(II.5)

but the complex amplitude at the output is given by:

\[ U(x_3, y_3) = \iiint P(x_2, y_2) \exp \left( -\frac{jk}{z_2} (x_2 x_3 + y_2 y_3) \right) dx_2 dy_2 \]  

(II.6)

This second equation represents the Fourier transform of the function \( P(x_2, y_2) \), in which the object space is \( (x_2, y_2) \) and the Fourier space is \( (x_3/\lambda z_2, y_3/\lambda z_2) \), ie.

\[ x_2 \leftrightarrow \frac{x_3}{\lambda z_2} \]  

(II.7)

By Parseval’s theorem,

\[ \iiint |P(x_2, y_2)|^2 dx_2 dy_2 = \iiint |U(x_3, y_3)|^2 \left( \frac{dx_3}{\lambda z_2} \right) \left( \frac{dy_3}{\lambda z_2} \right) \]  

(II.8)

\[ = \frac{1}{(\lambda z_2)^2} \iiint |U(x_3, y_3)|^2 dx_3 dy_3 \]

For the shared microlens system, \( P(x_2, y_2) \) is the pupil function such that:

\[ P(x_2, y_2) = C \quad \text{inside the aperture } S \]

\[ = 0 \quad \text{outside} \]  

(II.9)

From eqn (II.8),

\[ |C|^2 \cdot (\pi a^2) = \frac{E}{(\lambda z_2)^2} \]  

(II.10)

\[ |C|^2 = \frac{E}{\pi (a \lambda z_2)^2} \]

The on-axis intensity is given by:

\[ I_o = |U(0,0)|^2 = |C|^2 (\pi a^2)^2 \]  

\[ = \frac{\pi a^2 E}{(\lambda z_2)^2} \]  

(II.11)
Appendix II

The fraction of the total energy contained within a circle of radius $q_0$:

$$\eta(q_0) = \frac{1}{E} \int_0^{2\pi} \int_0^{q_0} I(q) q dq d\phi$$  \hspace{1cm} (II.12)

where $I(q) = I_o \left[ \frac{2J_0(kaq/z_2)}{kaq/z_2} \right]^2$ \hspace{1cm} (II.13)

Equation (II.12) becomes:

$$\eta(q_0) = \frac{8\pi \cdot \pi a^2 E}{(\lambda z_2)^2} \left( \frac{z_2}{ka} \right)^2 \int_0^{kaq/z_2} J_0^2(kaq/z_2) dq$$

$$= 2 \int_0^{kaq/z_2} J_0^2(w) dw$$

$$= -\left[ 1 - J_0^2(w) - J_1^2(w) \right]_{0}^{kaq/z_2}$$

$$= 1 - J_0^2\left( \frac{kaq}{z_2} \right) - J_1^2\left( \frac{kaq}{z_2} \right)$$

whereby the change of variable $w = kaq/z_2$

At the nulls, $J_1 = 0$,

$$\eta(q_0) = 1 - J_0^2\left( \frac{kaq}{z_2} \right)$$  \hspace{1cm} (II.15)

The efficiency of the shared microlens system when $q_0 = 1.22 \lambda z_2/(2a)$ is found to be about 83.6%.

The efficiency for the pinhole system can be found in a similar way although no analytical solution can be determined.
(3) Phase function for collimated rays passing through the lens at an angle

Suppose that only one dimension is considered here for simplicity, in the x-direction. The equation of the circle for the lens curvature is given by:

\[ x^2 + z^2 = R^2 = X^2 + Z^2 \]

or

\[ Z = \sqrt{R^2 - X^2} \] (II.16)

The equation of the line \( l \) is given by:

\[ X = -\frac{Z}{\tan \theta} + c \]

or

\[ Z = -(X + c) \tan \theta \] (II.17)

The thickness of the lens tilted at an angle of \( \theta \) is thus:

\[ t(X) = Z_{\text{circle}} - Z_{\text{line}} \]

\[ = \sqrt{R^2 - X^2} + (X + c) \tan \theta \]

\[ = X - \frac{X^2}{2R} + (X + c) \tan \theta \] (II.18)

\[ = -\frac{X^2}{2R} + X(R + \tan \theta) + c \tan \theta \]

using the thin lens approximation (or when \( X \ll R \)). The resulting amplitude function of the lens is:

\[ e^{i\phi(X)} = \exp \left[ jk \left( -\frac{X^2}{2R} + X(R + \tan \theta) + c \tan \theta \right) \right] \] (II.19)
The first term gives the familiar quadratic phase change through a lens and $R$ is related to the focal length $f$. The second term merely shifts the output by a certain amount that depends on the magnification of the system, i.e. space-invariant. The third term gives rise to a constant which will be eliminated when taking modulus.
Appendix III - Seidel coefficients for Astigmatism and Spherical Aberrations

