Optical Phase-Only Correlation and Image Processing

Ian W. Chisholm

Ph.D.

University of Edinburgh

1996

Acknowledgements

I would like to thank the Science and Engineering Research Council for funding the first two years of this research, and the University of Edinburgh Physics Department (most specifically Dr. D.G. Vass) for arranging that I be awarded a Ritchie research scholarship to fund my third year.

Many thanks go to Andrew Garrie, the group technician without whose hard work the experimental sections would have been impossible, and to Dr Will Hossack for his many useful ideas and help - not least his suffrance during the proofreading of this thesis! I would like also to thank my supervisors, Dr D.G. Vass and Dr N.E. Fancey for their advice and help during the writing of this tome.

Irene of the Computing Service Bio-Support team deserves particular thanks for typing in my bibliography having had to decode my (awful) handwriting with the aid of Susan.

Last, and definitely not least, I would like to thank Duncan and Steven for tea at three, crosswords, pints and other traditions. It wouldn't have been nearly so much fun without them.

To Helen, my wife, for her patience and support, and babies Alex and Ewen, born during the course of this thesis, for showing me that physics isn't that important anyway.

It isn't every thesis that encompasses two lifetimes!

Abstract

The potential of the instantaneous Fourier transforming ability of optical systems is now being tapped. In particular, optical correlation is more and more becoming the focus of attention by current research groups with it's many applications in the fields of machine vision, pattern identification and target tracking. This thesis gives a comprehensive view of the current state of the science of optical correlation and describes the many factors affecting experimental results in typical correlation systems including algorithmical errors and limitations and optical effects such as noise or phase inconsistency.

The uses and types of spatial light modulator used in such systems are described and, in particular, the reflective electronically addressed liquid crystal over silicon modulator used in this thesis. Liquid crystal mesophases are examined and their light modulating effects analysed with respect to their applications in optical systems.

Current optimisation algorithms and techniques are analysed and an adaptive stochastically self-correcting correlator system is described. Results are presented as computer simulations and sample optical processing architectures are described. These illustrate the benefits of such a technique on physical systems with the capability of cancelling out the effects of imperfections in objects and filters and indeed in the optics themselves.

I've touch'd the highest point of all my greatness;

And from that full meridian of my glory

I haste now to my setting. I shall fall,

Like a bright exhalation in the evening

And no man see me more.

Cardinal Wolsey (From Shakespeare's King Henry VIII)

Contents

Introduction

The structure of this thesis is designed to take the subject of Spatial Light Modulator-based optical processing techniques from its inception as a primitive form through to its most recent technological achievements and methods. Examples are given where appropriate and experimental results are outlined to back up the theoretical standpoints. Computer simulations are also extensively used to illustrate the simpler points as well as the more complex processor-intensive methods of the latter chapters.

In Chapter 1 the basic tenets of Fourier transforming through optics are outlined, with proofs in appendices. The initial theoretical statements of the transforming power of a simple lens are extended to allow derivations of correlation and convolution equations and it is shown how these may be achieved using optical methods. Other applications of Fourier image processing are given and the versatility and speed of the procedures is emphasized.

Chapter 2 introduces the Spatial Light Modulator (hereafter referred to as the SLM) in its many and varied guises. Electrically addressed and optically addressed SLMs are described with their advantages and disadvantages as practical devices. Also, within this section, the physical mothods by which light is to be modulated are stated and the three main liquid crystal types are described with their physical

characteristics, and corresponding ability to affect an incident beam of light. Each has its own strengths and weaknesses and these are described briefly (with references) as they have been well studied elsewhere.

A description of the practical work carried out on the electrically addressed SLM designed by Douglas McKnight in this department is laid out with subsequent conclusions and the reasons for not continuing with the study of this device. The methods used in evaluating potential SLMs optically and the electrical and computer interfaces are described with diagrams. Some of the efforts made towards improving the manufacturing methods and overall quality of the devices used in the group are described.

The theoretical sections are concluded with Chapter 4 which enters the realm of optical correlation methods and techniques. The subject of (and problems associated with) using pixellated devices as Fourier plane filters is addressed and some solutions indicated and referenced with respect to sampling and the space-bandwidth product of the SLM used. The most important algorithms for generating optical correlations are described in detail with a more general and flexible iterative technique being the main subject for this thesis. A short introduction to simulated annealling in Fourier optics serves as a precursor to chapter 5.

Chapter 5 consists of a much more detailed discussion on simulated annealling together with a description of the computer simulations used to illustrate its use and general flexibility in optical correlation systems. Comparisons are drawn with other algorithms and results presented to substantiate these. It is shown that such iterative and adaptive techniques allow for infinitely more flexibility than previously used algorithms and that even major phase distortions in the filter SLM

may be (almost) anulled by the use of this technique. It is one of the aims of this thesis to illustrate that phase-correction is inherent in the simulated annealing scenario: regardless of whether the SLM is imperfect or not - the filter 'adapts' to the SLM and the optical system in general.

This work concludes with a brief chapter summing up the work outlined and presenting a case for further study in the area. Optical image processing is still a science in relative infancy and the potential for development is vast.

Chapter 1

Fourier Optics and Optical Image Processing

Introduction

Coherent image processing is entirely founded on the Fourier transforming ability of lenses (e.g. Goodman (1968), Gaskill (1978) and summarised in Appendix 1): Cutrona et al (1960) suggested the processing of 2-dimensional information using optical systems. The fact that a 2-dimensional input image may be transformed into its 2-dimensional Fourier transform at near instantaneous rates (literally "speed of light") has opened the door to allowing complex operations on images which were hitherto time consuming or impossible to perform digitally. This is not to say that electronic and numerical image processing techniques will become obsolete overnight. Evidently this has not happened and as is shown in later chapters such techniques are, so far, instrumental in the further development of this branch of science.

This chapter introduces the major concepts of the subject and discusses the factors which have an influence on optical image processing techniques. The theories supporting this study area are presented together with optical system descriptions and illustrations. Current developments in compact correlator design are

described. Other considerations are the effects of pixellation on image processing and a portion of this chapter is devoted to a description of the factors involved.

The 4-f coherent optical processor

The most commonly applied implementation of an image processing bench is shown schematically in Figure 1.1(a) and is referred to as a 4-f coherent optical processor, where f is the focal length of the individual lenses used (it should be noted, however that there may be several values of f in a single bench as different focal length lenses may be used in order to vary the overall magnification of such a bench). The value of 4-f represents the total distance between input and output planes (plane 1 and plane 3 in the figure)

A laser beam passes through a spatial filter assembly attached to the front of the laser and thus appears as a point source of monochromatic light. By placing a lens (L_1) such that the point source is at the front focal plane of the lens, a parallel beam results which provides a collimated and uniform illumination of the input plane (this, of course, is subject to the constraints of using Gauss's approximations to the wavefront). The parallel beam is allowed to pass through the object (positioned coincident with plane 1) which is situated at the front focal plane of the first transforming lens (L_2). Under these circumstances, the light field at the back focal plane of the lens has the form of the scaled two-dimensional Fourier transform of the transmission function of the input plane. The position of this plane is coincident with the front focal plane of the second transforming lens and is referred to, henceforth, as the Fourier plane of the system.

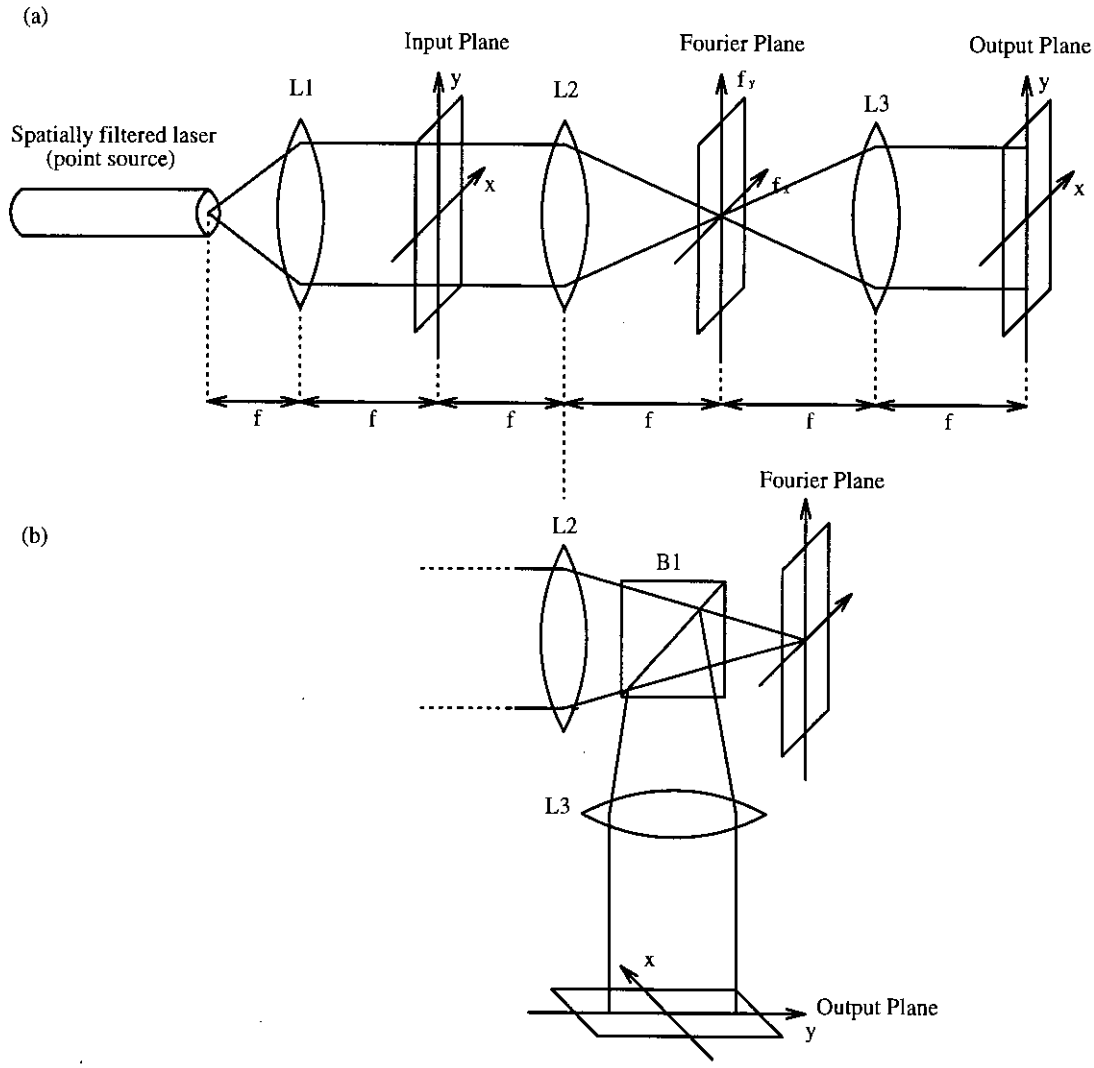


Fig 1.1 (a) The 4-f optical processing bench architecture; (b) reflective mode Fourier plane

The contents of the Fourier plane are then transformed once more (by lens L_3) to produce an inverted output image at the front focal plane of the second lens. This assumes an ideal optical system.

Any modification of the image is typically carried out at the Fourier plane where the object has been decomposed into its spatial frequency components which are arranged with the zero-frequency (DC) component in the centre and higher spatial frequencies occurring at larger radii; the relationship between spatial frequency and position in the Fourier plane (with respect to the optical axis) is

linear, as is also shown in Appendix 1. Figure 1.1(b) shows an alternative geometry which must be used when the modulating device used is a reflecting device (often the case when utilising a spatial light modulator). The lenses are labelled so as to indicate their equivalent rôles as described above but functionally the two geometries are identical (although aberrations in the form of astigmatism are introduced by the beam splitter).

If no modification has occurred, then the output image will simply be the same as the object but rotated by 180 degrees, as the lenses execute forward Fourier transforms only, and the effect of two successive forward Fourier transforms is to invert the image relative to the object. Of course the ability to modulate the light from the object in some subjective way is the sole reason for developing such a system and this is achieved by filtering at the Fourier plane. The function resulting from the use of a Fourier plane filter is derived in Appendix 1B and is shown to be the convolution of the Fourier transform of the filter function and the co-ordinate-reversed image of the object.

For the case where the object does *NOT* lie on the front focal plane of the first transforming lens a multiplicative quadratic phase factor will be included in the Fourier transform. The Fourier transform lies in fact on the surface of a sphere but the radius of this sphere will in most cases be sufficiently large for the planar approximation to hold (Goodman (1968)).

Modulation at the Fourier Plane

The discussion above is based on the assumption that ideal lenses are being used as Fourier transforming lenses. Dealing, as we are, in the real world, there are two more factors to be taken into consideration: the point spread function (or PSF)

of the lens and the bandwidth of both object and lens. The PSF of a lens is defined as the response of the optical system as measured in the back focal plane of the lens to a point impulse situated at the front focal plane of the lens and this is illustrated in Fig 1.2.

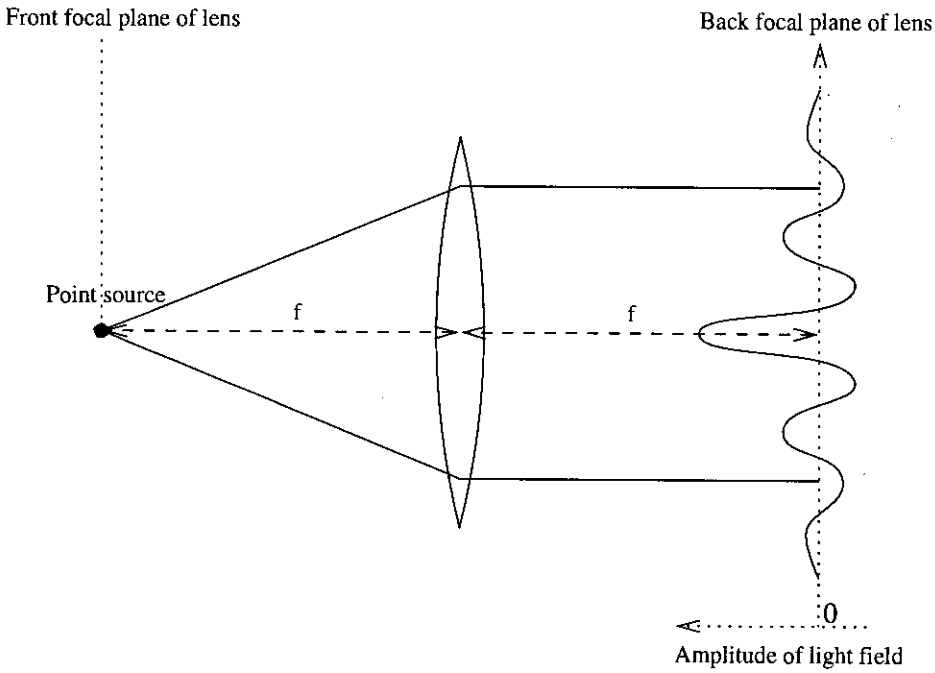


Fig 1.2. Illustration of the Point Spread Function of a typical lens

For a typical lens, the PSF is influenced by the presence or otherwise of lens aberrations as well as the diffraction effects brought on by the introduction of a finite aperture into the system. The bandwidth refers to the maximum spatial frequencies which may be supported by the object. For a system containing a finite aperture (or lens), the bandwidth of the optical system is limited also by the diameter of the aperture - spatial frequencies above a certain level will be blocked by such a finite aperture and object information may thus be lost.

The modulation of the Fourier plane may take many forms, including amplitude-only, phase-only, combined amplitude and phase or polarisation

modulation. In the following sections the case of amplitude modulation at its most basic (using aperture stops to block certain areas of the Fourier plane) is illustrated.

Image sharpening (high pass filtering)

The visibility of details within an object (edges or lines where there is a high intensity gradient) may be increased while that of uniformly transmitting areas is reduced by edge enhancement. The operation involves the attenuation of the lowest spatial frequency components of the Fourier transform of the object used as input to the system. By examining the reciprocal relationships defined in equation A1.3 (Appendix 1) this is implied implicitly as the lowest spatial frequency components are characteristic of large features in the object plane.

The spatial frequency component of most interest in this case is obviously the central component which corresponds to the zero spatial frequency and is commonly referred to as the D.C. component. By removing this, the mean of the amplitude distribution of the image subsequently formed is reduced to zero, the image being formed of positive and negative amplitudes as opposed to the (real) object whose amplitude components are all positive. Taking the squared modulus of the amplitude distribution (as all optical detectors are intensity sensitive) results in an image where positive and negative amplitudes give rise to positive intensities, and areas of zero intensity where the amplitude is zero. These areas correspond to those regions in the original object where the amplitude crosses the amplitude mean level (rising or falling) and therefore, in the image, edges will appear to be dark lines and the overall contrast will have been reduced.

One of the undesirable effects of this technique is the unavoidable exaggeration of any noise present in the object, as the high spatial frequencies are emphasised and noise terms tend to be within the high frequency range.

Image smoothing (Low-pass filtering)

The converse of the above method is one in which the higher spatial frequencies are clipped (using an annular filter rather than a block) and the object is low pass filtered. From the arguments above, this means that the image resulting will have large features enhanced at the expense of the fine details. In the field of electronics this technique is used to reduce the presence of noise and the optical applications are closely related to this.

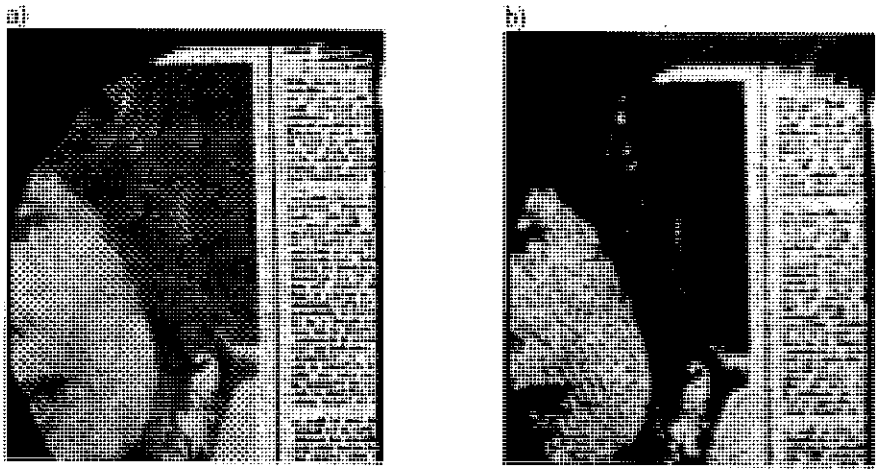


Figure 1.3(a) Half-toned newspaper image; (b) Low pass filtered version of (a)

An example is of the removal of the dot structure from within a newspaper photograph (Figure 1.3(a)) where the halftone dot structure has a higher spatial frequency than the image which it goes to create. When using such an object in a 4-f bench, the Fourier plane contains a grid of bright dots (referred to as spectral orders - see *Pixellation in the Fourier Plane.*) which are related to the dot structure of the picture. By passing only the central dot and spatial frequencies up to half the

distance to the neighbouring spots, the information about the dot structure is removed from the Fourier plane and on reconstruction of the image, only the main features of the object can be seen (Figure 1.3(b)). These pictures were obtained by taking a photograph of the output plane of the 4-f optical bench illustrated on page 1 and using an annular filter placed in the Fourier plane to remove the higher spectral orders from the frequency domain.

Edge removal

As the Fourier transforming property of a lens is a diffraction effect, it follows that a feature (or series of features) oriented at an angle θ° to the horizontal will generate spatial frequency components in the Fourier plane physically oriented at an angle $(\theta+90)^{\circ}$ to the horizontal. By blocking the $(\theta+90)^{\circ}$ portion of the Fourier transform, it follows that the reconstructed image will have had the θ° features removed.

This is illustrated using an image generated in this group (Fig 1.4(a)) of a house whose features (doors, roof, windows etc.) are formed from a series of transmitting lines at different orientations. The Fourier transform of the image is shown in figure 1.4(b). Blocking the horizontal components of the Fourier transform, using a simple narrow filter, will remove the walls from the image as they are composed of vertical lines (figure 1.4(c)). Rotating the filter block further to an angle perpendicular to the lines which form the house door will selectively remove the features perpendicular to that block and this is indeed shown to be the case (figure 1.4(d)).

Although this is an artificially generated object manufactured to illustrate precisely this point, the technique may nonetheless be utilised in practical situations

and with 'real' objects. In this case, once more, the optical bench from figure 1 was used, and a block consisting simply of a short, straight length of wire attached to a rotating mount replaced the annular filter from the previous example.

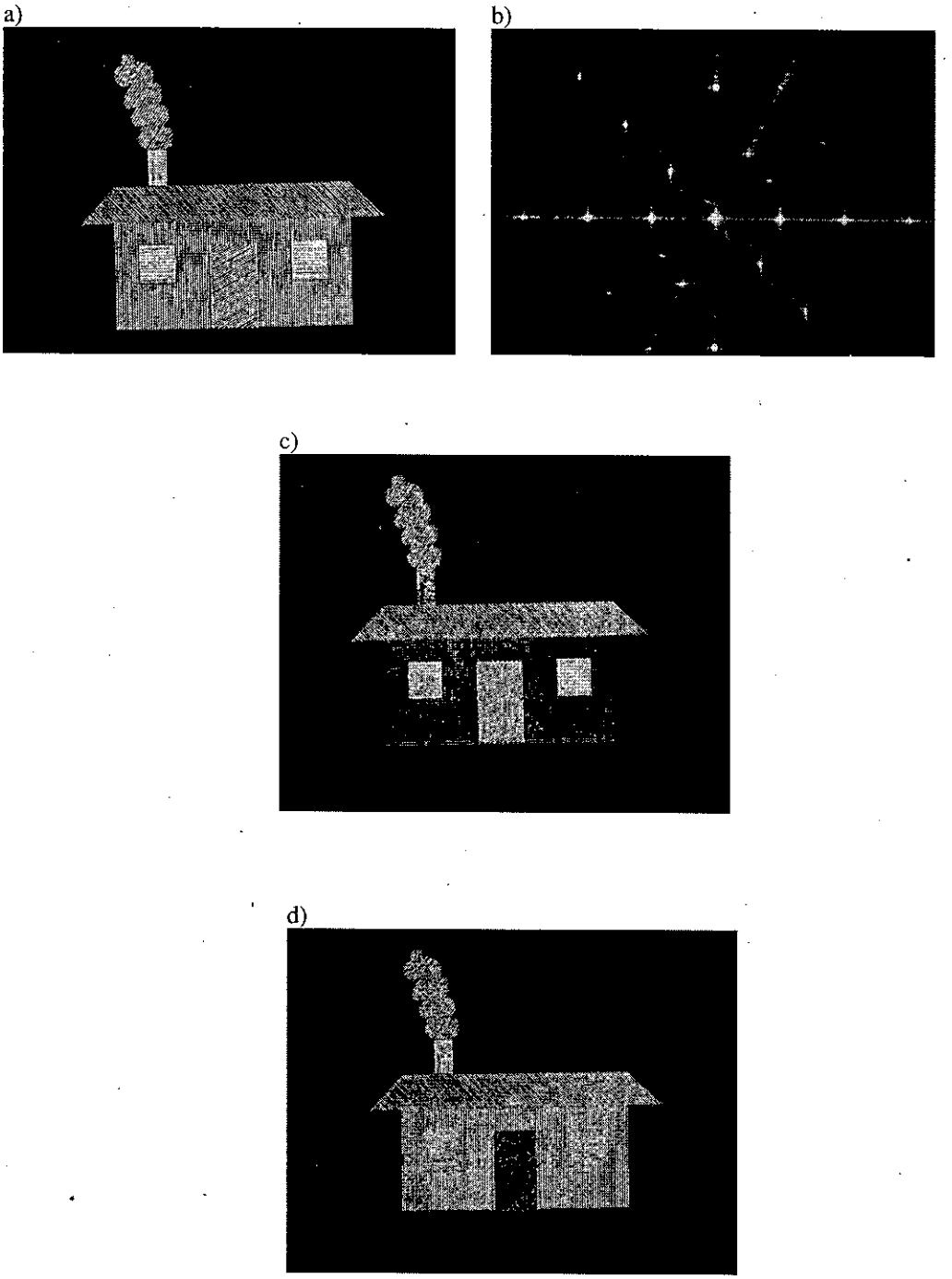
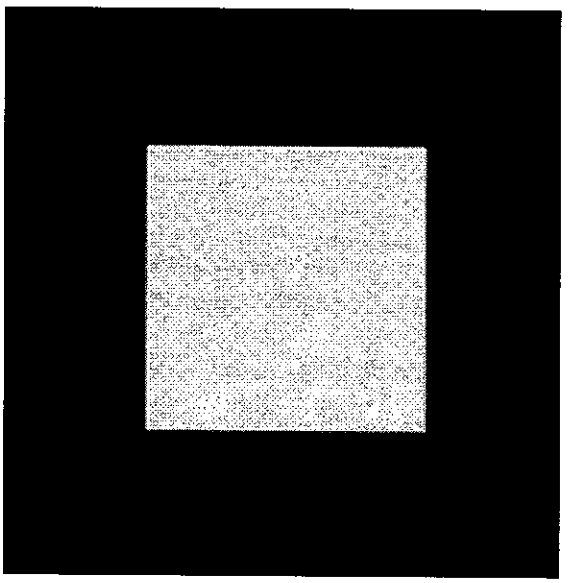


Fig 1.4(a) Object; (b) Fourier spectrum of (a);
(c) Output constructed after removal of horizontal $\theta = 90^{\circ}$ component of (b);
(d) Output reconstructed after removal of 30° component of (b).

Fig 1.5(a) shows a square object of uniform irradiance with fig 1.5(b) being the resulting output on blocking the horizontal component of the Fourier transform of the square (including the DC component). The spatial orders relating to the vertical edges are removed and the object is also edge enhanced. This has the effect of isolating the horizontal edges from the original object.

a)



b)

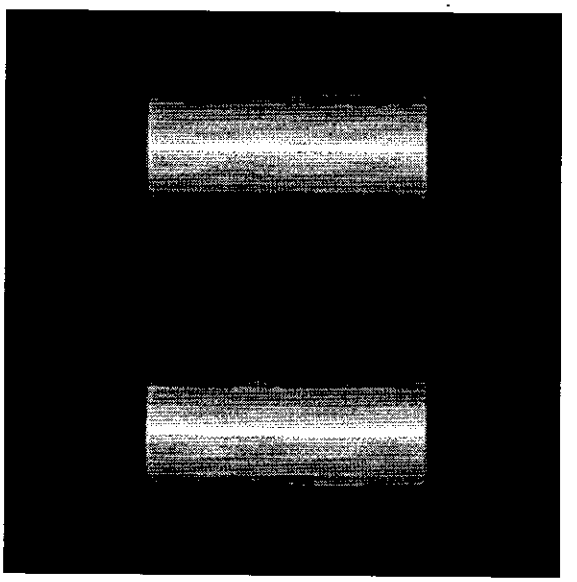


Fig 1.5(a) Uniformly illuminated square object; (b) Reconstructed image after blocking the horizontal component of the Fourier transform of (a) (simulation)

Pixellation in the Fourier plane

The advent of spatial light modulators as programmable optical filters has increased the flexibility of image processing arrangements in that a device has merely to be reprogrammed rather than physically replaced in order to achieve some alternative image processing function. However, these devices have tended to be pixellated and this leads to other (sometimes insurmountable) problems which must be taken into consideration.

In order to examine this fully, it is necessary to return to first principles in Fourier transformation theory and use the case of a one-dimensional transform which may be generalised into two dimensions. As we have shown previously, the output image from a Fourier filtering 4-f bench is the convolution of the input image and the Fourier transform of the filter transmission function. Consider a filter consisting of a regular (one-dimensional) array of identical pixels with dimension a and pixel spacing b (fig 1.6)

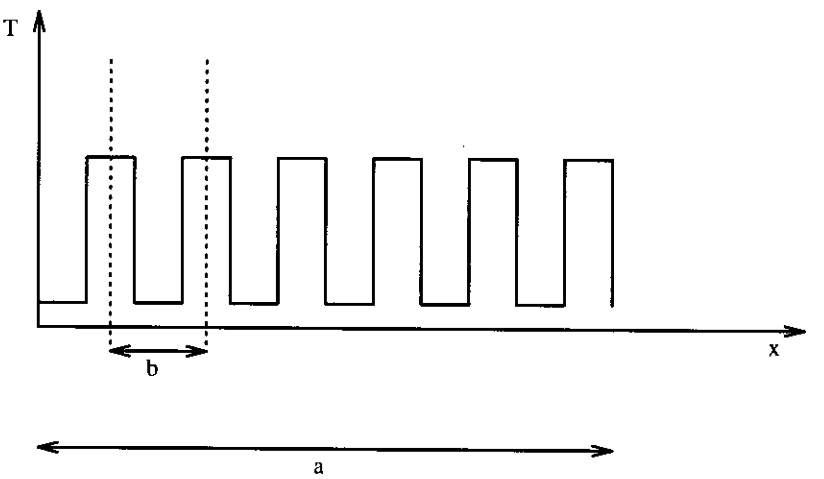


Fig 1.6. One dimensional pixel array: b is the pixel spacing; a is the array dimension

This may be mathematically represented by the function:

$$\tau(x) = \text{rect}\left(\frac{x}{a}\right) \times \left[\text{comb}\left(\frac{x}{b}\right) * p(x)\right] \quad \text{... equation 1.1}$$

where * represents convolution, × represents multiplication, $p(x)$ is a (possibly complex) function representing the transmission of a single pixel (and therefore of finite extent, having the same extent as the pixel with value zero outside) and the functions $\text{rect}\left(\frac{x}{a}\right)$ and $\text{comb}\left(\frac{y}{b}\right)$ are defined as :

$$\text{rect}\left(\frac{x}{a}\right) = \begin{cases} 1, & -\frac{a}{2} \le x \le \frac{a}{2} \\ 0, & \text{otherwise} \end{cases}$$

$$\text{and } \text{comb}\left(\frac{x}{b}\right) = \sum_{-\infty}^{+\infty} \delta(x - nb) \text{ (where n is an integer)}$$

$\delta(x)$ denotes the delta-function. Applying the convolution and similarity theorems (Bracewell (1986)), it is possible to obtain an analytical expression for the Fourier transform of equation 1.1 and this gives

$$T(f_x) = \text{sinc}(\pi a f_x) * \left[\text{comb}(b f_x) \times \mathcal{P}(f_x)\right] \quad \text{... equation 1.2}$$

where $\mathcal{P}(f_x)$ is the Fourier transform of the pixel function, and $\text{sinc}(x) = \frac{\sin(x)}{x}$

The remaining variables retain their meanings.

The net result of using such a filter in the Fourier plane of an image processing system is that the output plane of the system will consist of the convolution of the coordinate reversed object with equation 1.2, resulting in multiple images due to the broadening of the delta function peaks caused by convolution with the sinc() function. The magnitude of the peaks are also attenuated by the Fourier transform of the filter pixel function p().

For large values of b, the pixel separation, the delta functions in equation 1.2 become closer. At some critical value of b, (referred to as the *critical sampling rate*) they become so close that the modulated sinc() functions with which they are

convolved touch and at larger values of b these actually start to overlap. This results in information from adjacent orders leaking into each other causing a corruption of the image. This is referred to as *aliasing*.

In order, therefore, to calculate the resultant effect on the optical system of introducing such a pixellated filter into the Fourier plane it is obvious that the pixel transmission function, pixel separation and device size all have a large part to play. It is then possible to deduce that for such a system:

i. The envelope of the function $T(f_x)$ is governed by the Fourier transform of the pixel function $p(x)$. For a simple pixel described by a simple $\text{rect}(x/c)$ function, this envelope is a wide $\text{sinc}(\pi c f_x)$ function with zeroes at $cf_x = 2n$ where n is an integer.
ii. The spacing of replications in the image plane (and therefore the maximum input image size) is governed by the pixel separation (b) as a direct consequence of the comb function produced in equation 1.2.
iii. The *point spread function* (PSF) of an ideal optical system is governed by the physical size of the filter (a) and results only from diffraction effects. Taking a wide (large a) device, the associated PSF will be a narrower sinc() function than for a narrow (small a) device.

The ideal filter device, it would appear, is infinitely wide with a δ-function for a pixel function and an infinite number of pixels. In practice this is not possible and there will obviously be trade-offs and corresponding losses in overall performance. The main one to be taken into consideration here is the conflict between i and ii above. The finite dimensions of the device restrict the amount of the Fourier plane information which may be transmitted on through the system - a

smaller device will pass fewer spatial frequencies than a larger device and so will act as a more severe low-pass filter. This may be countered by the use of a shorter focal length lens which has the effect of reducing the scale of the Fourier transform. The conflict arises as, in order to increase the replication separation in the output plane, the spacing of the pixels in the Fourier plane must effectively be reduced. As the pixellated filter is unlikely to be flexible enough physically to alter the pixel separation, in general a longer focal length lens is utilised which rescales the Fourier transform to cover a larger area of the Fourier plane (Appendix 1).

It can be shown that the product of the Fourier plane sampling rate and the Fourier plane pass band (the *space-bandwidth product* or *SBP*) is fixed and is equal to the number of pixels across the filter; a bigger dimensioned SLM is not necessarily better as the important factor is the sampling rate of the device.

The effect of sampling the Fourier plane using a pixellated filter is shown in figures 1.7(a) to (d). The objects used were two annuli with the same proportions but with different dimensions (from the above reasoning it can be seen that increasing the dimensions of the object is *exactly* equivalent to reducing the sampling frequency). In figure 1.7(b) the Fourier plane is sampled at the critical rate and the image contains replications of the object which *just* avoid overlapping. Figure 1.7(c) shows the result of using the same sampling frequency but with a larger object (exactly equivalent to reducing the sampling frequency for the original object). It is clear from the photograph that the object is being under sampled: the replications of the image overlap to a large extent and the output image is hopelessly complicated.

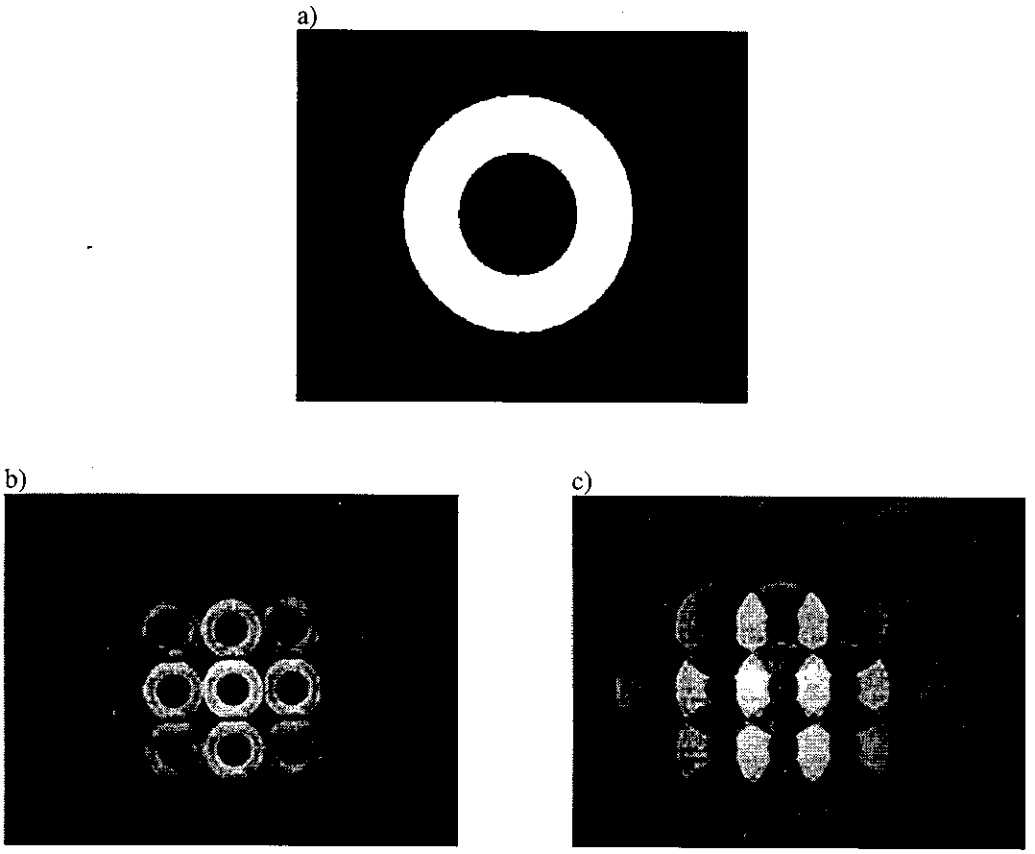


Figure 1.7(a) Annular object; (b) object (a) critically sampled in the Fourier plane; (c) Object (a) over sampled in the Fourier plane.

All of the above arguments may, with equal validity, be applied to a pixellated object. As is shown in the section on half-tone removal earlier in this chapter, a pixellated object gives rise to replications of the spatial frequency spectrum in the Fourier plane. From the work outlined above, this can be seen to be a direct result of sampling the input image in order to represent it on a pixellated filter and the replicated Fourier transforms are (separately) the transforms of the unsampled object. In order to counter this, one may either perform the same filtering operation on each of the spectra (in order to use all the available illuminance) which is an extremely complex task and impossible to carry out with the limited resolution and bandwidth of available devices, or more simply, block all replications but one as each contains all of the required information and the replications serve only to encode the pixellation of the object under study. Practically it is only the most

intense DC order[1] which is passed in these circumstances and this is then filtered conventionally. The highest spatial frequencies of the spectra overlap and should not be used to construct the output image as aliasing will result. As McKnight (1989) puts it - this is just another way of saying that one cannot write more detail onto a filter than there are pixels available.

Mapping considerations and computer simulations

There follows a brief discussion on mapping the Fourier plane within the optical regime and in computer simulations.

When placing a filter in the Fourier plane it is preferable to have the filter cover the Fourier spectrum exactly i.e. without sampling second order and greater spectra to any extent in order to avoid the problems of aliasing (see previous section). By examining the equations relating position in the Fourier plane to feature size, it is possible to deduce that there is a maximum object size which may be thus sampled by a given (fixed size) pixellated Fourier plane filter. The dimensions of this object may be calculated by arranging for the spatial frequency component corresponding to the minimum (non zero) spatial frequency of the object to fall on the centre of the pixel adjacent to the central pixel in the Fourier plane. i.e. $f_{min} = \frac{b}{\lambda \times f}$ using equation A1.3 (Appendix 1). The inverse of this equation gives an expression for the maximum dimension of an object which may be optimally sampled in the Fourier plane by such a pixellated object:

$$D_{max} = \frac{\lambda \times f}{b}$$

[1]Note this is *not* the same as the DC frequency. Here we are referring to the spatial frequency spectrum located on the optical axis and not the zero spatial frequency component of the Fourier transform.

where λ is the wavelength of light used, f the focal length of the lens and b the pixel spacing. It may also be seen that the spatial frequency cut-off of this system is given by:

$$f_{max} = \frac{Nb}{2\lambda f}$$

where N is the number of pixels in a given dimension and therefore Nb is the linear dimension of the SLM (Hossack et al (1991))

It is obvious from the above equation that with more pixels available in the Fourier plane filter (for a given SLM size), a larger object size may be supported due to the reduced pixel spacing, b. Conventional theory states that for optimal mapping in the Fourier plane the Fourier transforming lens must be chosen to suit the dimensions and features of the pixellated filter being used and studied.

Computer Simulations

One of the primary tools of the researcher in the field of optical image processing is the Digital Fourier transform (or DFT). This is merely a digital representation of the analogue Fourier transform function, with integrals replaced by summations. The Fast Fourier Transform (or FFT) was developed as the basic DFT was an extremely inefficient algorithm, with the number of calculations required scaling as n^4 for the two-dimensional calculations required for optical system simulation. The FFT (e.g. Spanier and Oldham (1987), pp 308-309) utilises the fact that many calculations are repeated in the original algorithm and pre-calculates many of the elements of the equation resulting in an algorithm within which the number of calculations requires scales as $n^2 \log_2(n^2)$ - obviously a huge saving in effort and computer cycles.

The use of computers in simulating optical systems requires that the various signals be rigidly pixellated. It is possible to use arbitrarily large precision in representing an analogue object on a pixellated plane but the object is, nonetheless, definitely pixellated - the requirement is simply for more pixels to represent the same object. For the purposes of the following discussions, the pixellation of the object under study will be referred to in terms of 'object pixels' whereas (as each object pixel may be represented by an arbitrary number of data points when the object is represented as a computer datafile) the term 'computer pixel' will refer to the sub-pixels forming the object pixels (Figure 1.8).

In the illustration, each 'object pixel' is represented by a 3x3 array of 'computer pixels'. For the more general case, each object pixel of an N×N pixellated object may be represented by an M×M array of computer pixels and embedded in a (larger) array of A×A computer pixels.

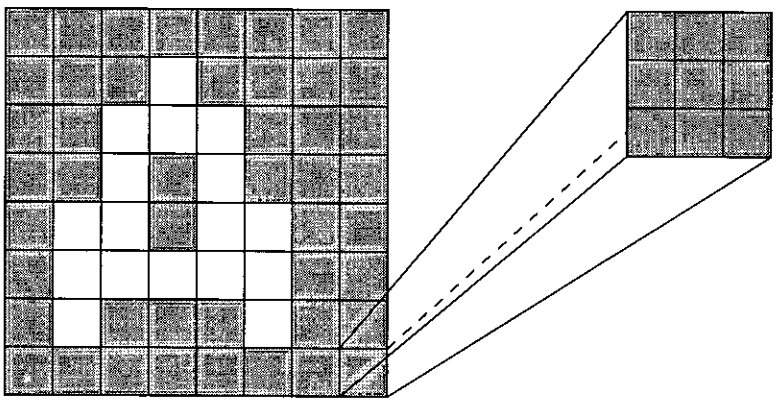


Fig 1.8. 8 by 8 pixel object represented by a 24 by 24 pixel computer array (N = 8; M=3).

The level of pixellation in the computer plane (magnitude of M in the general representation above) is of great importance as it determines the extent of the Fourier transform of the computer image. The space-bandwidth product applies such that the Fourier transform of an N×N object pixel array expanded to (M×N)×(M×N) and embedded in an A×A array of computer pixels will cover only A/M computer pixels,

so resolution has been lost in the Fourier plane. Ideally, M=1 and N=A as in that case the object is represented at its original resolution, and the Fourier transform plane is at optimal resolution also. This is not always possible when for example one requires more resolution to represent successfully a more complex pixel function in the Fourier plane or, when multiple inputs are supplied. In these cases, the object(s) must be embedded in a larger dimensioned zero intensity background and the Fourier plane filter correspondingly scaled to cover the entire plane.

Chapter 2

Liquid Crystals and Spatial light modulators

Introduction

The building block upon which all optical processing architectures and techniques are founded is the *spatial light modulator*. This is simply any device which can "impress information onto an optical wavefront"; this encompasses a huge range of possibilities, from photographic transparencies to the whole gamut of optically and electronically addressed devices available and under research at this time. They are currently in use in correlation experiments (e.g. Psaltis et al (1984)), optical neural networks (Burns et al (1994)) and as optical logic blocks (Weigelt (1987), Wherrett and Tooley (1989)) which allow simple logical functions (AND, OR, etc.) to be realised optically. The simplest of these devices, the blocks and transparencies are of little relevance or interest in the context of this thesis and attention will be concentrated on the classes of real-time programmable modulators.

Characterisation of the spatial light modulator must explore the various methods of light modulation possible, addressing problems and solutions and the many performance factors which affect them. This chapter will start as a generalised

introduction to the science with an examination of specific devices and applications in the latter sections.

In the context of this thesis, the spatial light modulator walks hand in hand with the science of the liquid crystal. Many of the most popular and successful SLM's reported are based on liquid crystal technology and it is informative to provide a brief summary of that particular area of study. More detail is not provided as it is inappropriate here and is in any case widely available in other references (eg Armitage et al (1987, I and II)

Liquid crystal technology

The liquid crystal state is a phase of matter exhibited by a large number of organic compounds over a limited thermal range - as the temperature rises, the material changes from a crystalline solid to a clear liquid. Within these limits, the material combines some of the optical properties of solids with the fluidity and behaviour of liquids. For all liquid crystals, however, a major characteristic is the elongated rod-like form of their molecules and obviously from this asymmetry the relative orientations of the individual molecules is very important. It is useful, therefore to speak in terms of the vector direction of the molecule, known as the *director* and referred to symbolically as $\underline{n}$.

Broadly speaking there are three basic liquid crystal types (or mesophases) and these are termed *nematic*, *cholosteric* and *smectic* - a summary of their characteristics and physical properties follows below. One of the most important electrical (and optical) characteristics exhibited by this class of materials is their dielectric anisotropy i.e. they show different dielectric constants $\varepsilon_{\parallel}$ and $\varepsilon_{\perp}$

depending on whether an electric field is applied parallel or perpendicular to the molecular axis. If $\varepsilon_{\parallel} > \varepsilon_{\perp}$ then the material is said to exhibit a positive dielectric anisotropy and applying an external electric field to such a material will cause the molecules to align themselves such that their directors are parallel to the applied field in order to minimise their energy. In practical situations the applied voltage tends to be a square wave A.C. voltage as a D.C. supply will tend to cause electrochemical degradation of the liquid crystal over time. Generally the aim is to maintain a zero mean of the applied voltage as far as is possible.

Nematic Liquid Crystals

In the nematic state, the molecules tend to orientate themselves such that their directors lie parallel to each other. This mesophase is, however, subdivided into two orderings:

1) *homeotropic*, in which the director of the molecule lies perpendicular to a liquid-solid interface (fig 2.1(a)), and

2) *homogeneous*, where the director lies parallel to the interface (fig 2.1(b)).

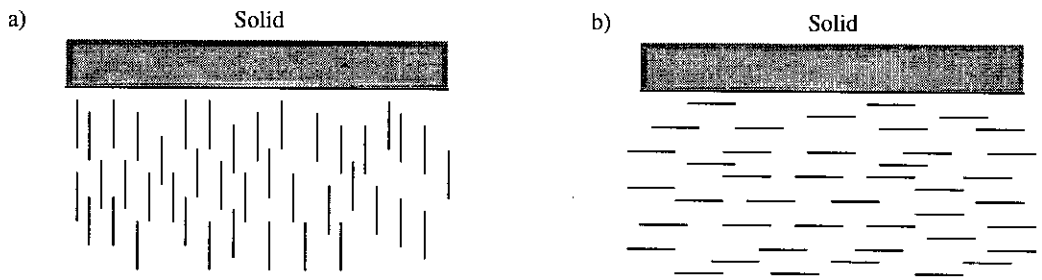


Fig 2.1(a) homeotropically ordered and (b) homogeneously ordered nematic liquid crystal

This orientational order may be altered by a suitable treatment of the neighbouring solid surfaces which induces the director to align accordingly (Cognard (1982)). Homogeneous alignment may be achieved by arranging for grooves of microscopic

dimensions to cover the liquid crystal-solid interface causing the crystal molecules to preferentially align parallel to the grooves. Two common methods of achieving this are rubbing microgrooves into a thin layer of polyvinyl alcohol or by evaporating a thin stratum of silicon monoxide at an angle to the substrate, forming a regular series of ridges.

Modulation Effects

As has been noted above, nematic liquid crystals exhibit a high degree of dielectric anisotropy. For liquid crystal molecules with a positive $\Delta\varepsilon\,(=\varepsilon_{\parallel}-\varepsilon_{\perp})$, the presence of an external electric field will cause the molecules to align themselves parallel to the field whereas those with a negative $\Delta\varepsilon$ will align themselves perpendicular to the applied field. There is an associated birefringence (variation in the refractive index, *n* depending on the polarisation state) which is defined by:

$$\Delta n = n_e - n_o$$

where n_e is the refractive index of the substance as seen by light polarised so that the electric field vibrations are parallel to the director (extraordinary), and n_o the refractive index as seen by light polarised with the electric field vibrations perpendicular to the director (ordinary). As it is possible to alter the alignment of the director, so it is possible to tune the birefringence of a layer of liquid crystal.

Field Induced Birefringence

Examining the simplest cell geometry for nematic liquid crystals, we take the situation of Figure 2.2. A liquid crystal exhibiting positive dielectric anisotropy is contained by two glass plates, with an alignment layer inducing a homogeneous ordering. In this case, an applied electric field will cause the director everywhere to

preferentially align parallel to the field[1], thereby changing the material to a homeotropic state.

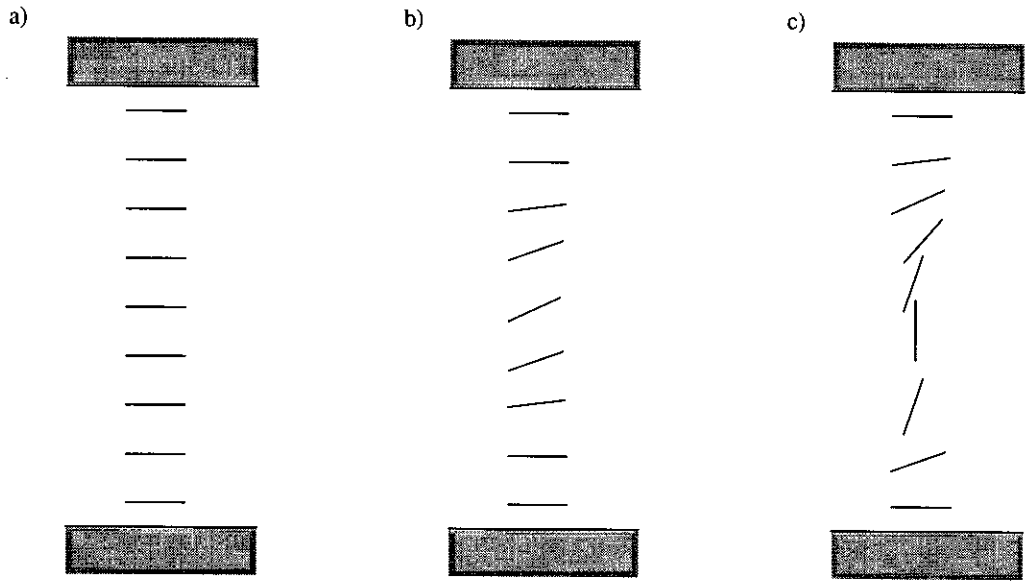


Fig 2.2. Homogeneously ordered nematic liquid crystal reacting to an increasing electrical field, ε

With no applied field, light incident with a polarisation parallel to the director will see a refractive index of n_e and light polarised perpendicular to the director will see a refractive index of n_o (Fig 2.2(a)). On application of an electric field above a critical magnitude, (Fig 2.2(b)) the molecules will tilt towards the direction of the field. As the applied field increases (Fig 2.2(c)), the refractive index seen by any polarisation of incident light will tend to the value n_o. The cell, in effect allows variation of the extraordinary refractive index within the range:

$$n_e \geq n \geq n_o$$

according to the magnitude of the applied field.