(1) Definition of the Seidel coefficients

Seidel coefficients determine the effect of optical elements on the propagation of light differs from the ideal and the phase imparted by the element is given in the form:

\[ W(r, \theta) = W_{\text{ideal}}(r, \theta) + \Delta w(r, \theta) \]  

(III.1)

where \( r, \theta \) are the cylindrical coordinates in the plane of the optical element (e.g. a lens) and \( \Delta w \) is the amount of phase imparted which is different from the ideal wave. The Seidel coefficients are the coefficients of an expansion of \( \Delta w \) in terms of cylindrical harmonics and powers of \( r \).

(2) Relative phase in the plane of the lens relative to the centre

A travelling light wave in 1-D can be described by \( A e^{ikx} \) where \( A \) is the amplitude and \( k \) is the propagation number and \( x \) is the distance travelled. The phase of the light at the focus relative to the centre of the lens is \( 2\pi f/\lambda \). The phase of the light relative to the position \( r \) in the lens plane is given by:

\[
\frac{2\pi z}{\lambda} = \frac{2\pi \sqrt{f^2 + r^2}}{\lambda} = \frac{2\pi f}{\lambda} \left( 1 + \frac{r^2}{2f^2} \right) \\
= \frac{2\pi f}{\lambda} \left( 1 + \frac{r^2}{2f^2} \right) \\
\approx \frac{2\pi f}{\lambda} + \frac{\pi r^2}{\lambda f}
\]

(III.2)

The relative phase in the plane of the lens relative to the centre is therefore \( \frac{\pi r^2}{\lambda f} \).
(3) Seidel coefficient for astigmatism

At off-axis, astigmatism causes light to be focussed into two different points in the sagittal and meridional planes. The focal lengths are given by \( f \) and \( f \cos^2 \phi \) where \( \phi \) is the off-axis angle. The effective aberration which results in the change in focal length is therefore the difference between the ideal phase and that from the altered focal length:

\[
\Delta w = \frac{\pi r^2}{\lambda} \left( \frac{1}{f \cos^2 \phi} - \frac{1}{f} \right)
\]

\[
= \frac{\pi r^2 (1 - \cos^2 \phi)}{\lambda f \cos^2 \phi}
\]

\[
= \frac{\pi r^2 \phi^2}{\lambda f} \quad \text{if assuming } \cos^2 \phi = \left(1 - \frac{\phi^2}{2}\right)^2 = 1 - \phi^2
\]

The symmetry of the astigmatism leads to the following expression for phase aberration:

\[
\Delta w(r, \theta) = \frac{\pi r^2 \phi^2 \cos \theta}{2\lambda f_{av}}
\]

(III.4)

where \( f_{av} \) is the average focal length. The Seidel coefficient of this term is given by:

\[
\Delta w(r, \theta) = A \rho^2 \cos \theta
\]

(III.5)

where \( \rho \) is the normalised coordinate given by \( 2r/d \) and \( d \) is the diameter of the lens.

This lead to the Seidel coefficient for the astigmatism being:

\[
A = \frac{\pi d^2 \phi^2}{8\lambda f_{av}}
\]

(III.6)
(4) Seidel coefficient for spherical aberrations

From [Hec87]:

$$\frac{n_1 + n_2}{S_o} - \frac{n_2 - n_1}{S_i} = \frac{n_2 - n_1}{R} + h^2 \left[ \frac{n_1}{2S_o} \left( \frac{1}{S_o} + \frac{1}{R} \right)^2 + \frac{n_2}{2S_i} \left( \frac{1}{R} - \frac{1}{S_i} \right)^2 \right]$$  \hspace{1cm} (IV.7)

where $n_1$ and $n_2$ are the refractive indices of the air and glass respectively, $s_i$ and $s_o$ are the conjugate image positions of a single refracting surface, $R$ is the radius of curvature of the lens and $h$ is the sagittal height of the lens.