By polarising the incident light at an angle ϕ it is at once obvious that the two polarisation components of the light will experience different path lengths within the

[1]Note that molecules nearest the glass surfaces will be held more strongly in their favoured alignment orientation and those in the centre of the cell will be affected most by the applied electric field.

cell depending on the applied field. The ordinary component (polarisation parallel to the director) undergoes an optical phase delay of:

$$OPD_o = \frac{2\pi}{\lambda} n_o L \text{ radians}$$

where L is the thickness of the cell. The case for the extraordinary component is much more complex as the refractive index n is a function both of position within the cell, and of the applied voltage. Here, however, we are interested in the macroscopic rather than microscopic situation and it simplifies matters to refer to an averaged refractive index function which is a function only of the voltage and is represented by $n_e(V)$ where V is the applied voltage. The optical path length traversed by the extraordinary component is, therefore:

$$OPD_e = \frac{2\pi}{\lambda} n_e(V) L \quad \text{...Equation 2.1}$$

The difference between these expressions corresponds to the phase difference introduced between the extraordinary and ordinary incident rays. For the situation where incident light is polarised using a polariser with an axis oriented at 45° to the alignment direction, analysing the emitted beam with a polariser at 90° to the first results in a transmitted intensity of :

$$I(V) = I_0 \sin^2\left(\frac{\Delta\phi(V)}{2}\right)$$

with I_0 being the maximum intensity transmitted and $\Delta\phi(V) = OPD_e(V) - OPD_o$, the phase difference between the extraordinary and ordinary rays.

For those values for which $\Delta\phi = 2m\pi$ where m is an integer, I(V) goes to a minimum. This point may be chosen by selectively setting V and so obtaining an amplitude modulating device.

By simply polarising the incident light parallel to the extraordinary optical axis, equation 2.1 indicates that the cell will function as a phase modulator with a *phase modulation depth* of

$$\Delta\phi(V) = \frac{2\pi}{\lambda}\left(n_e(V_0) - n_e(V)\right)L$$

and with $n_e(V) = n_o$ for sufficiently large V.

In the case of the E7 nematic liquid crystal from BDH, $n_e - n_o = 0.224$ and so, for maximum phase modulation ($\Delta\phi = \pi$) the cell need only be a few microns thick. It is the large birefringence which makes this a particularly good choice for experimental work.

Twisted Nematic Effect

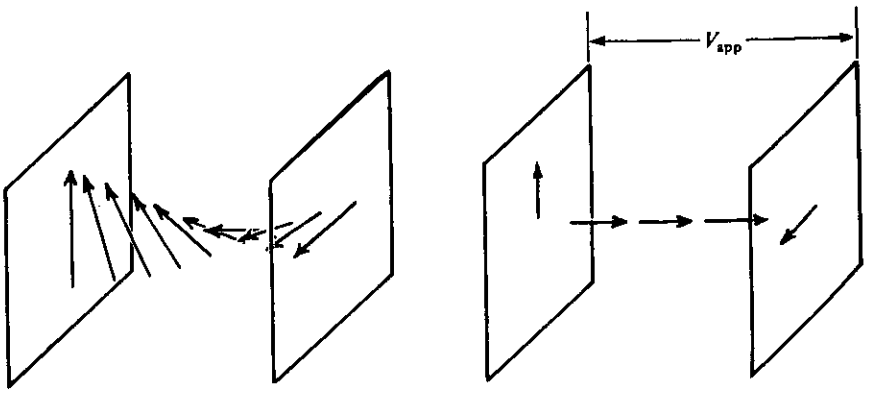


Fig 2.3. Twisted nematic effect (after Wilson and Hawkes(1983), p164).
(a) No applied field; (b) large applied field.

A common implementation of the nematic liquid crystal in optical displays utilises the twisted nematic effect. (Fig 2.3, Wilson and Hawkes (1983)) A cell is created such that the walls of the cell are treated to produce a homogeneous liquid crystal ordering in which the molecular directors at the walls are at right angles to each other. This means that the liquid crystal molecules undergo a 90° rotation

across the cell (Figure 2.3(a)). When a polarised light beam is allowed to fall on the cell, the anisotropy of the liquid causes the plane of polarisation of the light also to be rotated by 90°. On application of a strong enough electric field across the cell, the directors will align perpendicular to the cell walls and the incident polarised light beam will be unaffected (Figure 2.3(b)).

In practical situations, the cell is sandwiched between crossed polarisers (whose polarising directions lie parallel to the surface director on the nearest cell wall). Incident light is either rotated by the cell (zero V applied) and is therefore passed by the analyser, or for an applied voltage V_a the light is unrotated as the liquid crystal is forced into a homeotropic ordering, the polarisation plane of the incident light is unrotated as the medium exhibits no chirality and the light is blocked by the second polariser.

One way in which this operation may be impaired is if the handedness of the twist is reversed over areas of the cell. A simple solution is to apply some 'pre-tilt' to the molecules at the cell walls which favours one handedness over the other by reducing the associated strain energy. A second solution is to dope the nematic liquid crystal with a chiral liquid crystal mesophase which again breaks the symmetry of the nematic structure.

The hybrid field effect

The hybrid field effect (Grinberg et al (1975)) uses a half-twist cell where the directors at the opposite surfaces lie at 45° to each other and the cell is used in reflection (i.e. two-pass) mode. With no applied field, the input polarisation plane is rotated (as in the twisted nematic effect) due to the chirality of the cell. Reflecting

the light field back through the cell causes it to be rotated back to its initial state and by reflecting this beam through an analyser (at 90° to the input polariser) no light is transmitted. On application of an electric field, the birefringence is reduced and the polarisation state of light reaching the mirror is elliptical. A second pass through the cell in the reverse direction produces a larger change in the polarisation state and light will be transmitted through the analyser.

It should be noted that this is the most commonly exploited effect in current liquid crystal display technology (e.g. in watches, calculators and other such displays). Most practical SLM's do not use this type of liquid crystal.

Ferro-electric Liquid Crystals

While it is inappropriate to dwell too much on the alternative liquid crystal technologies in the context of this thesis it is necessary to mention the ferroelectric liquid crystal state as it has become extremely popular among displays research groups. This class of liquid crystal originates in the smectic state which are identified by short range positional ordering and long range orientational ordering characteristics; the molecules forming layers of similar orientations. It is possible to design the molecules such that the substance exhibits a chirality in the orientation of the directors of the molecules and in this case the liquid crystal will have a microscopic spontaneous polarisation which is non-zero, due to the overall molecular symmetry, and will therefore be ferroelectric (the direction of the spontaneous polarisation may be altered by the application of an external electric field). In general, the molecular directors of successive chiral smectic layers forms a helix and so the overall (macroscopic) spontaneous polarisation averages to zero but

it is possible to suppress this and so obtain a liquid crystal which exhibits a macroscopic non-zero dipole (and spontaneous polarisation).

Typically, the ferroelectric sub-classes most exploited are the smectic-A (S_A) and smectic-C (S_C) LC's - the differences between ferroelectric LC's and their nematic counterparts being:

i. FLC cell is more difficult to manufacture reliably due to the higher viscosity and reduced film widths utilised.
ii. Nematics are more resilient to cell thickness variations.
iii. Ferroelectrics are less tolerant of surface treatments on the electrode-LC boundaries.

At the time of study, the ferroelectric liquid crystals, although interesting, were in the primary stages of research. Devices were no more than prototypes and the science was still at an early developmental stage. There was no alternative but to attempt to use the existing nematic technology and, in the context of this thesis, which aims primarily at the simulations and potential of correlation algorithms, the choice of medium was not strictly relevant.

Spatial Light Modulators

Light Modulation

A beam of linearly polarised monochromatic light propagating in the z-direction (in a Cartesian coordinate system) may be modulated spatially or temporally or both by a general modulation function M(x, y, t). In the cases to be discussed, the modulation is not temporal but the spatial modulation is variable

within discrete time blocks. It may be said, therefore that the function M is a function of x, y and T_n where T_n is the time interval over which $M(x,y;T_n)$ is kept constant.

The cases for incoherent and coherent spatial light modulation are slightly different - incoherent light may have only its intensity or polarisation modulated with any effect, but coherent light may be modulated through combinations of phase or amplitude or polarisation.

Addressing Schemes

There are two main subsets of spatial light modulator: optically addressed and electronically addressed. Optically adddressed SLMs are activated by a 'write' beam which impinges on the device and causes the modulation to occur via some optical interaction. A 'read' image is then reflected from the device and is modulated according to the form of the write beam. Electronically addressed devices utilise electro-chemical, electro-physical, electro-optical or even opto-mechanical interactions to modulate a beam of light which may either be reflected from the device or transmitted by the device having been modulated in the process.

Many spatial light modulators used today are electronically addressed 'hybrid electro-optic' devices. This is primarily because of the dimensions involved which current electronic technology can easily conform to and the robustness of the electro-optic modulation techniques (which has been covered in the previous section on liquid crystals). In addition, the field of all-optical computing or all-optical image processing is still in its infancy and so at some stage a conversion to the more traditional electronic technology is almost inevitable.

These types of device tend to function by using some field dependant technique whereby the amplitude and/or phase are affected by the application of such a field over a pixel area. The optically addressed SLM must not be discounted, though - the earliest devices were of this class and the potential resolution of such devices is far larger than that for electrically addressed devices as it is ultimately limited only by the wavelength of the light and also, by definition, they can be used as incoherent to coherent light converters.

Examples

The *Hughes Liquid Crystal Light Valve (LCLV)* (Grinberg et al (1975)) illustrated in Fig 2.4 is an optically addressed SLM and one of the classic designs of reconfigurable devices. As shown in the figure, it consists of layers of photoconductor, dielectric mirror and uses a layer of liquid crystal as the active modulating medium. The voltage across the liquid crystal is varied by the intensity of the write beam which varies the impedance of the photoconductor layer. By choosing the liquid crystal and the thresholding conditions, it is possible to create an analogue modulating device or, with a sharper threshold, a binary modulating device.

The device suffers somewhat from charge leakage resulting in a more complicated pixel function (as defined in chapter 1) as the charge over the pixel area is smeared. More recently, photodiodes have been used in an array on the write side of the device to prevent exactly this problem by gathering the generated charge. Devices have exhibited a space-bandwidth product of around 2000 (Latham and Owen (1986)).

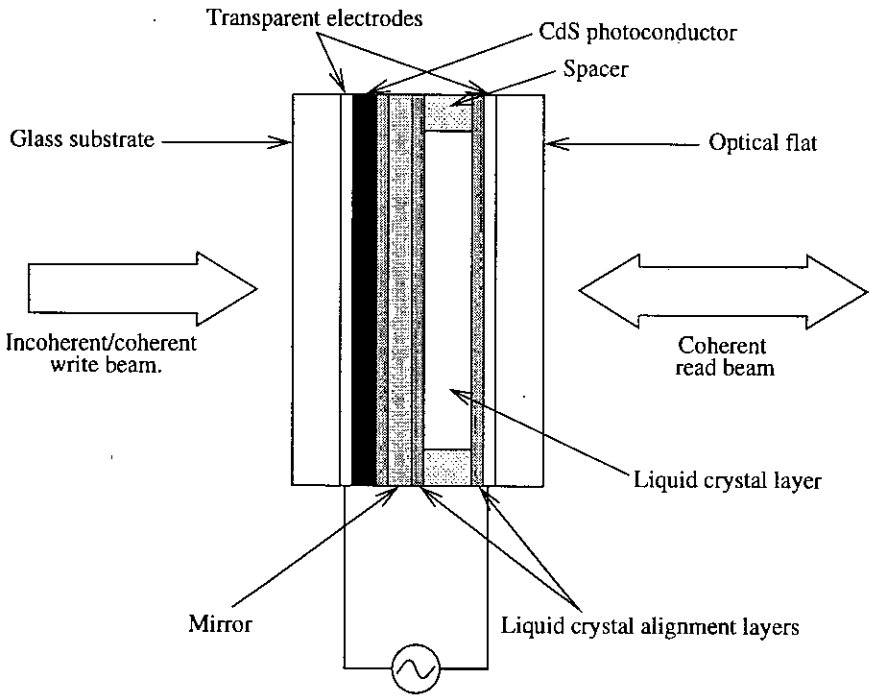


Fig 2.4. The Hughes Liquid Crystal light valve.

The *Magneto-optic SLM* (Ross et al (1985)) consists of an array of pixels which are electrically addressed. The application of a voltage across the pixel (which is made of a bismuth garnet film) causes the device to exhibit the Faraday effect, modulating the linear polarisation of transmitted light. By suitably orienting polarisers around the device, it may be used in binary amplitude or binary phase modulating modes.

The main problems involved have been due to the high drive currents and poor transmission of such devices although these have been addressed to some extent by the introduction of MOSLMs with far lower drive currents than previously required. A major advantage of this device is its speed of operation - 1988 figures (Flannery, Loomis and Milkovitch (1988)) quote 1100 frames/second when changing half of the elements in each frame.

A *deformable mirror SLM* (Hornbeck (1988)) is illustrated in Figure 2.5 and consists of a hinged mirror whose orientation may be altered by the application of a suitable voltage. In the form illustrated, incoming light is deflected from the normal when a pixel is switched "on". It would be possible, also to modulate the phase of the light emitted by altering the hinge arrangement allowing the mirror to be depressed in the on state rather than hinged. Reports of performance on this device suggest switching speeds of about 10 μs in devices 128 by 128 pixels square. Texas Instruments have recently developed a 768 x 576 pixel device for use as a projector display (Wilson (1993)).

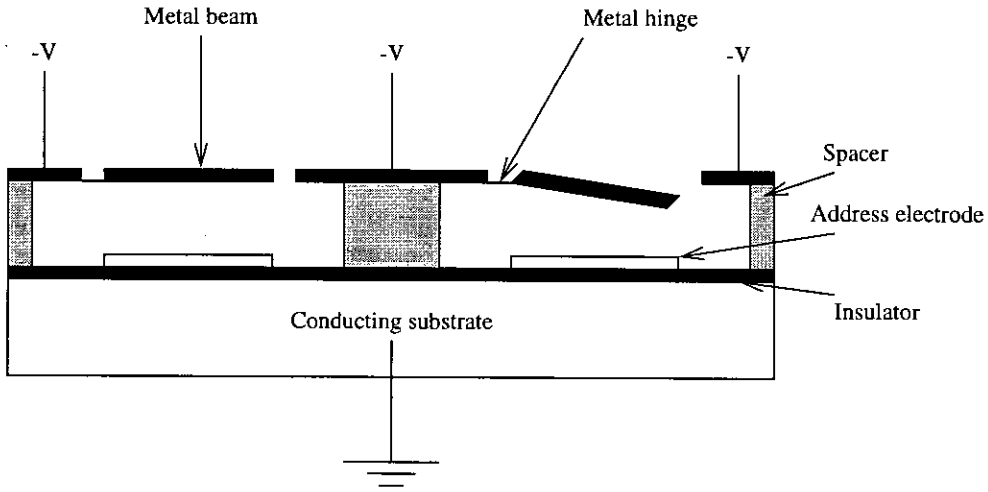


Fig 2.5. The Deformable mirror spatial light modulator (after Hornbeck).

Liquid Crystal SLMs

There are two categories of Liquid crystal SLMs- those matrix addressed devices which are spin-offs from the liquid crystal display industry and the Liquid crystal over silicon devices which are the speciality of this research group and of other groups (for instance at Boulder, Colorado). These are discussed separately below.

Liquid Crystal Displays

These are, in general, transmissive devices consisting of a layer of liquid crystal sandwiched between two glass layers upon which are embossed matrices of conducting electrodes. A more detailed description of a simple addressing scheme is presented in chapter 5. The characteristics of this type of device are low contrast and frame rates, with poor optical flatness and contrast uniformity over the device. However, they may be made reliably well and utilise the simplest of liquid crystal interactions for the modulation effects. It is also possible to correct for imperfections in the display through additional optics or specialised geometries (Casasent and Xia (1986), Kim et al (1988)). This form of device has been given considerable attention (e.g. Liu and Chao (1989), Kirsch et al (1992)) due to its low cost and ready availability. Miller (1993) used one such device successfully as the filter in a correlation architecture

Liquid Crystal over Silicon SLMs.

One of the first devices of this type was pioneered by Ian Underwood of this research group in his Ph.D (Underwood (1987)). It consists of an array of 16 by 16 electronically addressable pixels on a silicon backplane and the modulation was achieved by exploiting the birefringence of the liquid crystal and its response to applied fields.

This form of device is made up from a thin layer of liquid crystal constrained by two conducting surfaces one of these being an array of mirrors upon which an electrical field may be applied. Due to the mechanics of the design, the device is necessarily pixellated and a schematic diagram of this device is illustrated

in Fig 2.6. A voltage signal is applied to the substrate and a (different) signal applied to the conducting surface of the cover glass. The resulting potential difference across the liquid crystal cell (denoted V in the figure) then controls the modulation of the pixel. The basic physics of the device may be extended to cover the remaining devices in this genre, allowing for some differences in the liquid crystal medium and modulation technique used (e.g ferroelectrics are bistable and require a different drive scheme from nematics).

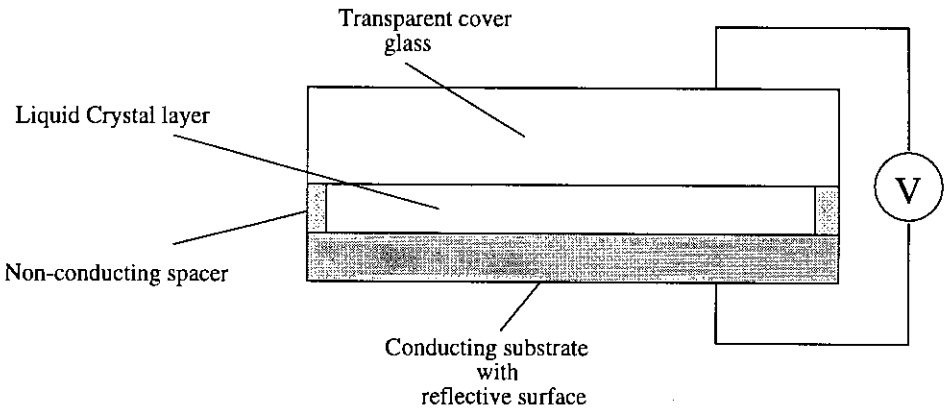


Fig 2.6. Schematic diagram of electronically addressed Liquid crystal over silicon spatial light modulator.

Devices in this family produced within this research group range from the original 16x16 pixels (Underwood (1987)) through the 50x50 and 128x128 pixel designs (McKnight (1989) - examined in detail later) up to more recent 176x176 (Underwood et al (1991)) and 256x256 (Burns (1994)) pixel devices. The liquid crystal medium utilised also changed throughout this evolution from simple nematic or guest-host liquid crystal to the faster switching ferroelectric liquid crystal where much current work is focusing (Underwood et al (1991)).

A technology attracting a great deal of research (e.g. Miller (1990)) is the multiple quantum well (MQW) self electrooptic effect device (SEED). This exploits

the quantum confined Stark effect - a non-linear effect describing optical behavioural changed of MQW structures on application of a perpendicular electrical field. As it is a quantum effect, response times are only limited by the applied electric field and typical frame rates are extremely high although reported contrast ratios are disappointing (~ 2:1).

Specific application devices

In the course of research many devices have been devised for general purpose use. In addition, though, specialty devices have been presented and these are attracting much interest.

The *randomly pixellated* transmissive Spatial light modulator (Heddle (1992)) was designed by Stephen Heddle of the Applied Optics Group of Edinburgh University. It is a simple device incorporating a multiplexed electronic drive scheme powering transparent electrodes evaporated onto the surface of the glass cell.

Heddle examined the performance in correlation geometries of the regularly pixellated device in comparison with a novel 'randomly pixellated' transmissive SLM. Fig 2.7 (a) shows a regularly pixellated design with a fill factor of 25%. Fig 2.7(b) shows the form of the novel design where each pixel takes up (randomly) one of four positions in each pixel area. The theory suggested that it would be possible to reduce, or remove, the second (or higher) order replications through choosing a suitable random positioning scheme to decide where to place the pixels on the device (Heddle and Sillitto (1993)). Such a device was shown to perform well in correlation architectures, removing the higher spectral orders of the Fourier transform and reducing the effects of false correlations.

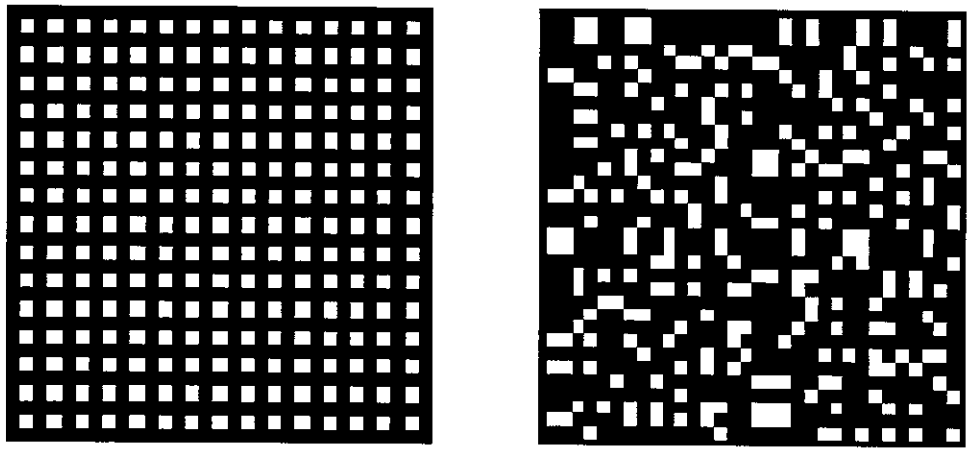


Fig 2.7 (a) regularly pixellated filter; (b) 'randomly' pixellated filter.

Related papers to this are the work of Miller (1993) and Kozaitis and Foor (1992) who both attempt to optimise the performance of reduced-resolution filters and hence maximise the potention for limited space-bandwidth product SLM and LCTV devices.

Other novel areas of interest include the *smart SLMs* (see the review paper by Johnson et al (1993)) where the device may contain photodetectors, amplifiers and memory elements. These devices may perform image processing functions on chip (image thresholding or edge detection, for example) or more abstract functions (producing a series of equal intensity pulses the number of which depends on the input intensity).

Performance Issues for SLMs

There are two approaches to the improvement of SLM design - these have either concentrated on new designs with increased bandwidth (larger pixel numbers, higher resolution and faster frame rates which have been partially covered in the preceding sections on SLM types) or the optical quality (review paper Underwood et

al (1994)). O'Hara et al (1993) concentrated on the polishing or "planarisation" of the silicon substrate of liquid crystal over silicon SLMs and has achieved some remarkable results, this particular paper reporting a reduction of the surface profile variations from around 16 000 Å to less than 200 Å. An additional benefit is the reduction in imperfections in the actual aluminium mirror surface evaporated onto the silicon. This is illustrated in Figures 2.8(a)-(d) below. Figure 2.8(a) illustrates the typical profile of a silicon chip surface upon which a thick layer of silicon dioxide has been deposited. This is polished flat (Figure 2.8(b)) before contact holes are etched through to the chip mirror surfaces (Figure 2.8(c)). Finally, a conducting (reflective) metal layer is deposited on the surface forming the final pixellated mirror surface with a resulting high reflectivity (Figure 2.8(d)).

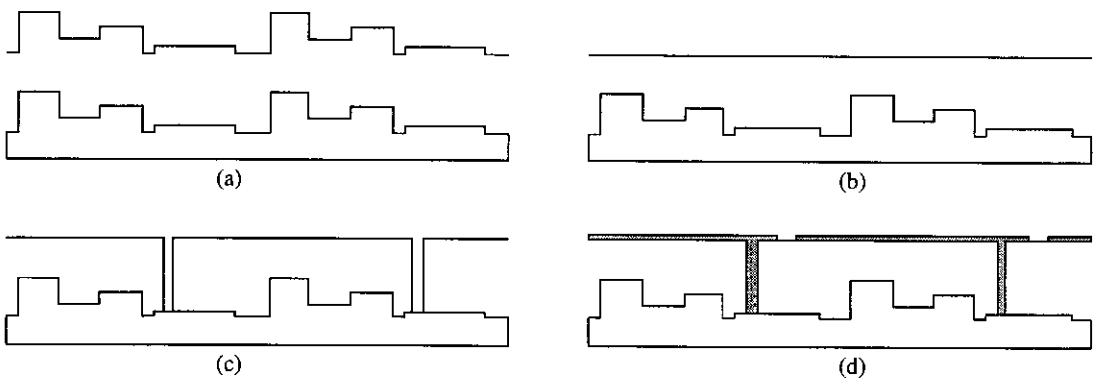


Fig 2.8 Cross sections illustrating planarisation process (After O'Hara et al (1993));
(a) SiO_2 deposition; (b) Planarisation ; (c) contact holes etched; (d) metal contact and mirrors deposited.