Define the focal length as the position of the image, $s_i$, when the object is at infinity (i.e. $s_o = \infty$) and the height of the lens, $h \ll 1$, therefore (IV.7) becomes the usual relationship between $f$, $R$ and the indices for a thin lens:

$$\frac{n_2}{f} = \frac{n_2 - n_1}{R} \hspace{1cm} (IV.8)$$

As the size $h$ of the lens increases, the second term on the RHS of eqn (IV.7) can no longer be ignored. The focal length becomes aberrated and is denoted by $f_{SA}$. This is the result of spherical aberrations. The new focus is described by:

$$\frac{n_2}{f_{SA}} = \frac{n_2 - n_1}{R} + \frac{n_2 h^2}{2f^3} \left( \frac{1}{R} - \frac{1}{f} \right)^2$$  \hspace{1cm} (IV.9)

The difference between this and the ideal focal length is the second term and from observation, it can be seen that:

$$A \propto \frac{d^2}{f^3}$$  \hspace{1cm} (IV.10)
Correlation simulation program - Matlab program

function [z] = correl3(n,pat1,pat2)

% This function calculates the correlation of two 32 x 32 patterns
% pat1 and pat2 by the defining integral (or summation)
% n is the figure number

z=zeros(64);
for xi=1:32;
    for eta=1:32;
        z(xi,eta)=sum(sum(pat1(l:xi,l:eta).*pat2((33-xi):32,(33-eta):32)));
    end
end
for xi=1:32;
    for eta=1:32;
        z(31+xi,31+eta)=sum(sum(pat1(xi:32,eta:32).*pat2(1:(33-xi),1:(33-eta))));
    end
end
for xi=1:32;
    for eta=1:32;
        z(32-xi,32+eta)=sum(sum(pat1(1:(32-xi),(eta+1):32).*pat2((xi+1):32,1:(32-eta))));
    end
end
for xi=1:32;
    for eta=1:32;
        z(32+xi,32-eta)=sum(sum(pat1((xi-1):32,l:(32-eta)).*pat2(l:(32-xi),(eta+1):32)));
    end
end
Appendix V - SLM control programs

(1) SLM display program

/* DISPLAY3.C (v. 3.0) */
/* This is a program for displaying (HEX & ASCII) patterns on FLC-SLM */
/* by Peter C.H. Poon and Mike G. Robinson */
/* */
/* 3rd August 1994 */

#include <stdio.h>
#include "simcmd.h"

unsigned char outbyte[128][16];

main()
{
  int i,
   j,
   choice;
  void Menu(void),
    DisplayAscii(void),
    SlmInit(void),
    Init(void),
    DisplayFrame(void);
  SlmInit();
  while(1) /* Initialise data array */
  {
    for(i=0;i<128;i++)
      for(j=0;j<16;j++)
        outbyte[i][j] = 0x00;
    Menu();
    scanf("%d",&choice);
    switch(choice)
      {
    case 1:
      DisplayHex();
      break;
    case 2:
      DisplayAscii();
      break;
    case 3:
      Init();
      break;
    case 4:
      DisplayFrame();
      break;
  } /* End of while loop */
} /* End of main */

/* End of DISPLAY3.C */
case 5:
    Command(InvertContrast);
    break;

case 6:
    exit(0);
}

void DisplayHex(void) /* Display pattern in hex */
{
    FILE *fp,
        *FileInput(void);
    int i,
        j,
        n,
        FrameInput(void);
    unsigned char inbyte;
    void SendPict(int);

    fp = FileInput();
    n = FrameInput();

    for(i=0;i<128;i++)
        for(j=0;j<16;j++)
        {
            fscanf(fp,"%x", &inbyte);
            outbyte[127-i][j] = inbyte; /* Make pattern upside down */
        }
    fclose(fp);

    SendPict(n);
}

void DisplayAscii(void) /* Display pattern in 32x32 ASCII text */
{
    FILE *fp,
        *FileInput(void);
    int i,
        j,
        k,
        n,
        FrameInput(void);
    unsigned char first,
                second,
                obyte,
                ReadChar(FILE *);
    void SendPict(int);

    fp = FileInput();
    n = FrameInput();

    for(i=0;i<128;i+=4)
    {
for(j=0;j<16;j++)
{
    first = ReadChar(fp);
    second = ReadChar(fp);
    if(first == '0' && second == '1')
        obyte = 0x0F;
    else if(first == '!')
        obyte = 0xF0;
    else if(first == '1' && second == '0')
        obyte = 0xFF;
    else if(first == '0' && second == '0')
        obyte = 0x00;
    for (k=0;k<4;k++)
        outbyte[i+k][j]=obyte;
}
fclose(fp);