Applications

In almost all fields of applied optical research SLM's have made their presence felt. They have been used as alternatives to transparencies, where the added flexibility of being able to program the display is invaluable especially when investigating complex filtering operations where the generation of successive transparencies would prolong such experiments. Now that the science of SLM

design is relatively mature, SLM devices are now designed around the requirements of optical researchers providing phase-only, intensity only and amplitude-and-phase display devices.

In short, there is a huge variety of spatial light modulating devices using a variety of modulating techniques and media. Research has now developed from the prototype devices of the 1980's to application specific devices being created now as interest (both from research and commercially) has been aroused. Current devices include the ferroelectric 256x256 developed within this research group (Burns et al (1994) which exhibits a 2khz frame rate and a 8:1 contrast ratio. As a comparison, MQW-SEED devices have been reported with frame rates of 2 GHz (Lentine et al (1994)) but with a poor contrast ratio of 2:1 for a 6x6 array, and the 50x50 nematic device designed by McKnight (McKnight (1989) had a frame rate limited by the nematic LC technology to around 0.1kHz. It is conceivable, therefore to talk of an ideal LC SLM which may be fabricated with current technologies with an array size of better than 1000x1000 and frame rates of hundreds of kHz.

Chapter 3

Practical SLM Work

Introduction

The following is an account of the practical work undertaken towards the goal of using a working SLM in an optical correlation system, with the ultimate aim of performing an optical simulated annealling process (Note - some elements of the system are necessarily electronic as optical implementation of the necessary feedback system is potentially very difficult and exceeded current equipment capabilities).

The workings of the 50 by 50 liquid crystal over silicon device are described together with drive systems and the computer control system. Some time is spent describing the techniques and problems involved with the construction of such devices and the efforts made to improve quality and reliability. The extension to the 128 by 128 device is also described although as the design was predominantly based on the 50 by 50 design, the fundamental principles are essentially the same. The experimental work carried out on the 128 by 128 device is described together with the reasons for not continuing with study on that device causing a fall back to studying the 50 by 50. This is followed by a description of the SLM manufacturing

process together with the efforts made to improve reliability and quality of the resultant devices.

A short note is required to explain some of the terms used in the following sections. The computer generated (and tested) design is used to create wafers consisting of neighbouring dies (a die being one complete SLM circuit). These are, in turn, sawn apart to provide separate dies and these are glued onto chip mounts with all relevant electrical connections provided to allow the drive circuitry to communicate adequately with the SLM circuit. This does not become a working spatial light modulator until a liquid crystal layer and cover glass are added.

The 50x50 Spatial Light Modulator

The 50 by 50 SLM consists of an array of total dimension 3.7mm by 3.7mm (giving a pixel pitch of 74μm). Each mirror has dimensions of approximately 43 μm by 42 μm giving a fill factor of 32.9%. The estimated frame rate for complete changes of displayed patterns (every pixel being changed) is given as 1.8 kHz but this is over optimistic as the nematic liquid crystal used on the device exhibits an expected switching time of between 10 and 13 ms (75 - 100 Hz) and it is the response time of the modulating medium which is the most limiting factor. For the purposes of this thesis, however, the performance of the SLM was not an issue as long as a device with sufficient contrast could be obtained for use as a Fourier plane filter. Indeed, for the purposes of simulated annealling, where just a single pixel is required to be switched each frame, this is perfectly adequate for prototype experiments.

The pixel drive circuitry

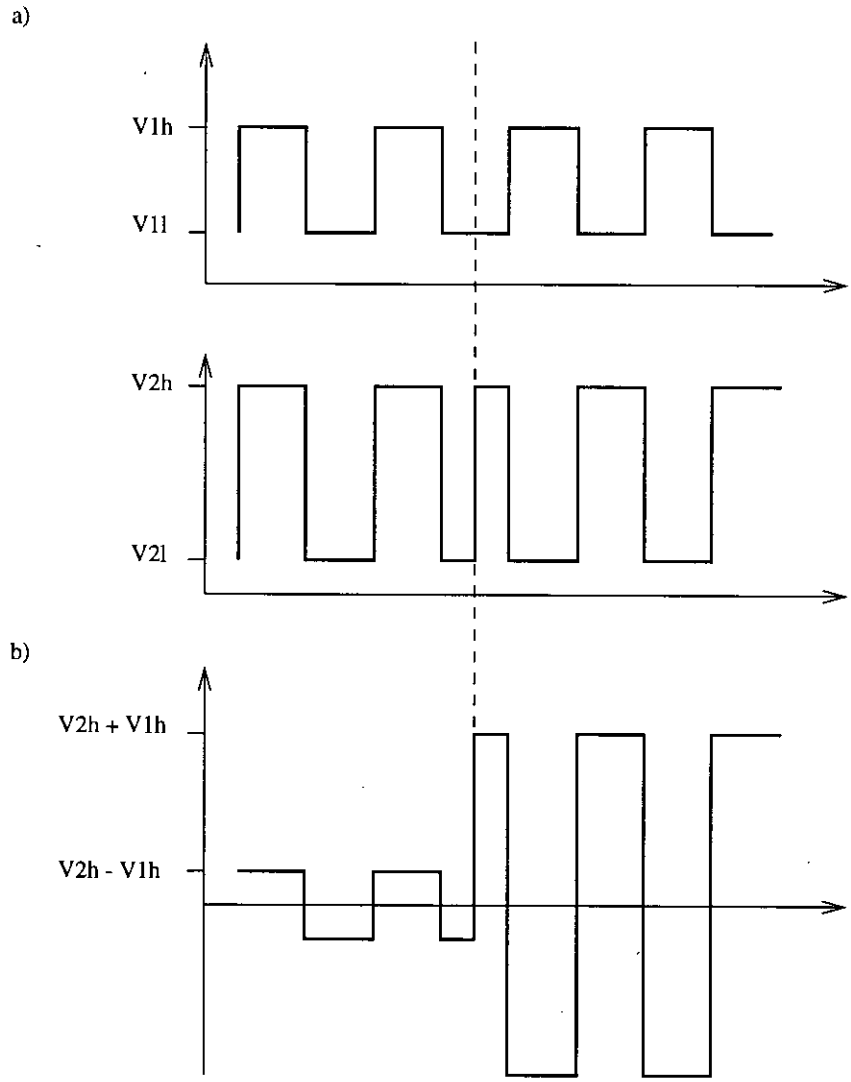


Figure 3.1. Pixel drive voltages. (a) magnitudes of drive voltages. (b) potential difference across LC layer

The drive circuitry of this device is very simple in principle although the implementation in silicon backplane design adds many more complexities to the problem (which are discussed in McKnight (1989)). A schematic of a typical liquid crystal over silicon SLM is illustrated in figure 2.6, chapter 2 and more information may be obtained from McKnight (1989) or Underwood (1987).

Essentially, the transparent glass cover electrode is driven by a constant-phase square wave pulse train (Fig 3.1(a)) and the pixel mirror driven by a square wave pulse train of the same frequency and magnitude. Changing the state of a single pixel involves changing the relative phases of the electrode pulse train and the mirror pulse train (Fig 3.1(b). In this example, prior to the switch (marked S in the diagram) the two trains are in phase, resulting in a net zero voltage across the sandwiched liquid crystal. At the position marked by the vertical dotted line, the mirror voltage becomes out of phase with the electrode voltage and there is a resultant square wave AC voltage of magnitude 2V over the liquid crystal. In conjunction with some of the liquid crystal effects described in chapter 2, this allows the device to modulate light.

Implemented in silicon, this design consists of a data register, which presents data to the chip *rows* and an enable register which controls the signals sent to the chip *columns*. Data is loaded into the data and enable registers simultaneously under the control of two non-overlapping electronic clock signals[1] Oncethe registers are filled with data electronic pulses are sent to the device causing the data to be sent to the pixels. The chip design was such that a relatively simple drive circuitry was all that was required, consisting of a number of integrated chips, power supply and a clock signal generator.

A computer program was written in a combination of BBC BASIC and BBC Assembler language to allow an Archimedes 3000 microcomputer to drive

---

[1] The clock signals may have any period but must never be high simultaneously, and must not be low for periods of greater than 1 millisecond. These are stipulations in McKnights design (McKnight (1989)).

the 128 by 128 and 50 by 50 SLMs using a graphical user interface and the microcomputer serial port to control the chip drive circuitry. The signals from the serial port were generated within the software according to the pattern developed on screen (including some optimisation to reduce the addressing times) and then sent directly to the drive circuitry. This program allowed arbitrary pattern design and a simple single key action which sent the data to the SLM.

The 128x128 Spatial light modulator

The 128x128 electronically addressed VLSI spatial light modulator was based on the 50 by 50 design as described briefly above. The essential differences between the designs are listed below:

i. The array is approximately 9 mm square with a pixel pitch of 70 μm.
ii. The pixels are 47 μm square giving a fill factor of 45 %. It should be noted, however, that this fill factor was achieved by covering the silicon circuitry wherever possible (without shorting any circuitry) with aluminium to create a mirror. The mirror area is not by any means flat and this must be considered in terms of complicating the pixel function.
iii. The 128-bit data and enable registers consist of 4x32-bit registers which could be addressed in parallel or serially, allowing the potential for improved addressing rates

It was proposed to use it in amplitude-modulating hybrid field effect mode and later to apply it to optical systems where its large space-bandwidth product

(=128, the number of pixels in a given dimension) suited it more than previous experiments in the group which were based on the 16x16 modulator developed by Ian Underwood or even the 50 by 50. The initial stages in the experimental portion of this thesis were to test the SLM designs electronically and then optically before finally placing them into an optical processing setup.

The state of affairs with regards to the 128x128 at the start of this work was that the design stage had been completed and a quantity of the raw chips had been produced in the Microfabrication Facility in the Department of Electrical Engineering at the University of Edinburgh. These had not as yet been electrically tested and the primary objective was therefore to ensure that sufficient dies were functioning adequately enough to support this Ph.D. and thesis.

In order to achieve this a probe card, consisting of electrical connectors ending in microscopic points which are lowered into direct contact with the electrical inputs of individual dies was created and used to check each die in turn. A typical wafer as fabricated is illustrated in Figure 3.2 with 28 complete dies per wafer and 10 of these wafers were made in total. The probe card was in turn connected to a switch box which allowed manual clocking of data through the device according to the basic drive circuitry (consisting of the non-overlapping clock signals, data, enable and read control signals) described in the previous section and it was this arrangement which was used both to (electronically) test the 128x128 wafers and the bonded 50x50 chips.

Once the probe card was in contact with the wafer, a data (or enable) signal was set and manually clocked according to the requirements of the design. An oscilloscope and logic analyser were used in turn to examine the output from the

data (or enable) register being tested, allowing an instant impression of whether the register was functioning as expected or not. This process was repeated many times using different voltage levels.

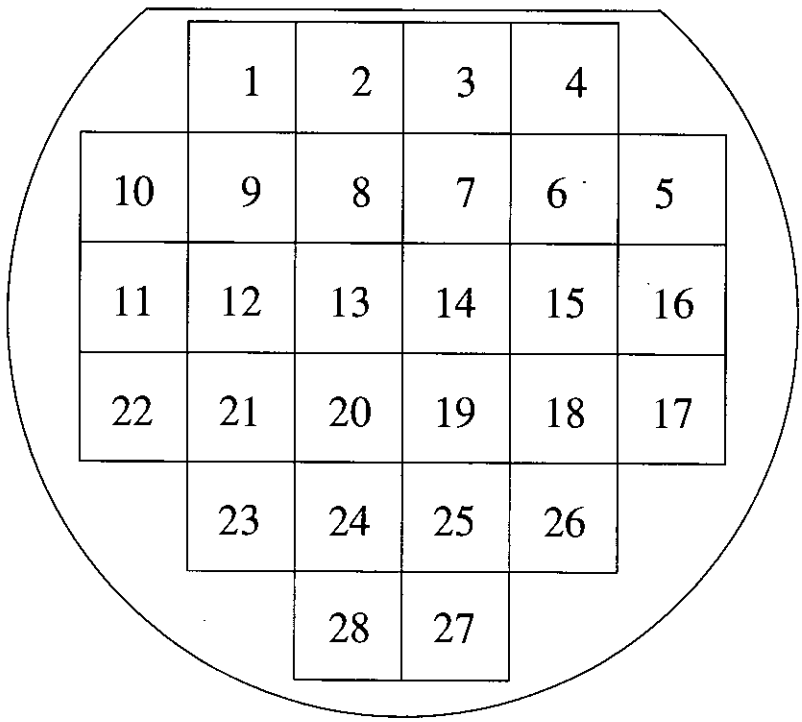


Figure 3.2. Geometry of complete wafer after microfabrication of electronic circuits.

128 by 128 SLM Test Results

All of the 128x128 modulator wafers were examined under a high resolution microscope, and those dies which were still undamaged by this stage were electronically tested as described above. Typical results for an individual wafer are noted in Table 3.1 - wafer two is taken as representative. Only the main features of the SLM design were tested at this stage - specifically these were the data and enable registers (as mentioned previously, the 128 bit data and enable registers were split into 4x32 bit registers in order to allow improved frame addressing rates). If either or both of these 128 bit registers did not function correctly then the device (or at least a

portion of the device) would be useless. The minimum acceptable situation would be for a single 32 bit section of the data register and a single 32 bit section of the enable register to function in which case one sixteenth of the device would be usable.

Wafer Two - test results summarised	
Dies 1-4	Very high current (1A) - no output
Die 5	Fair current (238 mA) Data register 32bits/128 working Enable not functioning
Dies 6-7	Very high current - no output
Die 8	Data register 32bits/128 working Critical to supply and substrate voltages
Die 9	Data register 32bits/128 working Enable not functioning
Die 10	Data register 128bits/128 working Sensitive to substrate voltage Enable not functioning
Die 11	Data register 32bits/128 working Enable not functioning
Dies 12-14	Damaged
Die 15	Data register 128bits/128 working Enable not functioning
Dies 16-19	No response
Die 20	Low current, no response
Dies 21-26	No response
Die 27	Current 600 mA, no response
Die 28	No response

Table 3.1. Electronic test results for wafer two (128x128 chip design).

In addition, the magnitude of the current drawn by the device was taken as an indication of whether the device would potentially work: currents of over 0.75 A were considered to be far too high (taking the design and computer simulations of the SLM into account) and suggested some bleed to ground or short circuits in the die. Zero currents also suggested that the die was not functioning normally.

As can be seen from the table, the design met with little success. No dies functioned in entirety and most did not function at all. In the cases in which either the data or enable registers partially worked, the die was extremely sensitive to the voltage level at which the silicon substrate was held and often results obtained were not reproducible. At the time, this was taken to be because of chips being destroyed somehow during the test procedure - an assumption later discounted.

After months of probing, with very little success, Burns having just joined the group from the department of Electrical Engineering applied his VLSI knowledge and located a bug in the design of the 128 by 128 chip (Burns, 1991) (the ground line was incomplete and therefore floated free). As the devices clocking and data shift mechanisms were highly dependant on the relative applied voltage levels, this effectively rendered the design useless. This particular bug explained extremely well the sensitivity to the applied and substrate voltages of the chips which partially and sporadically functioned, and also the irregular or irreproducible twitches in one or other of the major shift registers of other dies.

In conclusion, it became impossible to continue with this device as it was, and although it would require only a minor correction to the design to complete the faulty ground line another wafer run through the Microfabrication Facility was out of the question. Attention shifted, therefore, technologically one stage backwards to the 50 by 50 spatial light modulator which had worked at least to some extent during the course of McKnight's thesis. This would still give significant improvements over the 16 by 16 device designed by Underwood (1987) and studied by Potter (1992) and Ranshaw (1988), allowing for more resolution in the input objects, due to the improved space-bandwidth product.

Experimental work on the 50x50

The status of this device was of proven capability. McKnight had used it in simple optical processing benches and had fully tested the design electrically although practically, it had exhibited a very poor contrast ration of about 10:1 However, in the process, no working devices survived and it was necessary to refabricate the SLMs from the remaining wafers and the remaining mounted dies. These were classified as the 'probable' and 'possible' working devices, according to how well they had passed the elecrical tests (using a probe card and manual electrical inputs as described above). The 'definites' had all been used to destruction during the course of McKnight's thesis.

The stages for the SLM assembling process are listed and illustrated in the following section.

SLM assembly

Figure 3.3 illustrates schematically the process through which a bare chip must go to manufacture an electrically addressed spatial light modulator. The weak points are emphasised, and the attempts made to improve the procedures are described. The main components of the final device are the overglass, the spacers, the liquid crystal and the electronic circuitry (with its external drive circuitry also).

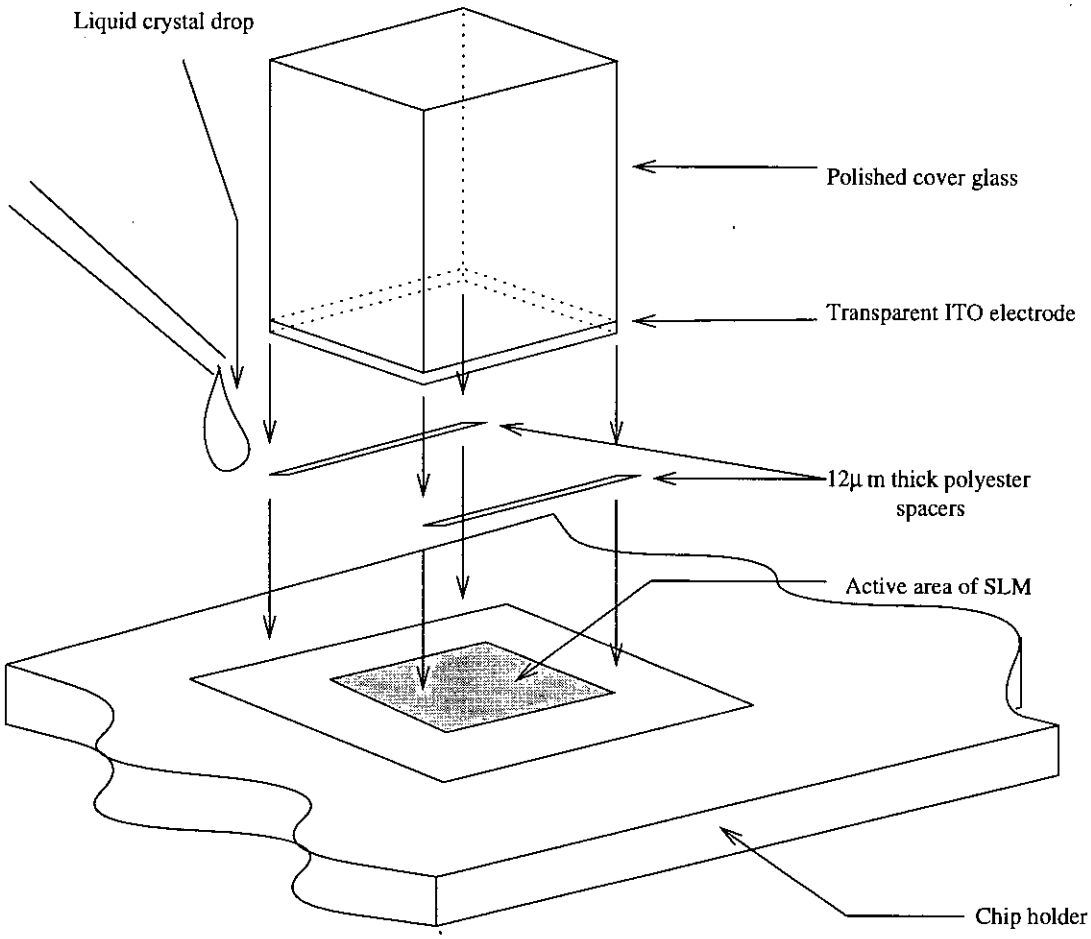


Figure 3.3. SLM Assembly

The manufacturing process involves coating the base of the coverglass with a transparent conducting medium to act as an electrode. In the chips produced in this group this has always been evaporated indium-tin-oxide (ITO) due to its good conductivity and transmissivity. On top of this, a thin alignment layer (of silicon oxide SiO_x is applied (where $1 \leq x \leq 2$) by obliquely evaporating it onto the ITO coated substrate. This forms a characteristic highly directional ridge pattern and by chosing the angle of evaporation the ridge frequency may be tailored in order to induce the nematic liquid crystal molecules to align along the ridges. A similar layer is evaporated onto the surface of the chip to tie down the liquid crystal at the opposing surface also.

A thin layer of aluminium is applied to one surface of the coverglass in order to obtain an electrical contact with the transparent ITO electrode. This extends some way up the back of the coverglass to allow an adequate electrical contact to be made with the drive circuitry. Thin (12μm) polyester spacers are cut from suitable sheeting and placed on the coverglass surface before lowering the glass onto the chip thereby leaving a thin empty space into which the liquid crystal can flow and the coverglass is secured to the chip using UV quick curing glue.

The chip is then heated to around 140°C before allowing a drop of liquid crystal to fall on the edge of the coverglass. Aided by the relatively high temperature, the liquid crystal flows freely into the space available by capillary action until the cell is completely filled. Once cool, Araldite Rapid™ is applied around the cell to counter the potential for the cell to leak and render itself useless.

There were several obvious weaknesses in this construction process, the main one being that no guard was taken against pollution of one or both conducting surfaces by grease or dust. As a single 5μm dust mote covers one and a half SLM pixels this must be taken into consideration. In addition such objects affect the overall thickness of the modulating liquid crystal layer resulting in variable and poor light modulation. A group of researchers in the research group, myself included, fabricated a clean room facility within which all such fabrication would henceforth take place. In addition, current working conditions in industrial establishments and methods of fabrication were studied before arriving at the final improved fabrication method. This was essentially the same as has been outlined but now with a rigorous solvent cleaning process for both cover glass and silicon chip. The materials used for deposition (SiO and ITO) were kept in a clean environment as were the liquid crystal

samples. Finally, Heddle and the author investigated the use of vacuum environments for liquid crystal filling of SLM cells - the method eventually used being filling the cell within a high vacuum chamber using an arm which could be manipulated externally. The results and reliability of test cells increased minimising damage effects from the construction process.

Conclusions

While all reasonable efforts were made to maximise the likelihood of obtaining a working SLM, these efforts did not result in success. The blame for this is twofold. Primarily the devices studied were the last ones remaining from the original stock and were by definition the least likely to produce good quality working SLMs. Secondarily, when close to the end of the supply of devices, a bug was discovered in the electronic drive circuitry which had resulted in a gradual destruction of even the devices which worked. Unfortunately at around this time, base funding for the electrical microfabrication facility at the department of Electrical Engineering was withdrawn by SERC which prevented further runs to correct the design errors. It became necessary to cut losses and change the onus of this thesis from the practical systems side to a more theoretical objective.

Chapter 4

Optical Correlation

Introduction to Optical Correlation Techniques

The potential of light-speed Fourier transforms has largely been exploited in the field of correlation. Computer techniques, even with the advances which have been made in parallel processing over the past years still place a heavy reliance on electronics and the inherent limitations of addressing times and linear mathematics. Optical techniques execute a Fourier transform effectively instantaneously through the phase effects introduced by a suitably designed and positioned lens (see for example Goodman (1968) or Appendix 1).

In Appendix 1b we see that the combination of the Fourier transforms of two input objects, when recombined by a Fourier-transforming lens, produces an energy distribution in the image plane which is the cross-correlation of the two objects. This obviously implies that speed--of--light pattern recognition is possible, however, once one is committed (due to technological limitations in an infant science) to using a pixellated, and possibly binary-only SLM with limited dynamic capabilities, there has to be a trade-off between the ideal and what is physically and technically feasible.

This chapter aims to cover a range of correlation techniques and theories, emphasising the concessions which are made for each and illustrating with examples and simulations.

Classical matched correlation

The original optical correlator (that proposed and demonstrated by Vander Lugt (1964)) was the classical matched correlator. The discovery of holographic (complex image recording) techniques had allowed phase and amplitude information to be recorded on transparencies and the first matched filter correlation techniques were proposed directly as a result. This technique involves the insertion of a complex filter in the Fourier plane of the device which is designed and chosen to cancel out the phase information of the Fourier transform of the target scene and, ideally, also flattening the amplitude at the Fourier plane.

More recently the advent of SLMs allowing phase and/or amplitude modulation has accelerated research in these areas (see for example Horner and Bartelt (1985), Psaltis et al (1984)). Due to difficulties in controlling both the phase, ϕ and the amplitude, A in complex filters, work has generally targeted on *single parameter* modulatory techniques whereby a binary SLM is used which modulates EITHER the phase OR the amplitude. This subject is covered in later sections of this chapter under *Binary Phase-Only Correlation* and *Binary Amplitude-Only Correlation.*

Figure 1.1 (Chapter One) shows a diagram of a typical 4-f image processing bench. The arrangement is exactly the same as for a classical matched filter correlator bench. The object is presented at the plane marked "input plane" in Fig

1.1 and the representation of the Fourier transform of the target is presented as a filter at the Fourier plane of lens L2.