SendPict(n);
}

*******************************************************************************

void Init() /* Initialise a particular frame */
{
    int n,
    FrameInput(void);
    void SlimInit2(int);

    n = FrameInput();
    SlimInit2(n);
    Command(SetDisplayFrame,n);
}

*******************************************************************************

void DisplayFrame(void) /* Display a particular frame */
{
    int n,
    FrameInput(void);

    n = FrameInput();
    Command(SetDisplayFrame,n);
}

*******************************************************************************

FILE *FileInput(void) /* Input filename */
{
    char fn[32];
    FILE *fp;

    while(1)
    {
        printf("Enter name of data file to be stored on SLM? ");
        scanf("%s",fn);
        if((fp = fopen(fn,"r")) != NULL)
break;
printf("Cannot open %s\n", fn);
}
return(fp);
}

void SendPict(int n) /* Pre-interleave data and send to SLM */
{
int i,
j;
Command(SetWriteFrame, n);
for(i=0; i<64; i++)
  for(j=0; j<16; j++)
  {
    SendByte(outbyte[2*i][j]); /* Send the even rows */
    if(outbyte[2*i][j] == OxlB) /* 1B is a reserved byte, */
      SendByte(0x1B); /* must send twice */
  }
for(i=0; i<64; i++)
  for(j=0; j<16; j++)
  {
    SendByte(outbyte[2*i+1][j]); /* Send the odd rows */
    if(outbyte[2*i+1][j] == OxlB)
      SendByte(0x1B);
  }
Command(SetDisplayFrame, n);
}

unsigned char ReadChar(FILE * fp)/* Read the '1' & '0' only */
{
unsigned char inbyte;
while(1)
{
  fscanf(fp, "%c", &inbyte);
  if
}
{

:(ooxo)3)^ap"9S
(++r-9i>r‘o=DJoj
(++î‘831>?‘0=!)Joj

!(U‘3UIBJd3JU^J3S)pUBUIUI03
î(0‘3UTJ3SnBtP3S)pUBUIU103
:(0‘3SJÜ033SnBJJ3S)pUBUIUI03
! ( ç ‘3u re Jj X B j d s i( p 3 S )p u B u iu i 0 3
Î(a q o jJ S F U U 0 > jj 3 S )p iiB u iu i 0 3

t(upuo3î3S)puüuiuio3
:(UB3S0A\2)3S)P"T31UIU0^
;(3‘3uijb3Jji9S)puBumio3
î(0‘9SJBO3b3Jjj3S)puBuimo3
!(j3JUUd~9ZI|BpiUl)pUBUmi03
!f
‘I }U1
'}

/* 93IA9P UI|S 9qi UI 9UIB^ B9SI|BIJIUI */

(U)Ul)^)IUIUqg pIOA

/* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * /

{

:90ioq3„)jjuud
-(..U\U\U\U\U\U\1:X3 (9)l\.,)pupd
•(..^ A u \W lS
uo iu9WBd 9qj y9AUi (ç)j\..)jjuud
•(..«\«\W1S
UI P9J0JS 9UIBJJ B XBldsiQ (t7)j\JjJUUd
Î(..U\U\W1S SSIIBIJIUI (e)j\..)jjuud
‘.(..U\U\S91IJ 1X9J nDSV z m £ UIOJJ
uo XBidsiQ (3)}\..)jjuud
!(„u\u\S9Iij X9q UIOIJ I ^ is uo XBjdsia (l)j\..)puijd
;(„U\U\U\U\U\U\U\U\U\lI\U\U\U\U\U\lI\U\U\.,)jJUUd
'