Consider a target image, t(x,y) which Fourier transforms to produce an optical field of $T(f_x, f_y)$. If a filter may be produced with a transmission function of $T^*(f_x, f_y)$, where * denotes complex conjugation, then placing this filter in the Fourier plane will cause the output of the processor to be the (coordinate reversed) convolution of the input with the Fourier transform of the filter as derived in Chapter 1. Using the convolution theorem, this output may be denoted:

$$\begin{aligned} X(x,y) &= \mathcal{F}\left(T(f_x, f_y).T^*(f_x, f_y)\right) \\ &= t(-x,-y) \otimes t^*(x,y) \\ &= t(x,y) \otimes t^*(-x,-y) \end{aligned}$$

X(x,y) is the auto-correlation of t(x,y), and may be denoted by $X_{tt}(x,y)$. For a different input function, f(x,y), the output of the same processor will be $X_{ft}(x,y)$, the cross-correlation of f(x,y) with t(x,y).

The output of the correlator (marked as "output plane" in Fig 1.1) can then be seen to be the cross-correlation of the object and target.

Classical joint Fourier transform correlation

Figure 4.1 shows a joint Fourier transform correlator geometry. This type of geometry has been well documented (e.g. Weaver and Goodman (1966)) over the past years. The input and the target scenes are both presented at the input plane and simultaneously Fourier transformed by the lens L1. The interference intensities of the Fourier transform are recorded by some non-linear (most usually square law)

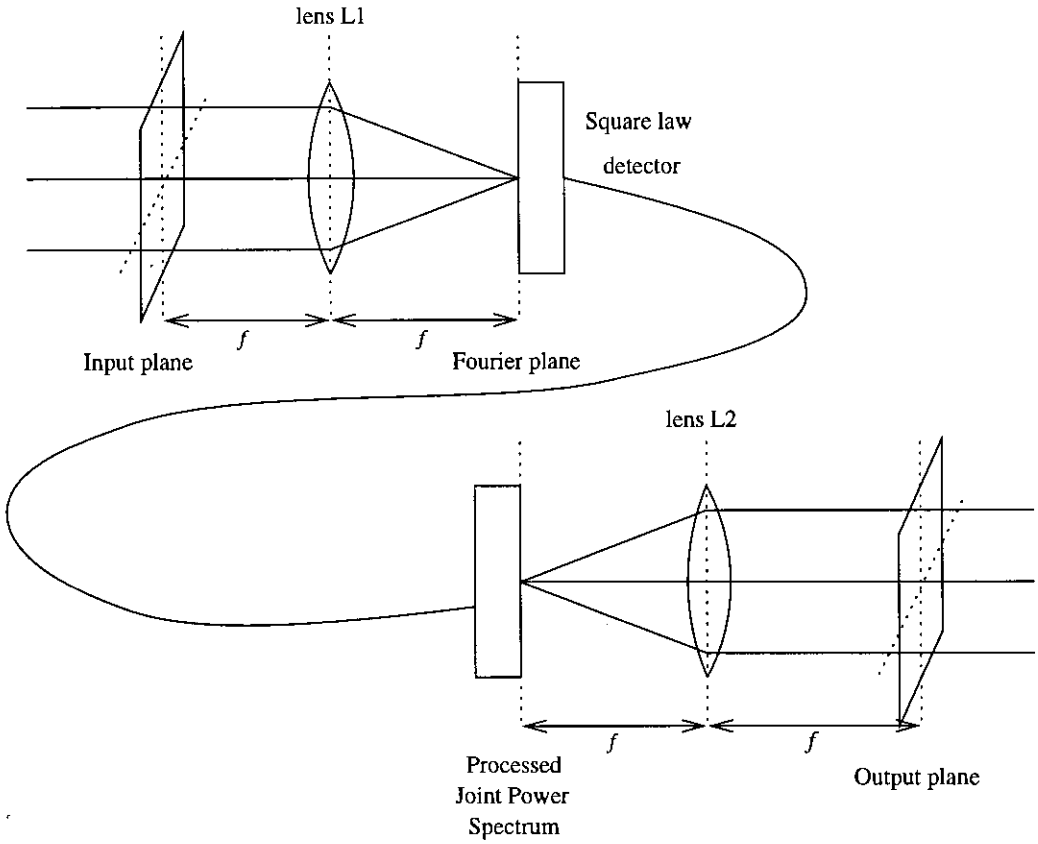
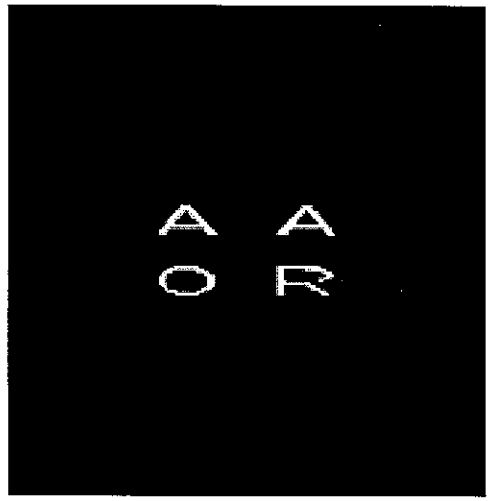


Fig 4.1 Classical joint Fourier transform correlator (JFT)

detector e.g. a CCD array or an optically addressed spatial light modulator. The resulting intensity field (known as the *Joint Power Spectrum*, or *JPS*) is displayed as the input to a second Fourier transforming lens, L2 which produces at the output plane a dominant on-axis peak and lower intensity side-lobes corresponding to the cross-correlations between the targets presented and the reference image.

The two main advantages of this type of correlator are that the two images to be cross-correlated may be presented in the same plane and secondly, that the Fourier plane representation of the Joint Power Spectrum is entirely real, so that a Fourier plane SLM need only be a single parameter modulator (a device which modulates only one of the possible light characteristics - amplitude, phase or polarisation).

a)



b)

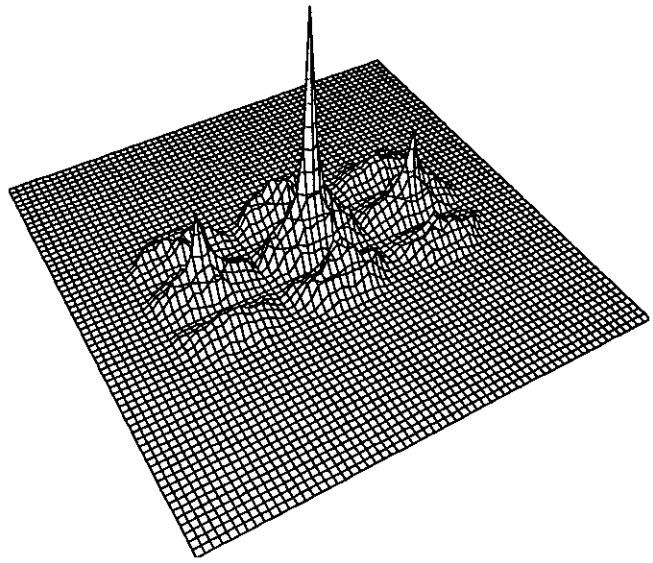


Fig 4.2. Simulation of output of Joint Transform Correlator.(a) object and (b) output.

A simulation of such a correlator was run on a UNIX workstation, using a simulated 64x64 pixel SLM, and with the object plane as shown in figure 4.2(a). Figure 4.2(b) shows the simulated correlator output when the objects of Fig 4.2(a) are presented in the input plane.

Disadvantages of this architecture are obvious: The central peak in the output is NOT the cross-correlation of the input images, but the auto-correlation of the

COMBINED image. The cross-correlations appear off-axis, arranged diagonally in pairs. It indicates clearly how the output plane may easily become cluttered with unwanted cross-correlations - in this case, a reference object presented with only two candidates produces six cross-correlations together with a strong central peak.

Javidi and Kuo (1988) proposed binarising the JPS in order to allow binary SLMs to be used in the Fourier plane and showed the output to have higher peak intensity and better cross-correlation discrimination when using such filters.

Phase-only and Binary phase-only matched correlation

Kermisch (1970), published a paper which analysed the effect of image reconstruction after discarding the amplitude information and using ONLY phase information. His treatment was statistical and investigated the cases of both perfect phase matching and imperfect phase matching. His conclusions were that in the perfect phase matching case, 78% of the total image radiance reconstructs exactly the original image and that the phase information dominates the process.

Oppenheim and Lim (1989) repeated his investigation and showed conclusively the relative importance of phase over amplitude in image transmission and reconstruction, and, indeed found that an image could be reconstructed surprisingly well using the Fourier plane phase information and an average envelope of amplitude information taken over several (different) objects.

Horner, in 1982 had defined a measure of the optical efficiency of a correlation by the equation:

$$\eta_H = \eta_M \cdot \frac{\iint | f(x,y) \otimes g(x,y) |^2 \, dx.dy}{\iint | f(x,y) |^2 \, dx.dy}$$

where η_H denotes the Horner efficiency and η_M denotes the efficiency of the filter material , i.e. η_H is the ratio of the total energy in the correlation plane to the energy in the object plane (or a basic measure of the amount of energy arriving at the output plane). The integrals are taken over the respective planes. f(x,y) is any input function in real space, g(x,y) is any filter function, and $\otimes$ represents the action of complex correlation. He derived the result that, for a classical matched filter consisting of a 2-D rect function, $\eta_H \cong 44.4\%$ assuming the filter was recorded on a material capable of 100% diffraction efficiency and was used in a Vander Lugt correlator (Horner (1982)).

Horner and Gianino (1984) extended this mode of thinking to propose the use of phase-only filters to maximise the Horner efficiency as defined above. They reasoned that, as phase-only objects merely redirect light rather than absorbing it, the energy passing through to the correlation plane would be increased, giving a potentially sharper and higher correlation peak and a value for η_H of up to 100%. They demonstrated that phase-only filters showed great improvements in discrimination and correlation clarity over more conventional matched filters and, in addition, they suggested that phase-only filters would outperform the other main correlation methods in the presence of noise. This may inpart be explained by the flattening of the amplitude factors in the Fourier plane as the phase-only filter then effectively acts as a (relative) high pass filter (Horner and Leger (1985)).

In 1984, the use of binary phase-only filters for pattern-recognition in an optical system was proposed and demonstrated i.e. filters in which the amplitude information of the Fourier transform of the reference object was discarded (set to unity) and the phase binarised somehow (in this paper (Horner and Gianino (1984))

Fig 4.3 shows this graphically; pixels whose phase is below the threshold line (shaded area) are assigned a phase of 0 (i.e an amplitude of +1) and those above the threshold (unshaded area) are assigned a phase of π (i.e. an amplitude of -1). The thresholding line is arbitrarily chosen in this case.

There have been many studies on the optimal value of ϕ about which to binarise the filter. Psaltis et al (1984) , Horner and Leger (1985) and Cottrell et al (1987) postulated conflicting theories in this respect. These are summarised in Figure 4.4.

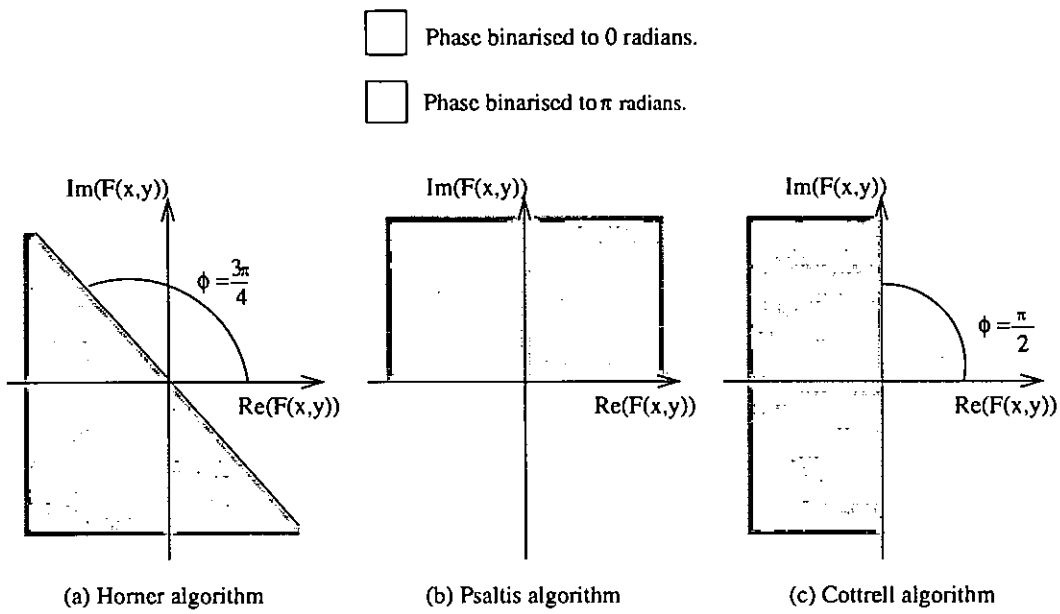


Fig 4.4. The three main phase binarising algorithms.

Horners filter (Fig 4.4(a)) represents a filter with a transmission function of :

$$H(\omega) = \begin{cases} +1, & \mathrm{Im}(F(\omega)) > 0 \\ -1, & \text{otherwise} \end{cases},$$

where $F(\omega)$ represents the Fourier transform of the object. This is obviously matched to the odd component of the object function, and a different approach was mooted by Psaltis et al (1984) (Fig 4.4(b)) whose postulated algorithm was:

the phase was set to π or to 0 giving possible amplitudes of -1 or +1 respectively. The remarkably good results displayed there opened the door for a generation of high discrimination and high performance filters using existing technology to provide the phase modulation required (e.g. Psaltis, Paek and Venkatesh (1984)).

Goodman and Silvestri (1970) provided an excellent study into the effects of quantising the phase in the Fourier plane. They took an arbitrary input function and transformed it. Once the Fourier spectrum was quantised it was re-transformed to indicate the effect produced by the quantisation by comparison with the original input image. It was noted that the resulting image was composed of an attenuated version of the original together with contributions from false images. The number and positions of these false images varied according to the number of phase quantisation levels - in the special case of phase binarisation, the output plane is complicated by a single false image, rotated by 180° with respect to the primary image, which may or may not overlap with the primary image depending on the position of the primary image in the input plane.

Now examining the case of binarisation to phases of 0 and π radians, the resultant transmission function has corresponding amplitudes 1 and -1 respectively (which constitutes a purely real function). Gaskill (1978) defines the symmetry property of Fourier transforms, stating that if an object $f(x, y)$ is purely real, then its Fourier transform, $F(f_x, f_y)$, will be Hermitian (the real component of F() will be even, and the imaginary part of F() will be odd). The presence of the false (or ghost) image therefore maintains the symmetry which is imposed by the act of binarisation. A binary phase-only filter with phases of 0 and π will be a perfect representation of the Fourier transform of any Hermitian function.

Kim and Guest (1990) conducted a theoretical investigation into the optimal value of phase in a binary phase-only filter after Farn and Goodman (1988) pointed out the erroneousness of the assumption that a modulation depth of π was optimal for a general filter. They concluded that under certain conditions, the optimal phase would not be π but could only be calculated by the use of a stochastic relaxation technique such as simulated annealing. It is interesting to note how few researchers have examined this aspect of the subject.

A general binary phase only filter pattern is calculated by binarising the phase information of the Fourier spectrum of the target according to some (variable) phase threshold. According to the reasoning above, the correlation will consist of two components: the cross-correlation with the primary target, and the cross-correlation with the ghost of the primary target. This is illustrated later in this chapter.

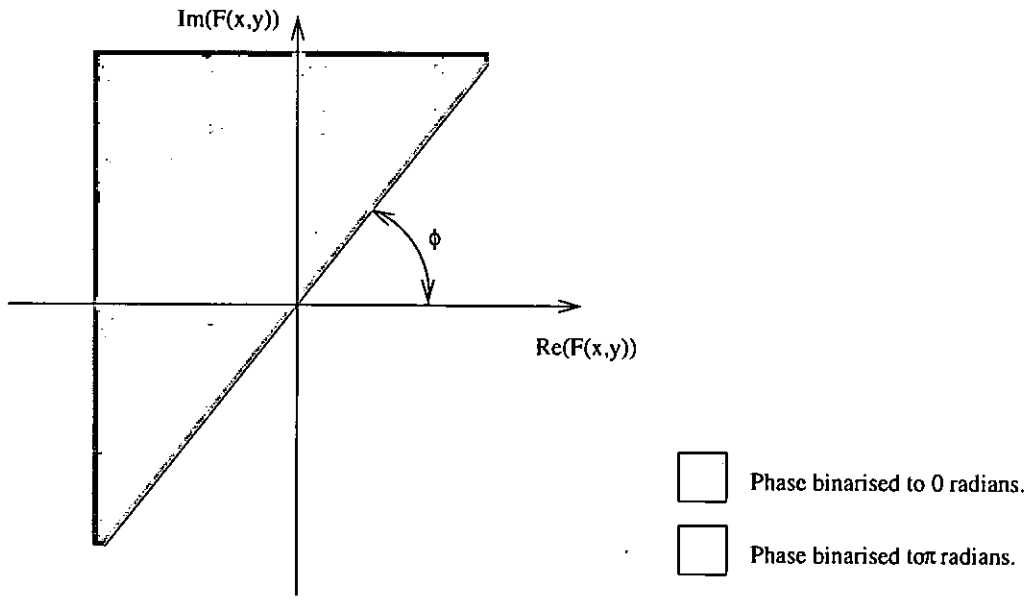


Fig 4.3. Phase binarisation through thresholding at an (arbitrary) line in phase space.

$$H(\omega) = \begin{cases} +1, & \mathrm{Re}(F(\omega)) > 0 \\ -1, & \text{otherwise} \end{cases}$$

which is matched to the even part of the function.

Cottrell et al (1987) (Fig 4.4(c)) based their algorithm on the Hartley transform viz:

$$H(\omega) = \begin{cases} +1, & \mathcal{H}(F(\omega)) > 0 \\ -1, & \text{otherwise} \end{cases},$$

where $\mathcal{H}(\omega)$ is the Hartley transform and is defined by:

$$\mathcal{H}(\omega) = F_E(\omega) + F_O(\omega),$$

with $F_E(\omega)$, and $F_O(\omega)$ representing the even and odd portions of the Fourier transform respectively. $\mathcal{H}(\omega)$ is therefore NOT a complex function. The Psaltis and Cottrel views of binarisation coincide when the object is purely even and the Horner and Cottrell views coincide if the object is purely odd. In terms of the phase angle ϕ as defined in Figure 4.3, Horner corresponds to $\phi=0$, Psaltis to $\phi=\frac{\pi}{2}$ and Cottrell to $\phi=\frac{\pi}{4}$.

Dickey, Stalker and Mason (1988) briefly discount the use of the Horner-and Leger (1985) algorithm due to the even part dominance of the energy for positive functions (i.e. optical irradiance images). Their proof is very short and is reproduced here.

If f(x) is a positive function ($f(x) \geq 0$) then the odd and even parts of the function are given as:

$$f_E(x) = \frac{1}{2}[f(x) + f(-x)]$$

$$\text{and } f_O(x) = \frac{1}{2}[f(x) - f(-x)]$$

The fraction of energy in the even component of the function can now be calculated by calculating the ratio of the integral of the modulus squared of $f_E(x)$ (above) to the integral of the modulus-squared of $f(x)$. This becomes:

$$\frac{\int_{-\infty}^{\infty} f_E^2(x)\, d^2x}{\int_{-\infty}^{\infty} f^2(x)\, d^2x} = \frac{\frac{1}{2}.\int_{-\infty}^{\infty} f^2(x)\, d^2x + \frac{1}{2}.\int_{-\infty}^{\infty} f(x)f(-x)\, d^2x}{\int_{-\infty}^{\infty} f^2(x)\, d^2x}$$

This has a maximum when f(x) = f(-x) and, as $f(x) \geq 0$, the minimum of the above fraction occurs at f(x)f(-x) = 0. The fraction is bound by:

$$\frac{1}{2} \leq \frac{\int_{-\infty}^{\infty} f_E^2(x)\, d^2x}{\int_{-\infty}^{\infty} f^2(x)\, d^2x} \leq 1$$

with the remainder of the fraction being the odd portion. Obviously, then, $f_E(x) \geq f_O(x)$ for a positive function $f(x)$.

By Parseval's Theorem, these bounds also apply in the Fourier domain and the lower bound is realised when the function and its reflection about the origin do not overlap. This in turn implies that the choice of origin in the input plane is of great importance in these circumstances.

Figures 4.5 and 4.6 illustrate the phase distribution in the Fourier plane for two different letter objects. The graphs are analagous to that in figure 4.3 above in representing phase space excepting in that the phase component of a particular pixel in the Fourier plane is also scaled in order to reduce the number of points which overlap and clarify the diagram (as $64^2 = 4096$ separate points are being plotted here).

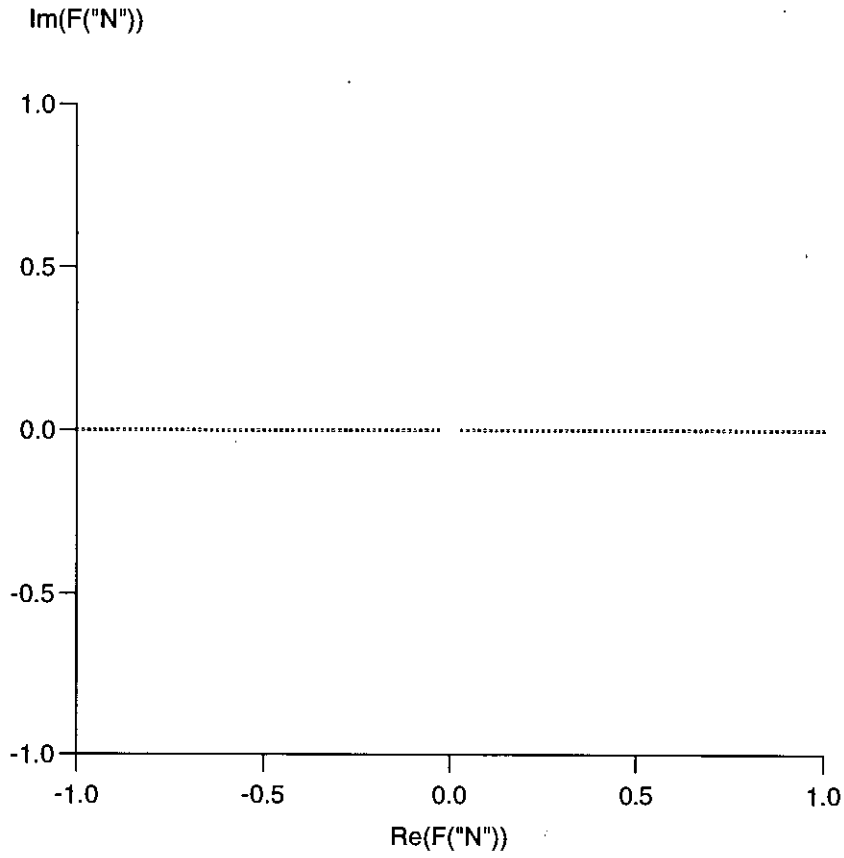
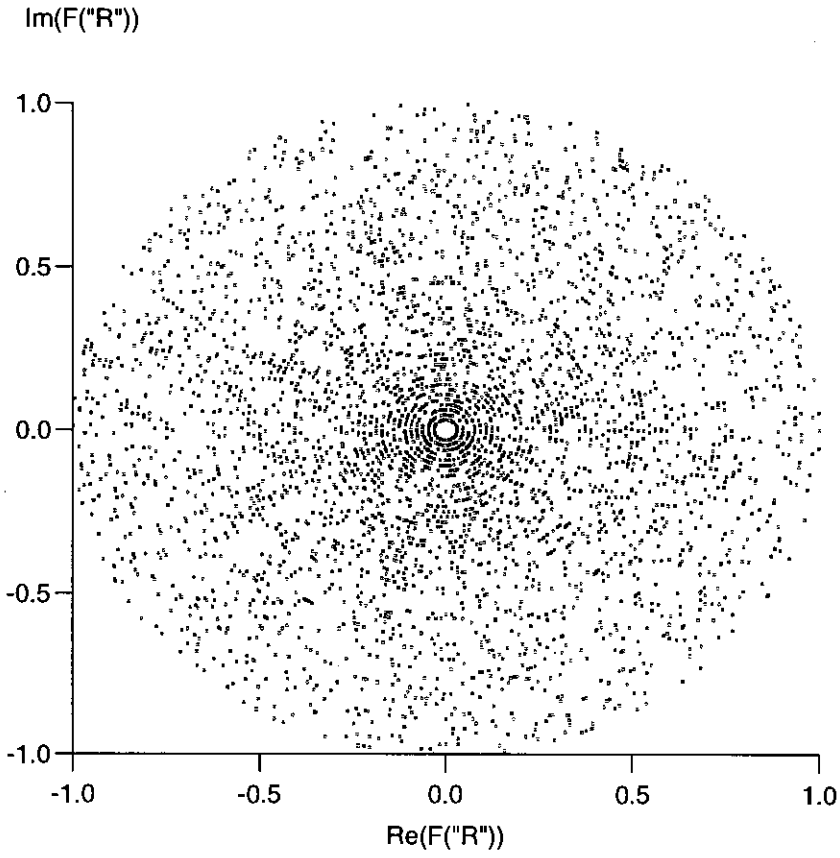


Fig 4.5. Phase content of the Fourier transform of the letter N, O

Fig 4.5 corresponds to the phase spectum of the Fourier transform of the letter N which exhibits a rotational symmetry (of degree two) in the object plane (this symmetry must be looked for diagonally - comparing top left pixel with bottom right and so on). As expected, the phase in the Fourier plane is limited to $\pm\pi$ and the Fourier plane is entirely real. Obviously this is identical to the Fourier phase spectrum of the letter O for the same reasons.

Fig 4.6(a) corresponds to the Fourier phase spectrum of the letter R. This has very little rotational symmetry and the phase components in the Fourier plane are evenly distributed through all possible values. Similarly, that of the letter A (Figure 4.6(b)) exhibits the same even distribution although it may be seen that there is some greater weighting towards the zero phase line (Im(F())=0).



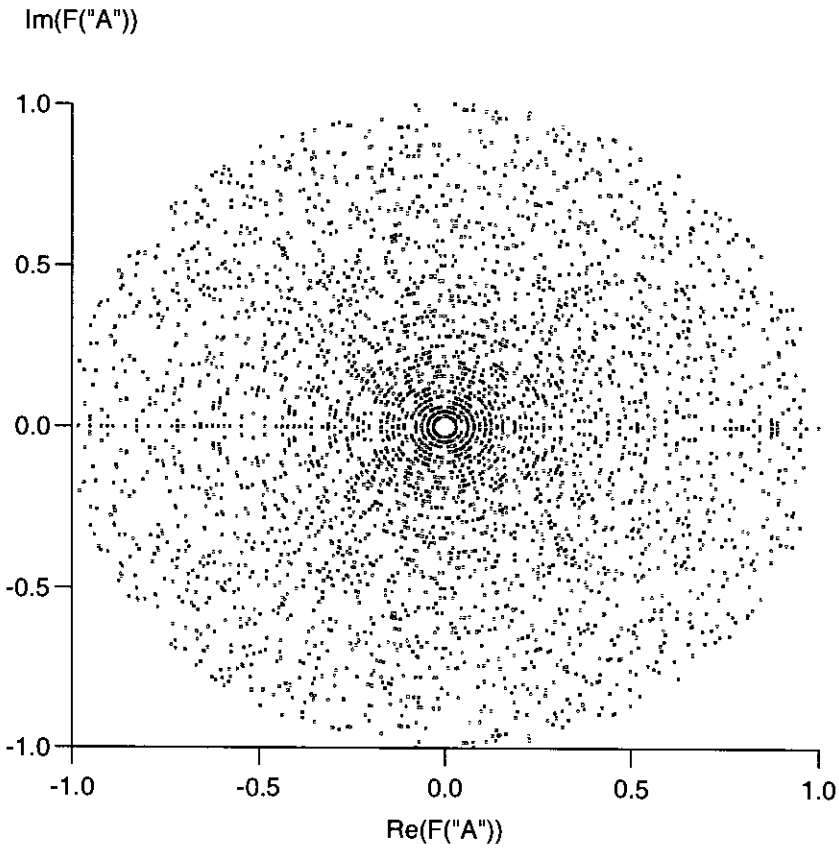


Fig 4.6 (a). Phase content of the Fourier transform of the letter R;(b) Fourier phase content of the letter A

The conclusion that may be drawn is that for different phase distributions in the Fourier plane different binarising algorithms may be more suitable. It is difficult to examine an arbitrary object and surmise what its ideal binarising algorithm might be without similar analysis to that carried out above and involving human decisions somewhere in the procedure as to the relative symmetries which an object may exhibit. This is a fundamental disadvantage of the so-called static binarisation algorithms.

Amplitude encoded binary phase only matched correlation

This is a simple technique by which a binary phase-only SLM is represented by a binary amplitude only SLM (Flavin and Horner (1989)). The binarised phase-only filter, $\phi(x,y)$, generated by any of the above (or other) algorithms and consisting of the two phases, ϕ_1 and ϕ_2 is represented by a transmission function $\tau(x,y)$ such that

$$\tau(x,y) = \begin{cases} 0, \text{ for } \phi(x,y) = \phi_1 \\ 1, \text{ for } \phi(x,y) = \phi_2 \end{cases}$$

This class of filters provide the high discrimination and sharp correlation peaks typical of the related binary phase-only filters, but as opposed to the phase-only case, these may be generated on low-cost amplitude modulators such as liquid crystal displays, or photographic transparencies.

The main advantage of moving (what seems to be retrospectively) back to amplitude modulation is that the resulting filter will not test for the rotated object as well. An object placed off centre will not produce a double correlation peak as in the case of the binary phase-only filter however, a ghost image of the original will be observed in the output plane. In addition, by the nature of the algorithm, half of the

incident optical energy is lost (on average) as it falls on OFF pixels in the Fourier plane. The phase-only filter is far more inviting from this point of view.

Other approaches to correlation

In addition to the above mentioned binary algorithms there exist others which examine ternary (or greater division) solutions to the matched correlation architecture (e.g. Lindell and Flannery (1990)) or combining amplitude and phase devices (Awwal et al (1990), Downie (1991)). This is beyond the scope of this thesis as it implies a SLM technology beyond that which was available for use.

Simulated annealing

The previous sections have attempted to indicate the many possible methods of generating the patterns required to produce binary phase-only filtering patterns which are used to generate optical correlations. There is no clear algorithm which is guaranteed to excel over the others in all situations and for all objects. Indeed, some references indicate that the choice of the phases used for the binarising levels is also important in optimising any filter-generating algorithm. Each has its merits, but none is ideal as the phase plane represented obeys different rules for different objects. In order, therefore to optimise the process, each object, or set of objects, must be taken individually into account in the calculation of its corresponding phase-only filter. Iterative techniques have been discussed (Bartelt and Horner (1985)) but the most recent developments have been in the field of simulated annealing (Kim and Guest (1990), Mahlab and Shamir (1989) and Nomura et al (1990)).

These papers indicate the potential for tailoring of the phase-filter to produce a pre-selected image for a given object. The method is by definition one which requires a large amount of number processing but produces clearly superior results to the other outlined algorithms and allows one to tailor the correlation peak as may be required for certain applications. The potential to move the numerical calculations into the optical domain is extremely tempting as this would speed up enormously the time-consuming Fourier transformation calculations.

The Simulated Annealing Algorithm

The simulated annealing algorithm iterates in order to find the values of the system variables (in this case, the phases of individual pixels on an SLM) for which the system has a global minimum for some chosen 'figure of merit' (Kirkpatrick) (usually some measure of the system energy, and here, for example, one may try to maximise the central correlation peak) . The analogy taken is of a metallurgical annealing procedure, whereby the minimum energy state of the metal is reached (and, hence, the positions of the individual atoms corresponding to that minimum energy) by cooling the system from a high temperaure to a low temperature. By cooling too hastily, the system enters energy minima which may be metastable states and thus prevent the global minimum energy state from being located. Alternatively, cooling too slowly is wasteful of resources and time.

The method outlined initially by Metropolis et al (1953) took as its subject a collection of atoms in equilibrium at a given temperature. During each step of the algorithm an atom is given a small displacement, and the corresponding change in the overall energy of the system ΔE is calculated. If the energy is reduced then the

new position of the atom is accepted, but otherwise the new position is treated probabilistically: the new position is accepted with a probability of

$$P(\Delta E) = \exp\left(-\frac{\Delta E}{kT}\right)$$

This is achieved by simply taking a random number in the interval (0,1). If this number is less than P, then the new configuration is accepted, otherwise the atom is returned to its original position and a new atom displaced for the next step of the algorithm. By repeating this procedure a number of times, the stable Boltzmann distribution is evolved for the chosen temperature.

The simulated annealing technique involves (maintaining the same analogy) first 'melting' the system at a high temperature into an energetic state. Then, once an equilibrium is achieved, cooling the system gradually until it freezes into a state where no further changes are possible (i.e. at T=0 when no perturbation of the system is accepted unless it further reduces the minimum energy (figure of merit)). The series of temperatures chosen is referred to as an annealing schedule (Kirkpatrick). Metastable states are avoided as for all temperatures T>0 transitions may occur which jump out of such local energy minima.

For a more light hearted introduction to the subject, see Appendix 2: *Optimisation by Kangaroo.*

Signal to Noise Ratio (SNR)

There have been many definitions of the signal to noise ratio in optical correlation architectures, some theoretical and others concentrating on the practicalities and applications. The main two cited in the literature are those defined

by Farn and Goodman (1988) and Horner and Bartelt (1985) and are defined as follows:

$$S.N.R = \frac{C_{max}}{N_{MS}} \text{ (Farn and Goodman)}$$

and

$$S.N.R. = \frac{C_{max}}{I_{rms}} \text{ (Horner and Bartelt)}$$

where C_{max} is the maximum correlation peak intensity, N_{ms} is the mean square output noise and I_{rms} is the rms of the intensity response outside a 50% threshold level in the correlation plane. These are justified in terms of practical correlation experiments as they promote high intensity peaks and suppressed side lobes.

Phase correction in light modulators

In 1986, Casasent and Xia published a paper investigating the effects of phase distortion in spatial light modulators. Their primary interest was in the low cost, low accuracy liquid crystal displays which are characteristically of poor optical quality and low contrast. They aimed to improve the performance of such devices by the insertion of a phase conjugate correction hologram after characterisation of the flatness of the device had taken place - typical variations were of the order of 3-10 wavelengths. Using the LCD together with the inserted hologram allowed simple correlation tasks to be performed by this poor quality device. Also, and extremely importantly, the phase correction technique allowed the system to retain its shift invariance. Correlation without the shift invariant property is not practicable.

A simpler solution put forward by Mok et al (1986) was to place the entire device within an optical liquid gate assembly, effectively padding out the phase inconformities to uniform levels. This has obvious disadvantages but suffices for

prototype systems where the inconvenience is outweighed by the need to remove such unconformities.

Horner (1988), however, looked into the effects of introducing phase errors into correlator systems in order to see if phase correction was in fact necessary at all and tried to define typical tolerances to the different correlation techniques to this phase variation. A definition of *tolerance* was chosen such that for a certain distortion k, the output SNR[1] decreased to its 3dB points (70.7% of the undistorted value). It was found that the classical matched filter configuration tolerated 3.7π of distortion whereas the phase-only and binary phase-only filters fared equally well with a tolerance of 1.6π. Translating the input image showed graphically the effects of such distortions for the classical matched filter correlation which developed two large broad sidelobes. This was not observed in the phase-only filters which (as described in earlier sections) act as high pass filters. The background noise level simply increased in magnitude.

Summary

This chapter has introduced the major algorithms utilised in the field of optical correlations, indicating the overall performance indicators and problems of the individual cases. The binary phase-only filters are the main subject of this work and it is perhaps useful to recap on the main features. Primarily, they are simple devices but with many possible algorithms with which to program them (as the choice of the discriminating angle ϕ is unconstrained there are potentially infinite such algorithms, the best choice of which will vary from object to object). Problems

[1]Note the definition of SNR used in this paper is the first one defined above (Farn and Goodman)

include noise, which is amplified by the phase binarisation process and the enforced symmetry of an object with a Fourier transform consisting of phases of 0 and π resulting in false correlations with the rotated object, but in general this is outweighed by the excellent performance in correlation systems as they produce characteristically sharp correlation peaks. Optimisation techniques have been proposed by several workers who have looked into phase correction of the filter, altering the binarized phase levels and also stochastic algorithms to achieve the desired improvements.

The field of the following chapter has limited itself to the investigation of binary phase-only filters and the matched filter correlator architecture but it can be seen that the simulated annealling technique may be extended to arbitrary degrees of freedom (incorporating optimisation of multi-level phase-only filters, combined amplitude-phase filters or alternative architectures such as the joint transform correlator).

Chapter 5

Simulated Annealing in Optical Correlation.

Introduction

The field of binary phase only filtering has been described in the previous chapter as well as elsewhere (e.g. Horner and Bartelt (1985), Barnes, Matsuda and Ooyems (1988) and others). It is clear that where generation of such filters is concerned that there is a huge variety of algorithms from which to choose. In addition, if the filter device upon which the calculated pattern is placed is poorly designed or constructed, then the experimental representation of such an algorithm may be very inaccurate, to the general detriment of the correlator's performance. It is the purpose of this chapter to show that where static algorithms fail to perform well, that, in general, a dynamic algorithm such as simulated annealing is preferred.

A bonus feature of the simulated annealing algorithm is the ability to control features in the correlation plane rather than taking the only possible solution that a static algorithm can give. It becomes possible to dictate the shape of the peak desired, to discriminate against certain classes of objects or train a correlator to discriminate between very similar objects. Importantly, too, it will take some

account of (unknown) defects in the experimental optics which would otherwise cause spurious results. This will be illustrated in the following sections.

In the computer simulations the favoured base computer used was a Sun Sparcstation 1000 with four processors. In practice, however, so many simulations were simultaneously being executed that as many computers as were available (including a Sun Sparc 2 and a Silicon Graphics Power Indigo 2) were used in order to maximise simulation throughput. It should be mentioned that no parallel use was made of the four processor computer - it was simply used to run four simultaneous simulations in order to fully exploit its potential. The number of pixels used on the Fourier plane SLM was $64^2 (= 4096)$ and the number of calculations made per simulation scaled (approximately) as n^2 with n being the number of pixels on a side of the SLM. This implies that to increase the pixel resolution to, say, 1024^2 pixels on an SLM would require an increase in the computing expense of the order of 256 - a not insignificant factor.

The results presented in this thesis are generated after of the order of 100,000 iterations per target-object pair, that number being required for the system to reach a satisfactory stable state at 'zero' temperature. By optimising the code as much as was possible, optimising the performance of the program executable through features incorporated in the Sun fortran compiler and running simultaneous simulations on separate computer processors the throughput was maximised. With approximately 7 iterations per second, each one consisting of two 4,096 point fast Fourier transforms, complex multiplication, subsidiary figure of merit calculations and array manipulations, it can be seen that a single simulation for a single figure of merit will take about 14000 seconds or 4 hours to complete. For four objects, there are sixteen cross-correlations to be performed per chosen figure of merit and it

becomes obvious that the computational aspect of this technique becomes non-trivial extremely quickly.

Computer Implementation

During the annealing process, a randomly generated SLM pattern is altered, one pixel at a time. Under certain conditions (e.g. by comparing some (suitably chosen) figure of merit with that calculated for the previous iteration) the pixel is kept at its new value[1]. If the relevant conditions are not met, however, the pixel is returned to its original value. Another (random) pixel is then chosen and the process repeated. A 'temperature' is assigned to each run of the generating program which is highly significant. A pixel change is accepted if *EITHER* the figure of merit is improved upon *OR* with a probability of $\left[1-\exp(-\frac{1}{T})\right]$ where T is the temperature of the anneal (i.e. always accept an improvement and accept deterioration with a probability which decreases with decreasing temperature). Note that this differs slightly from the description above in that the probability of acceptance of a detrimental perturbation is not related to the loss in system energy (ΔE). Whereas the above case considers a large loss in energy as a higher risk than a small loss in energy, this does not differentiate the two cases. Any energy loss is as bad as any other. Both were used for a period, but no significant differences could be seen in the results and in an effort to speed up the computer code, the extraneous calculation of ΔE was omitted.

By using a schedule of temperatures it is possibly to avoid merely arriving at some local maximum figure of merit, but to find the global maximum. An intriguing

[1]N.B. In most implementations, the system energy (cost function, C) is minimised - in most of the cases illustrated in this chapter, the system energy is maximised. This is achieved by defining a cost function of C^{-1} and minimising that. This can be seen to be directly equivalent to maximising C

effect noted during the course of this research was that in the simple cases (where, for example the figure of merit was just the central intensity) the same final iterated result could be obtained when omitting the temperature schedule entirely, and commencing the simulation with a temperature of 0, thereby vastly reducing the computational requirements. On discussing this with other workers (Samus (1994), Smith (1995)) it is noted that this is not a unique discovery and that these others have also noted this effect. This suggests that the mathematical form of the figure-of-merit function as a function of the SLM filter pixel values is well behaved and without local minima in which to stop the iterative process. This reasoning was not pursued further and it should be noted that all of the simulations presented conformed to the full temperature schedule.

In these sets of simulations, a figure of merit was chosen and calculated from the cross-correlation of the object of interest and an object whose binarised (phase-only) Fourier transform was represented in the Fourier plane by the (iteratively changing) computer generated SLM pattern on a binary phase-only SLM. The results could then be compared with cross-correlations produced using patterns generated by the other main methods (Psaltis, Horner and Cottrell algorithms as described earlier).

A schedule of temperatures for the process is shown in Table 5.1. Once the zero temperature was reached, the figure of merit was monitored by the computer program and the simulated anneal halted once no change in the figure of merit can been made for any flipped pixel (at zero temperature any acceptable such change must necessarily constitute an improvement in the system energy).

Anneal Number	Temperature (T)	$1-e^{-T}$
1	1.0	0.6321
2	0.5	0.3935
3	0.25	0.2212
4	0.125	0.1175
5	0.0625	0.0606
6	0.03125	0.0308
7	0.015625	0.0155
8	0.0078125	0.0078

T is halved repeatedly until at some predetermined value it is set to zero.

n	0.0	0.0000

Table 5.1. Simulated annealling temperature schedule

Comparison with Other Algorithms

Several representative objects were chosen and these appear in figure 5.1 overleaf. The criteria involved in choosing the set of objects were as follows: four letters were chosen (A, R, O and N) to represent regular objects with no degrees of symmetry (R), one degree of symmetry (A), one degree of rotational symmetry (N) and circular symmetry (O). These objects were rendered on a 64x64 pixel grid and each comprised 1501 ON pixels and the remaining 2595 ($64^2 - 1501$) pixels OFF in order to normalise the energies in the simulated optical systems and allow direct comparisons between the letter objects.

In the following sections, the description 'object' refers to the pattern placed in the input (object) plane of the correlator whereas the term 'target' refers to the

pattern whose binarised Fourier transform is positioned in the filter plane of the correlator.

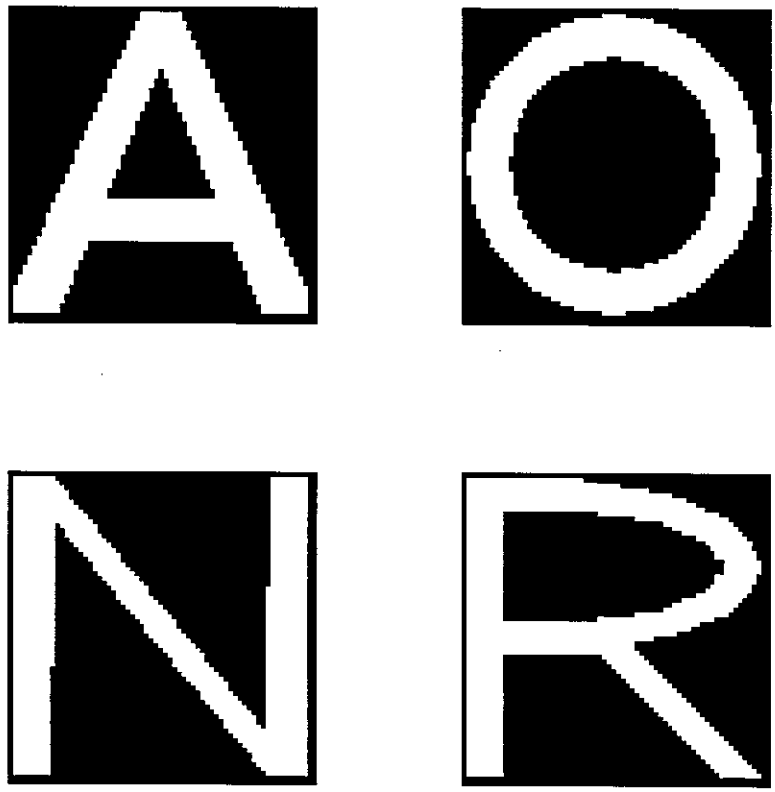


Figure 5.1. Input objects A, N, O and R

Tailoring the Correlator Output

By its nature this is an inherently adaptive algorithm - its advantage over other algorithms is precisely the fact that it is not static. By a suitable choice of the Figure of Merit (or *cost function*), it is possible to tailor the correlator output according to the application in question. In the majority of papers concerning optical correlation, the aim is to obtain as output a large intense peak - for theoretical demonstrations this may be adequate, but the practicalities are often quite different. For example, when the pixel dimensions of a detector are small a very sharp peak (of possibly lower intensity) is preferential to one which covers several of the detector's pixels or alternatively, a peak which is broader but may contain a larger proportion

of energy in the correlation plane may be more suitable depending on the detector sensitivity.

An alternative scenario is that of discerning one of several similar targets where a conventional correlator will detect similar targets with a similar peak (Kim and Guest (1990)) which implies a third case in which an object is specifically being discriminated against by the system (an anti-correlator).

Sample (Illustrative) Simulation Output

Figures 5.2 to 5.5 illustrate some typical aspects of the simulated annealling process. In this case, the cost function was simply chosen to be the central peak intensity[2] and the object used was the figure O illustrated above (figure 5.1). This figure-of-merit is understood to be a crude measurement of the correlation efficiency and it is informative to examine it as an example. Figure 5.2 is a graph of Iteration number versus Figure-of-Merit and is shown for the temperature being set at a value defined by the temperature schedule until some equilibrium is reached (the system energy reaches some maximum for this temperature), then geometrically changed (halved) and the iteration to equilibrium repeated as described earlier. The temperature changes stand out as regions of the graph where the gradient is higher than in the preceding region, as would be expected, and these are highlighted at the points marked by a short horizontal line on the graph. In this case, the figure of merit is being maximised - it is a simple extension to the algorithm to minimise a particular figure of merit.

---

[2]The central peak intensity is the maximum point intensity in the correlation plane.

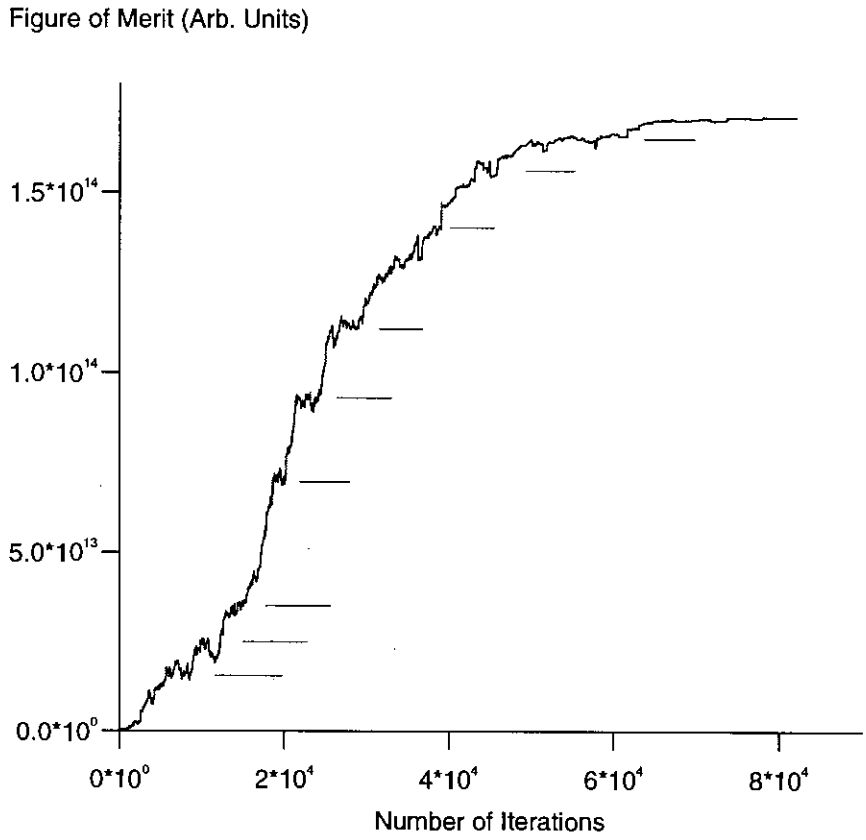


Figure 5.2. Graph indicating the progress of the figure of merit function with respect to the iteration number.

As can clearly be seen, the overall form of the iteration and the convergence of the graph proceeds through a series of characteristic steps. The approach to an overall maximum is asymptotic and the final temperature of the schedule (0) is maintained until each pixel has been flipped at least once and the system tested for potential improvement at each 'flip'. Once no further improvement is possible by any pixel being flipped, then the iteration is considered to be complete, the program halts and the results may be analysed.