/* nu9pq XBjdsiQ */

}

(piOA)nu9j/^ piOA

/* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * /

{
î(9 jX q u i)u jm 9 J

{

t5fB9jq
(.1. =

fS Z

A x ip u d d d y

II .0. =

3)Xqui);i


(2) SLM command program

```c
#include <stdio.h>
#include <bios.h>
#include "simcmd.h"

extern unsigned slm = 0; /* 0 parallel port */

void SendByte(unsigned char byte) /* Send data to SLM */
{
    unsigned int status;
    status = _bios_printer(_PRINTER_WRITE, slm, byte);
    if(status & 0x0001 >0)
        printf("Error in sending byte to slm device \n");
}

unsigned short Command(int command, int optParm) /* Send command to SLM */
{
    unsigned short status;
    unsigned short data;
    switch(command){
    case Initialize_printer:
        status = _bios_printer(_PRINTER_INIT, slm, data);
        printf("Parallel port initialized, status = 0x\n",status);
        break;
    case SendEsc:
        SendByte(Ox1B);
        break;
    case SendNull:
        SendByte(Ox00);
        break;
    case SetLeadingPulse:
        Command(SendEsc);
        SendByte(Ox01);
        Command(SendNull);
        break;
    case SetTrailingPulse:
        Command(SendEsc);
        SendByte(Ox02);
        Command(SendNull);
        break;
    case SetFrameBlank:
        Command(SendEsc);
        SendByte(Ox03);
        Command(SendNull);
    }
break;
case SetTwoScan:
    Command(SendEsc);
    SendByte(0x04);
    Command(SendNull);
    break;

case SetACDataZeroStrobe:
    Command(SendEsc);
    SendByte(0x05);
    Command(SendNull);
    break;

case SetZeroDataStrobe:
    Command(SendEsc);
    SendByte(0x06);
    Command(SendNull);
    break;

case SetNormalStrobe:
    Command(SendEsc);
    SendByte(0x07);
    Command(SendNull);
    break;

case SetSingleScan:
    Command(SendEsc);
    SendByte(0x08);
    Command(SendNull);
    break;

case SetContin:
    Command(SendEsc);
    SendByte(0x09);
    Command(SendNull);
    break;

case SetSingleScanInit:
    Command(SendEsc);
    SendByte(0x0A);
    Command(SendNull);
    break;

case InvertContrast:
    Command(SendEsc);
    SendByte(0x0B);
    Command(SendNull);
    break;

case SetFreqCoarse:
    Command(SendEsc);
    SendByte(0x81);
    SendByte(optParm);
    break;

case SetFreqFine:
    Command(SendEsc);
    SendByte(0x82);
    SendByte(optParm);
    break;

case SetWriteFrame:
    Command(SendEsc);
    SendByte(0x83);
    SendByte(optParm);
    break;

case SetDisplayFrame:
    Command(SendEsc);
SendByte(0x84);
SendByte(optParm);
break;
case SetPauseCoarse:
    Command(SendEsc);
    SendByte(0x85);
    SendByte(optParm);
    break;
case SetPauseFine:
    Command(SendEsc);
    SendByte(0x86);
    SendByte(optParm);
    break;
case SendDataByte27:
    Command(SendEsc);
    break;
default: ;
}

void SImInit(void) /* Initialise the slm device */
{
    int i,j;

    Command(Initialize_printer);
    Command(SetFreqCoarse,0);
    Command(SetFreqFine,2);
    Command(SetTwoScan);
    Command(SetContin);
    Command(SetNormalStrobe);
    Command(SetDisplayFrame,5);
    Command(SetPauseCoarse,0);
    Command(SetPauseFine,0);
    Command(SetWriteFrame,0);
    for(i=0;i<128;i++)
    for(j=0;j<16;j++)
        SendByte(0x00);
    Command(SetDisplayFrame,0);
}
(3) SLM command header file

```c
enum{
    Initialize_printer,
    SendEsc,
    SendNull,
    SetLeadingPulse,
    SetTrailingPulse,
    SetFrameBlank,
    SetTwoScan,
    SetACDataZeroStrobe,
    SetZeroDataStrobe,
    SetNormalStrobe,
    SetSingleScan,
    SetContin,
    SetSingleScanInit,
    InvertContrast,
    SetFreqCoarse,
    SetFreqFine,
    SetWriteFrame,
    SetDisplayFrame,
    SetPauseCoarse,
    SetPauseFine,
    SendDataByte27
};

void SendByte(unsigned char);
unsigned short Command(int, int);
void SlmInit(void);
```
LIST OF PUBLICATIONS AND PATENT