Output correlations checkpointed at times during the iteration are shown in Figure 5.3 to 5.5

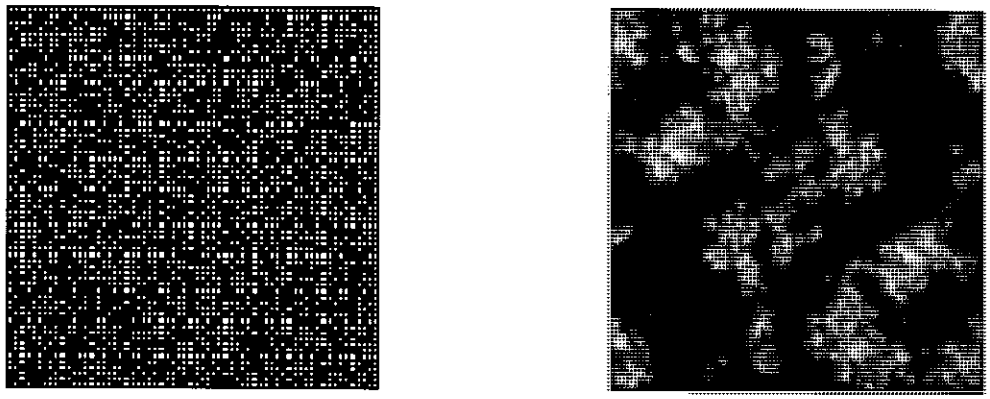


Figure 5.3. Initial state of system; (a) random binary SLM pattern (target pattern for letter O)
(b) Output correlation plane (normalised central peak intensity=I.0)

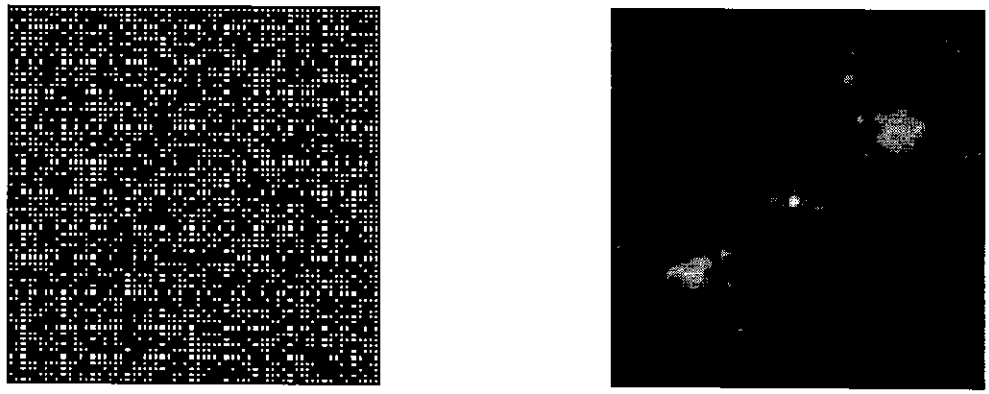


Figure 5.4. Intermediate state of system; (a) random binary SLM pattern (target pattern for letter O)
(b) Output correlation plane (normalised central peak intensity=53 (relative to Fig 5.3))

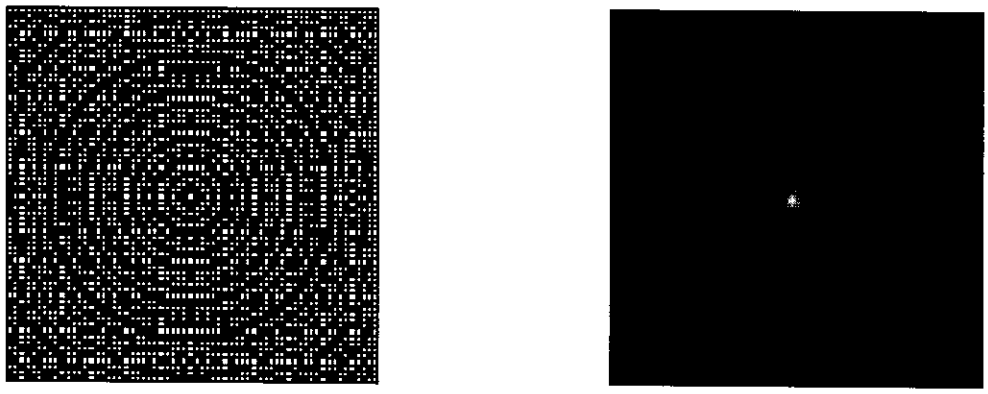


Figure 5.5. Final state of system; (a) random binary SLM pattern (target pattern for letter O)
(b) Output correlation plane (normalised central peak intensity=I130 (relative to Fig 5.3.))

Figure 5.3(a) illustrates the initial state of the filter SLM in the Fourier plane of the simulated correlator and Figure 5.3(b) shows the resultant correlation plane. Figures 5.4(a) and 5.4(b) show the same planes respectively but at a midpoint of the iteration procedure - the central peak is now dominant in Figure 5.4(b), and the binary SLM pattern shows signs of losing the initial random scheme of pixel states. In this case, the checkpoint was taken at the second temperature change indicated in figure 5.2 (after approximately 12000 iterations).

The final Figures (5.5(a) and 5.5(b)) show the final state of the SLM and corelation planes as the anneal terminates. The peak is extremely sharp and dominant as is expected with a binary phase-only matched filtering system and the SNR is obviously vastly improved from the previous figures. The SLM has also converged to a distinctive pattern corresponding to the optimal binarised Fourier transform for this particular object and chosen figure of merit.

Attention should be brought to the figure descriptions: the intensity scale in the first two figures (5.3 and 5.4) has been exaggerated in order to bring up the details in the output plane. The peak correlation intensities noted for figures 5.3 to 5.5 are calculated relative to the peak intensity of the worst case (the random scheme of Figure 5.3). The annealling process has improved the peak correlation intensity by a factor of over 1000 and produced a sharp central correlation peak as desired.

Note that in this case, only a single pixel is examined in order to calculate the figure of merit and here, it is the central pixel. By choosing to maximise the intensity at another pixel, it is possible to iterate to a non-central correlation peak.

Choice of Figure of Merit

Of primary importance in the simulated annealing scenario is the choice of which output plane characteristic to optimise. One of the characteristics of this type of algorithm is that the optimised figure may be completely arbitrary and it is dependent on the system requirements as to what to choose. A number of different figures of merit were chosen in this series of experiments and these varied from simple central correlation plane intensity to more complex ratios of the correlation plane energy distributions. These, and computer generated results, are tabulated and illustrated below. Direct comparisons were drawn between the performances of the static algorithms and the simulated annealling algorithm for the primary case (maximum correlation peak intensity) in order to ascertain some measure of the effectiveness of the algorithm.

Simulations

A basic measure of the ability of a correlation algorithm is the peak output correlation intensity, and it is useful to examine the relative performances of the static and simulated annealing algorithms with this simplest case. The figure of merit in the computer program is simply the maximum point intensity in the output plane of the correlator. For the case of the autocorrelation of the letter O the results have already been presented graphically in figures 5.2 to 5.5 but for clarity, the numerical results are tabulated in table 5.2 below:

Autocorrelations of the four object letters were calculated using the main methods described and the figure of merit calculated after each calculation. The results tabulated are relative to a figure of merit calculated when using a binary

phase only filter for that particular correlation. For information, the performance of the phase-only (not binarised) correlator is included, with correlation peak again calculated relative to the best performing binary algorithm. This phase only filter was calculated by taking the Fourier transform of the object, setting its amplitude to 1 but keeping all of the phase information as opposed to binarising it. Theoretically, by keeping more information about the object, this should provide the highest peak intensities in the autocorrelation plane relative to the binarising algorithms.

	R⊕R	O⊕O	N⊕N	A⊕A
Phase Only	*2.26*	*1.00*	*1.00*	*2.29*
Cottrell	0.94	1.00	1.00	0.91
Psaltis	1.00	1.00	1.00	1.00
Horner	0.79	0.55	0.30	0.89
Sim. Anneal	0.98	0.97	0.97	0.98

Table 5.2: Relative comparisons of autocorrelations of the letters R, O, N and A. (X⊕Y in the upper row of the table indicates cross-correlation between 'X' and 'Y'). Accuracy of results calculated to ±0.005.

The results show that the Horner-Bartelt algorithm suffers in this comparison (as expected when considering the arguments presented against it in Chapter 4) with the Psaltis algorithm slightly shading the Cottrell algorithm and that the simulated annealling algorithm fares consistently almost as well as the other two main binary phase-only algorithms. The fact that it does not equal or outperform them in all cases may be accounted for by two factors. Firstly, the annealling temperature schedule may not be quite "slow" enough i.e. at the final stages of the iteration, the system falls into a slight energy minimum out of which it cannot leap. Secondly, the other phase-only algorithms are expected to perform extremely well anyway and especially in the case of the letters N and O where the symmetry of the object

dictates that the Fourier transform will be EXACTLY represented by a Psaltis or Cottrell based filter so these are guaranteed to be optimal in the binary phase-only regime anyway under these (idealised) conditions.

Cross-correlations, too, are interesting to compare as these emphasize the ability of a correlator system to differentiate between classes of object. Tables 5.3, 5.4 and 5.5 show how the Cottrell, Psaltis and Horner algorithms respectively fare (again using the four standard letters as objects and targets). In these cases, the peak intensity of the cross-correlations are calculated relative to the peak auto-correlation intensity. A low value for the cross-correlation is obviously preferred.

Target letter / Object letter	R	O	N	A
R	1.00	0.07	0.17	0.30
O	0.02	1.00	0.04	0.02
N	0.09	0.04	1.00	0.03
A	0.05	0.13	0.07	1.00

Table 5.3. Cottrell algorithm performing cross-correlations

Target letter / Object letter	R	O	N	A
R	1.00	0.07	0.16	0.05
O	0.03	1.00	0.04	0.03
N	0.19	0.04	1.00	0.03
A	0.06	0.11	0.06	1.00

Table 5.4. Psaltis algorithm performing cross-correlations

Target letter / Object letter	R	O	N	A
R	1.00	0.07	0.04	0.07
O	0.03	1.00	0.04	0.05
N	0.11	0.10	1.00	0.08
A	0.06	0.07	0.05	1.00

Table 5.5. Horner algorithm performing cross-correlations

Table 5.6 repeats the above using the phase-only (not binarised) filter algorithm described earlier in this chapter.

Target letter / Object letter	R	O	N	A
R	1.00	0.03	0.07	0.02
O	0.04	1.00	0.04	0.05
N	0.12	0.04	1.00	0.04
A	0.04	0.05	0.03	1.00

Table 5.6. Phase only (not binary) algorithm performing cross-correlations

A point to which attention should be drawn is that X correlated with Y is not *necessarily* the same as Y correlated with X (for X, Y any input objects and correlation involving the binarized Fourier transform of the target object). This is due to the differing symmetries of the target objects which implies a less accurate representation of the target Fourier transform as a filter in the Fourier plane of the correlator. The cross-correlation of X and Y in these simulations is not the actual cross-correlation but the binary phase only cross-correlation. Thus, with X the object and Y the target, the cross-correlation is actually $X.Y_B$ with Y_B being the representation of the target Y with a binary phase only pixellated Fourier transform

(i.e. the filter function). Similarly, with Y the object and X the target the cross-correlation calculated is $Y.X_B$.

It is now less easy to determine which binarising algorithm performs best when considering the new quality figure for the system. Although the Horner-Bartelt algorithm does not produce as high a peak in the output plane, its discriminatory power is at least as good. Each binarising algorithm has a weak point - Cottrell fares badly (ratio of cross-correlation to auto-correlation greater than 0.10) with R.N and R.A and A.O, Psaltis with R.N, N.R and A.O and Horner with N.R and N.O

A more customised figure of merit was chosen to be the ratio of the central peak energy to the sum of the energies of the surrounding pixels (i.e. the intensity in the centre of the correlation plane divided by the sum of the intensities of those pixels which form a square centred on that central pixel). This is expected to result in a non-infinite expression with a maximum value of 1 (when the central peak contained all of the energy in the central defined square) but this, as stated, does not satisfactorily define the desired effect as an extremely large figure of merit may occur with only a small central peak and in fact the expression does become infinite with a zero intensity central region. To counter this behaviour, the importance of a high central peak was emphasised by squaring the central peak intensity. This was expected to optimise the filter to produce a more isolated high intensity peak with a low intensity 'depletion zone' surrounding it. This was seen to be potentially useful when utilising an array of smart photodetectors (or a smart SLM (Johnson, McKnight and Underwood (1993))) to detect the correlation peak output. In those cases some equation relating the intensities of groups of pixels rather than solitary pixels might be realised electronically or even optically.

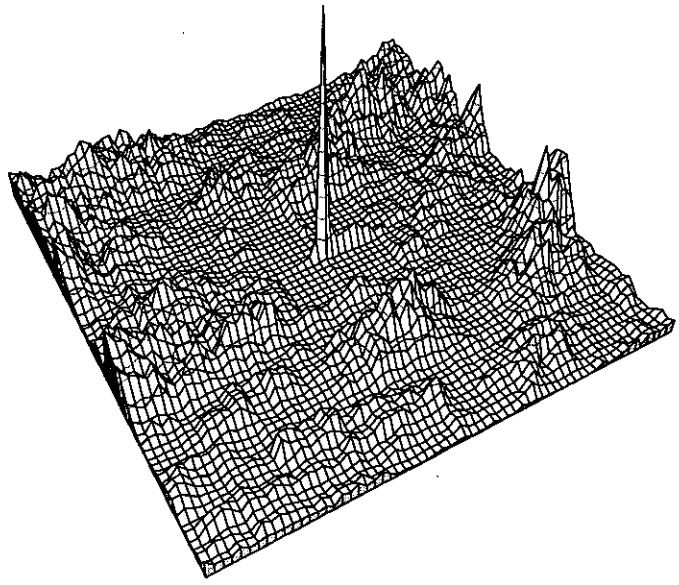


Figure 5.6. Correlation plane for 'depletion zone' filter. Autocorrelation of the letter R.

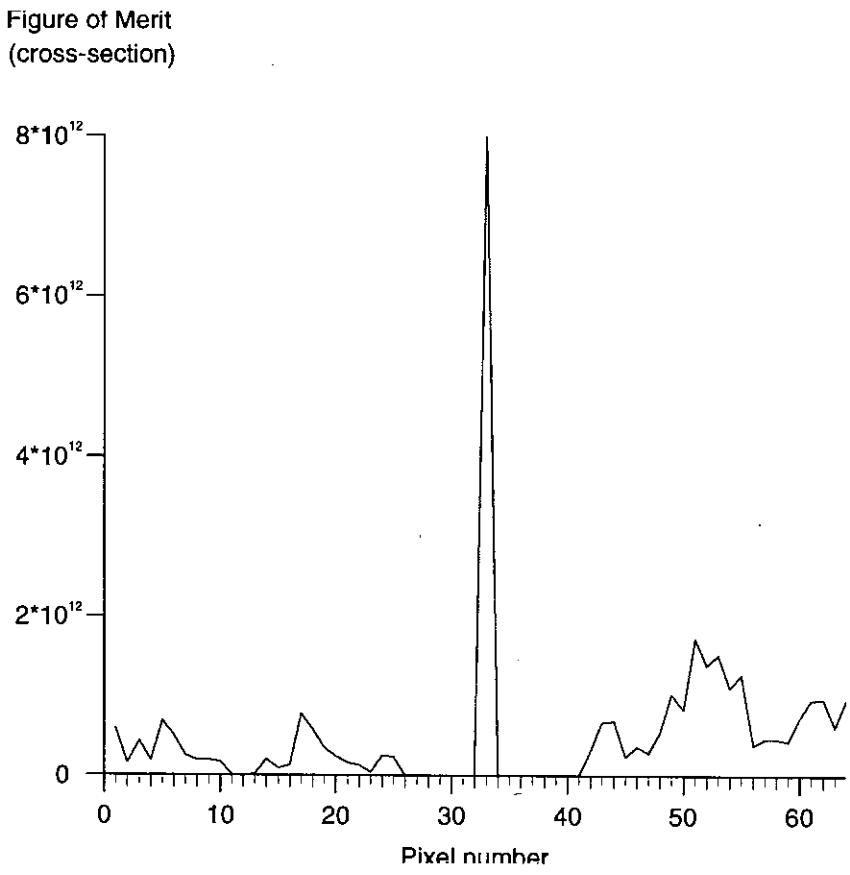


Figure 5.7. Cross section of figure 6 emphasising the salient features of the correlation plane (letter R).

Taking the object R as a sample (i.e. the binary filter generated selects for an R) a simulated anneal was performed and figure 5.6 illustrates a typical output plane. The central, sharp peak is obvious as is the surrounding low level plateau. Figure 5.7 shows a graph of the diagonal cross section of the correlation plane through the central pixel, again emphasising the sharp peak and flat surroundings.

The depletion zone is not in fact zero intensity although the graph may suggest that: the pixel energy levels within that zone are approximately one thousandth that of the central intensity and simply do not show up on this scale. Figure 5.8(a) shows the state of the SLM after the annealling process and figure 5.8(b) the Fourier transform of such a filter. Figure 5.8(c) indicates the generated pattern optimising for central peak intensity and 5.8(d) the Fourier transform of 5.8(c). The 'depletion zone' filter corresponds to a similar but more diffuse object than the standard case, and it is this diffuse nature which produces the characteristic form of the output plane.

It is perhaps interesting to compare these results with those obtained by Heddle (1993) for the case where a SLM with randomly placed (within certain strictures) pixels was used to suppress the nth order replications in the power spectrum of the SLM. The SLM pattern generated through simulated annealling has attained a characteristic 'randomness' when compared with the patterns generated for simpler figures of merit.

Noise must obviously have an effect here. Without further simulation it is difficult to estimate the extent to which the effects of noise would be annulled by such an algorithm but it is expected that noise (where we are talking about static rather than dynamic noise i.e. a noise field which is constant in time and space) on

the Fourier filter would have its effects reduced as it is the output plane when using that *particular* filter (where such noise is effectively a part of the filter) which is used as a measure of the device. Noise in the object plane would be incorporated into the calculated filter for that object and so, if the noise is indeed constant in time and space then this too will be reduced in its effects. Noise which fluctuates with time will affect the output of such an optical processor by an unknown amount.

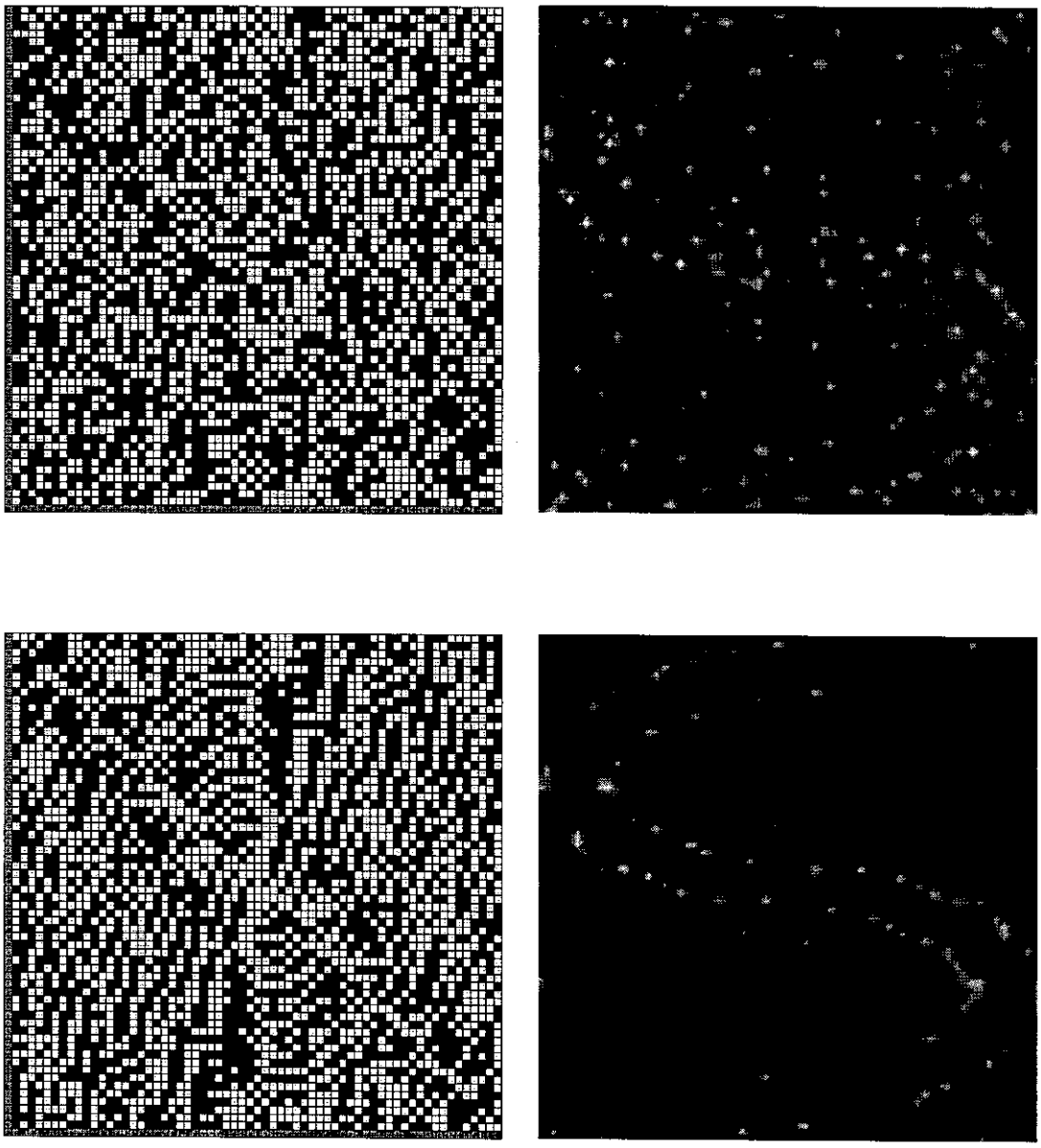


Figure 5.8. (a) State of 'depletion zone' filter SLM after annealling. (b) Fourier transform of (a) (c). 'standard' filter; (d) Fourier transform of (c)

This fairly artificial figure of merit is tabulated below for auto-correlations using the Psaltis, Cottrell, Horner and Simulated Annealing algorithms (Table 5.7). Again, the static algorithms perform as well as each other and, in this case, to approximately the same level as the phase only algorithm. However, it is very obvious that the simulated annealing algorithm has 'trained' the filter to optimise the chosen figure of merit and far outstrips the competition. This is emphasised in the following tables which repeat the format of Tables 5.3 to 5.6 earlier in this chapter in displaying cross-correlation values of the chosen figure of merit.

Target letter Object letter	R.R	O.O	N.N	A.A
Phase Only	*0.366*	*0.305*	*0.206*	*0.212*
Cottrell	0.300	0.305	0.206	0.202
Psaltis	0.295	0.305	0.206	0.199
Horner	0.350	0.269	0.102	0.107
Sim. Anneal	1.000	1.000	1.000	1.000

Table 5.7. Auto-correlation values using depletion zone' figure of merit

Table 5.8 uses the simulated annealling algorithm and at once suggests that this figure of merit has improved the inter letter discrimination one hundredfold.

Target letter Object letter	R	O	N	A
R	1.0000	0.0001	0.0009	0.0003
O	0.0006	1.0000	0.0007	0.0002
N	0.0008	0.0001	1.0000	0.0007
A	0.0001	0.0001	0.0008	1.0000

Table 5.8. 'Depletion zone' SLM cross-correlation values using simulated annealling algorithm

Table 5.9 repeats the calculations for the Horner static algorithm which performed the best of the three tested and also implies an improved discrimination in all cases.

Target letter / Object letter	R	O	N	A
R	1.000	0.027	0.024	0.002
O	0.032	1.000	0.003	0.011
N	0.027	0.023	1.000	0.0001
A	0.008	0.004	0.041	1.000

Table 5.9. 'Depletion zone' figure of merit for Horner algorithm.

Whilst it is not suggested by any means that this is the ultimate discriminating algorithm (there is an infinity of variations possible), it is suggested that utilising such a customised merit figure may have uses in the improvement of correlator performance for specific applications and that such an iterative technique is the only way of generating such filters. The filters generated through simulated annealling perform approximately with a factor of 10 improvement in this discrimination factor.

A second application of the dynamic phase filter pattern generation is that of creating an 'anti-correlation' figure of merit. Kim and Guest (1990) with which to test simultaneously for a single object (or family of similar objects) whilst discriminating against another type of object (or family of similar objects). Similar simulations were performed using peak correlation intensity as the 'pro' figure and the inverse of the standard deviation of the energy distribution in the correlation plane for the 'con' figure. These results are not presented here as sufficient material

exists in the references to illustrate the point and as such this idea was not pursued beyond a feasibility simulation.

It was then decided to investigate further this 'customisation' of the correlation peak - after having successfully illustrated the sharpening effect on the peak of the previous figures a merit figure was devised to maximise the energy confined in the correlation peak: this was taken to be the ratio of the energy in the central spot (here arbitrarily chosen to be a circle of diameter six SLM pixels positioned at the centre of the Fourier plane) to the energy in the plane outside this spot. This is expected to reduce the central peak intensity but raise the energy concentrated in the central pixels - particularly useful in practical applications once photodetectors are being used and energy concentrations require to be improved.

A sample correlation peak is displayed in figure 5.9 below.

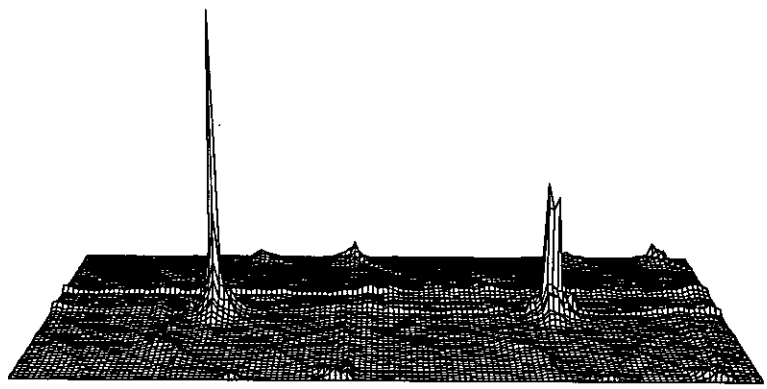


Figure 5.9. 'Sharp' peak autocorrelation of the letter R compared with 'Broad peak' autocorrelation (generated by simulated annealing)

In this case, the energy included in the central spot of the correlation peak of the broad peak relative to the energy outside has increased from 12% to over 14.5%. The change in the shape of the correlation peak is marked - no longer is there a

single sharp maximum, but instead the central energy has been smeared over the central area into a plateau. The central peak intensity has reduced by a factor of 2.3 yet the energy in the peak has increased.

Effects of phase errors in input and Fourier planes

As a method of examining the capability of such an algorithm successfully to anneal out optical imperfections from a correlation system, several scenarios were explored. Phase wedges were superimposed on either the input (object) plane or the filter (Fourier) plane. This simulates an object or filtering SLM with imperfections caused by manufacturing or inherent limitations (such as the LCTV devices described in Chapter Two).

Two such wedge forms were calculated and these are illustrated in figures 5.10 and 5.11. The lowest point in these figures coresponds to 0 phase distortion.

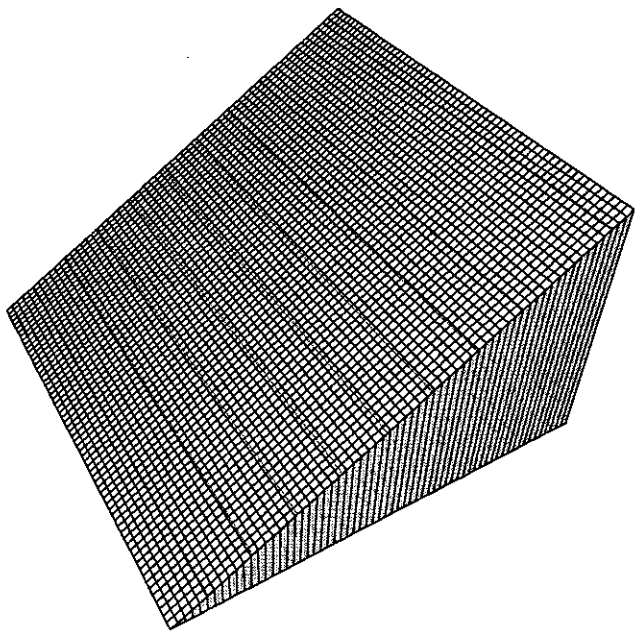


Figure 5.10. Simple linear wedge phase distortion. Maximum distortion π radians.

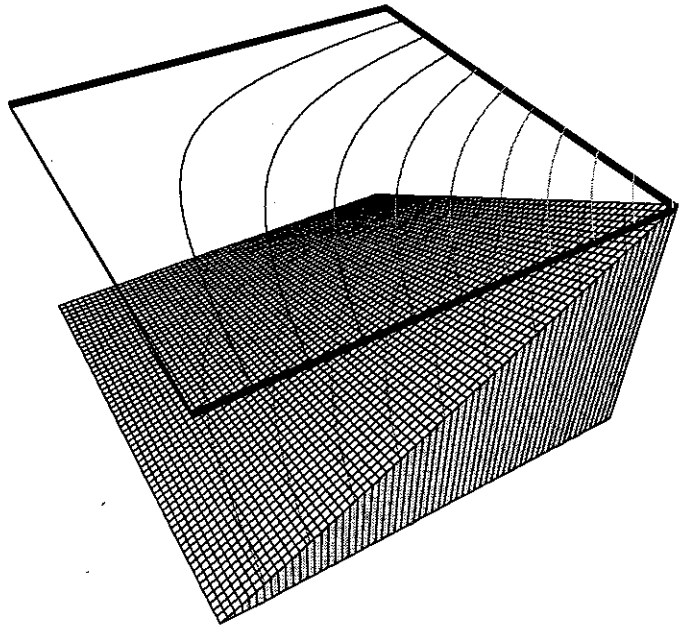


Figure 5.11. Quadratic wedge phase distortion. Maximum distortion π radians.

Figure 5.10 shows a simple wedge with the phase addition in the plane rising from 0 at the left hand end to $n\pi$ at the right hand end. The value of n can be arbitrarily controlled through the program interface and here is given the value 1.

When the Fourier transform of a uniformly illuminated object with this phase form is taken, the phase spectrum in the Fourier plane is predominantly 0, but the single point output obtained by Fourier transforming a uniform amplitude object is shifted by one pixel per π radians maximum distortion wedge. As the phase of the Fourier transform of an object varies rapidly over the plane it may be seen that shifting one transform relative to another by even one pixel in the Fourier plane where the complex filtering takes place will cause unpredictable effects in the correlation plane due to the positional mismatch. In addition, such a wedge in the Fourier plane will shift the correlation peak in the output plane.

Figure 5.11 shows a second such wedge again with a maximum distortion of $n\pi$ (here, n=1). In this case the phase variation is more complex, being a quadratic

rather than a linear variation. The projection over the 3-d graph indicates the phase contours and the exact form of the slope which was chosen to simulate experimental measurements of a LCTV as published by Horner (1988) and as such, this is a useful distortion to apply, being exactly of the form of measured phase variations. As the phase varies rapidly over the Fourier plane, it is obvious that the effects of applying such a distortion is particularly significant.

A third form of phase error encountered in this field is that which occurs when the device used in the study is not a pure binary amplitude or phase modulator, but some combination of the two. A potential situation where this may be useful is one in which a binary phase-only SLM does not binarise to $(\cos(\pi),\sin(\pi))$ and $(\cos(0),\sin(0))$ but to a more general phase $(\cos(a),\sin(a))$ and $(\cos(0),\sin(0))$ where a may vary between 0 and 2π radians. Similarly, an amplitude only SLM may have an additive phase factor.

It is now necessary to examine how the various algorithms fare on application of such phase distortions in either the object plane or the Fourier plane.

Phase Distortion in the Object Plane

Taking the case of the simple linear wedge (Fig 5.10) and superimposing this form onto the object, it is evident that for an 'auto-correlation', the target is no longer the same as the object and the 'auto-correlation' is in fact a cross-correlation of two functions whose amplitudes are identical but whose phase components are not. This has the effect of mismatching the Fourier plane filter with respect to the Fourier transform with which it is intended to match.

Figure 5.12 is a graph of correlation peak intensity versus n ($n\pi$ is the maximum phase distortion applied, as defined previously) where object and target are both taken as the letter R, O, N and finally A. It should be noted here that the number calculated is the maximum peak ANYWHERE in the correlation plane and not necessarily in the centre of the plane as would normally be expected.

The form of the four lines indicates a fairly rapid drop off in central peak intensity with a minimum at around n=1.75, rising to a secondary maximum and falling again at approximately n=3.75. The peak intensity then drops off towards a general background noise level at higher values of n (n > 6).

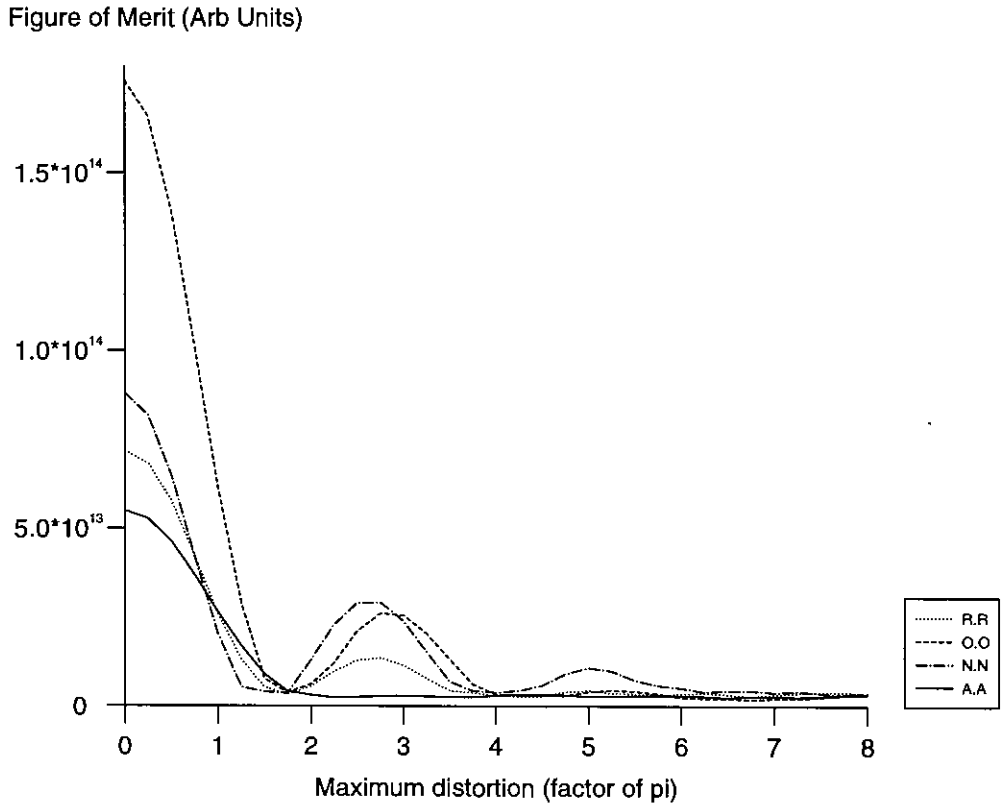


Figure 5.12. Autocorrelation peak intensity in the output plane versus maximum phase distortion (n) in the input plane (letters R, O, N and A).

Very similar results are obtained for all four letters. It should be noted that the fall off is very similar (independent of the level of symmetry of the input object) and is primarily a function of the amount of the phase distortion (and hence of the

level of mismatch of the filter in the Fourier plane) rather than of the object itself. In each case, the peak correlation plane intensity drops from its maximum at n=0 to a minimum at around n=1.75 with a second maximum occuring at around n=2.5.

It is also educational to examine what happens to the cross-correlation intensities for a varying wedge distortion in the input plane. A correlator is primarily a discriminatory device and must therefore perform well as a cross-correlator as well as an auto-correlator. To this end, Figure 5.13 was plotted corresponding to the same experiment performed for different target and object letters (here the target, for which the filter was calculated, was the letter R).

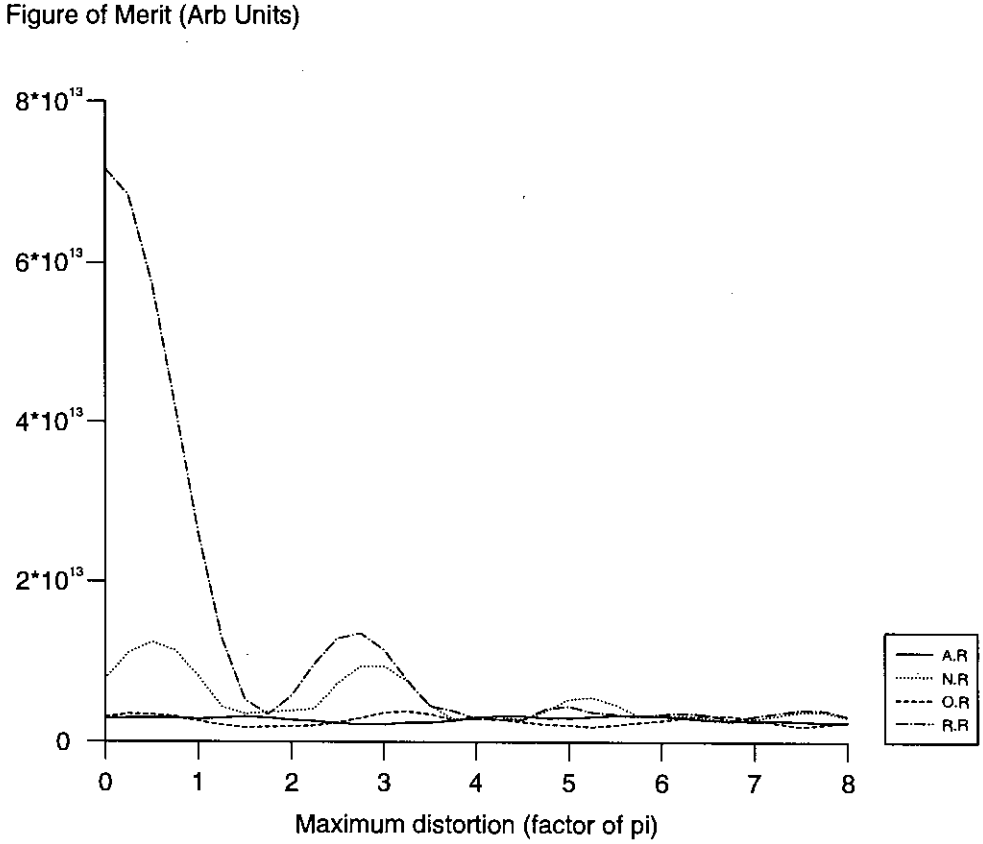


Figure 5.13. Central peak intensity in the output plane versus maximum phase distortion (n) in the input plane (letter R, cross-correlations, linear wedge in the input plane).

From this graph, it can be seen that beyond n=4 the device is essentially useless for the static algorithm. The cross-correlations are affected strongly by the addition of a phase wedge to the input plane and from these results it is possible to

generate a third set of graphs of the ratio between the autocorrelation peak intensity and the cross-correlation peak intensity as a function of the maximum phase distortion introduced (Figure 5.14). This may be taken as a measure of how the 'faulty' correlator actually performs in a discriminatory scenario.

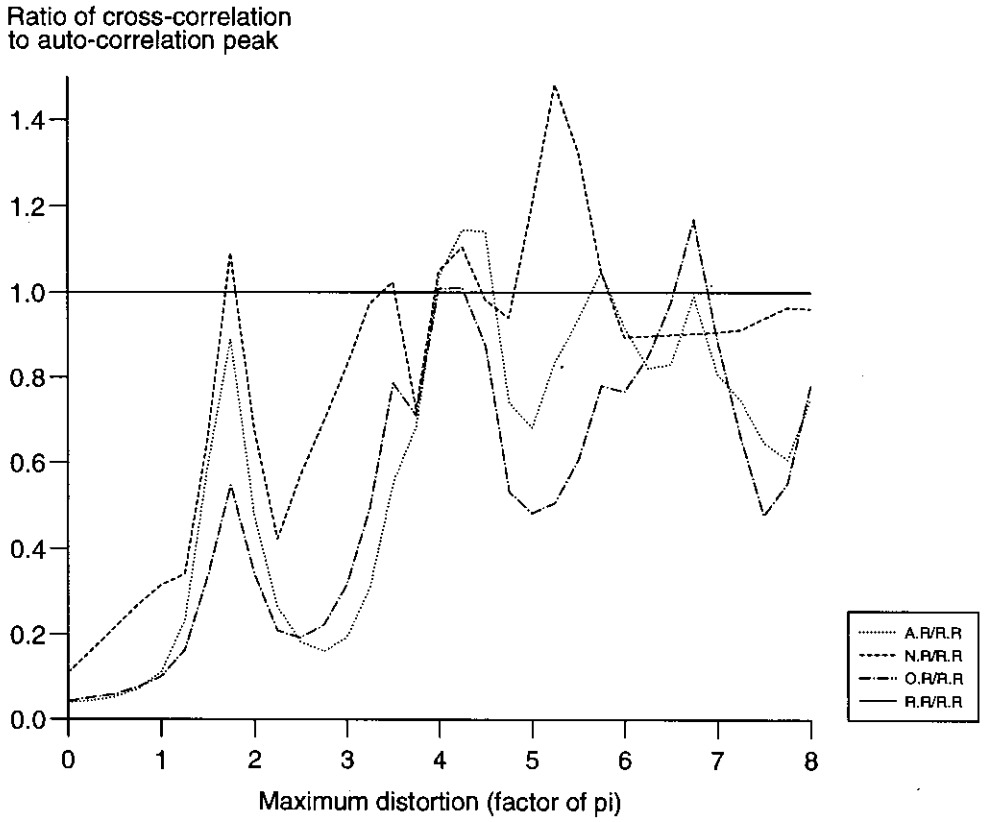


Figure 5.14. Ratio of autocorrelation peak intensity and cross-correlation peak intensityversus maximum phase distortion (n) in the input plane (letter R).

For points where the cross-correlation graphs cross the R.R/R.R line the correlator cannot discriminate between the letters offered as objects and the target.

Now, obtaining the final figure of merit for simulated annealling runs auto-correlating the various letters and using different phase distortions in the object plane results in the graphs of figures 5.15 to 5.18. For comparison, the equivalent autocorrelation results using the Psaltis algorithm are presented alongside. In all four graphs the simulated annealing algorithm has forced the final correlation peak to be both centralised and also at an equivalent level to that of the zero distortion case.

This is really as should be expected - the algorithm simply produces a binary phase only filter for the modified input object and this results in a true autocorrelation rather than the pseudo autocorrelation generated using static algorithms.

It is noted that there is a distinct periodic nature to the results produced from the simulated annealling algorithm in these figures. While the interactions within this process are complex, it is expected that this is a function of the pixellation of the correlation plane where, if the correlation peak is shifted slightly from a single pixel, the same intensity may be 'shared' over two or more pixels, thereby artificially seeming to flatten the correlation peak by some proportion. There is, perhaps a need for further investigation into this effect to determine whether it is infact an artifact of the computer simulation.

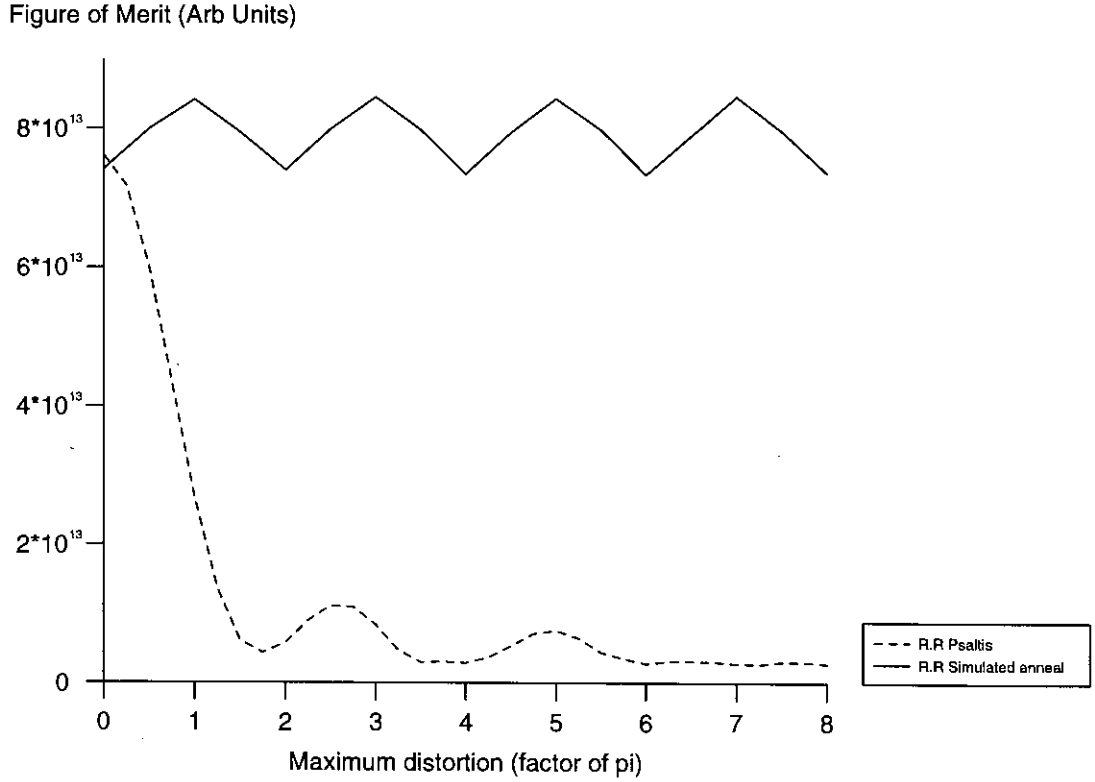


Figure 5.15. Performance of simulated annealling algorithm over Psaltis algorithm (R.R)

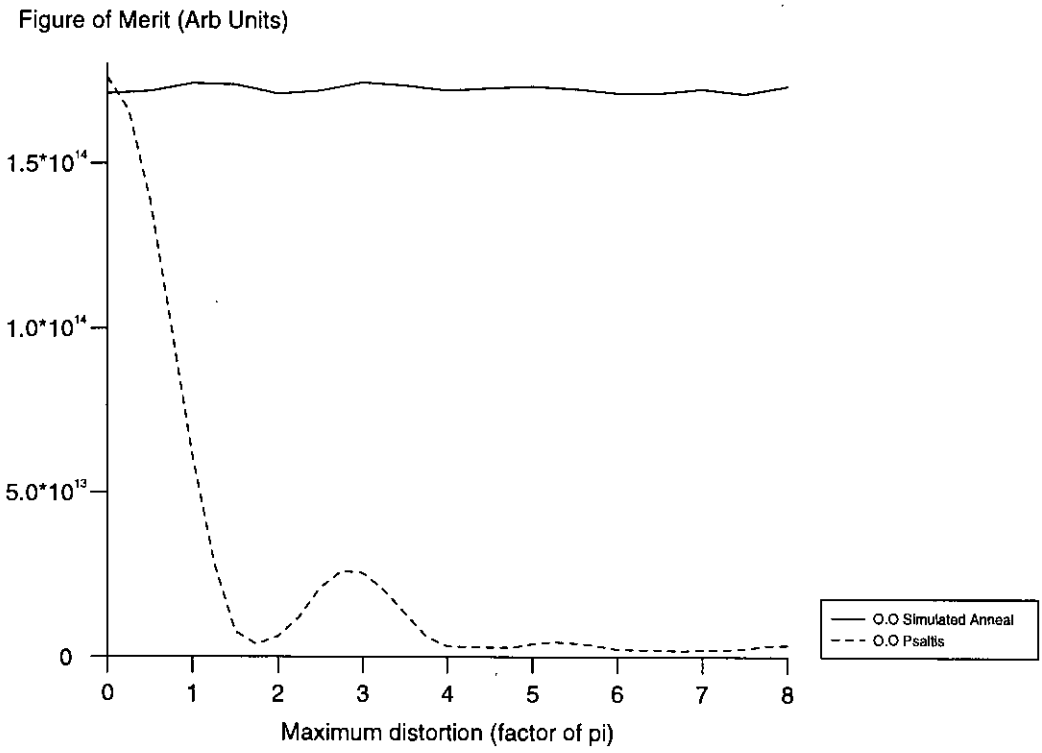


Figure 5.16. Performance of simulated annealling algorithm over Psaltis algorithm (O.O)

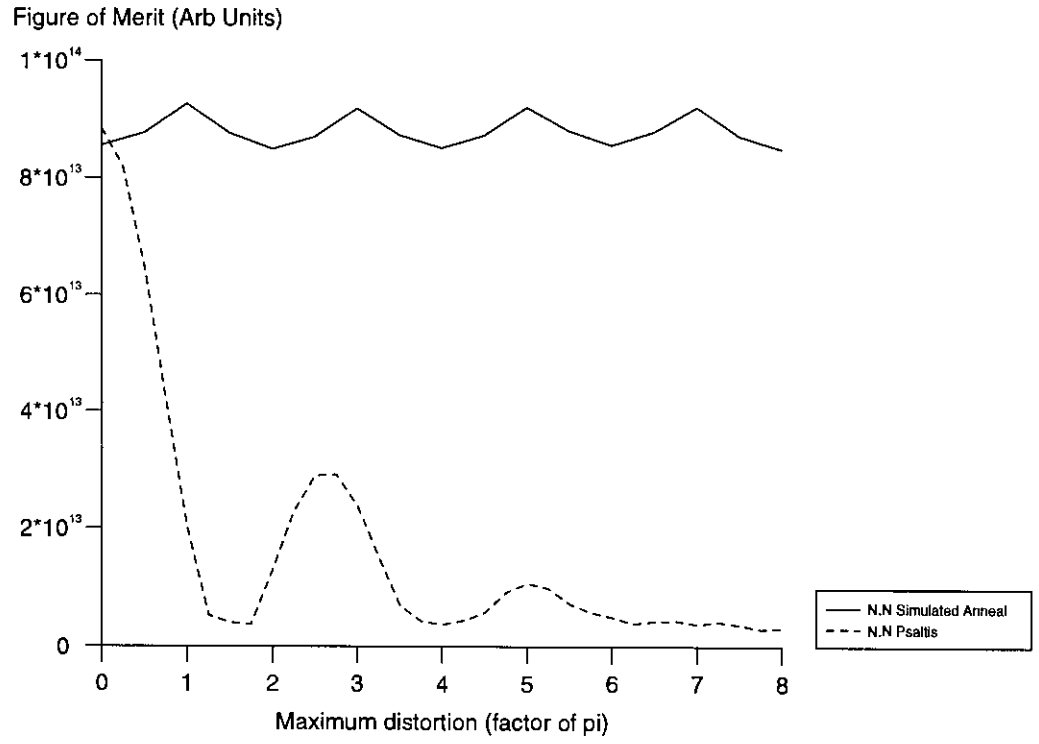


Figure 5.17. Performance of simulated annealling algorithm over Psaltis algorithm (N.N)

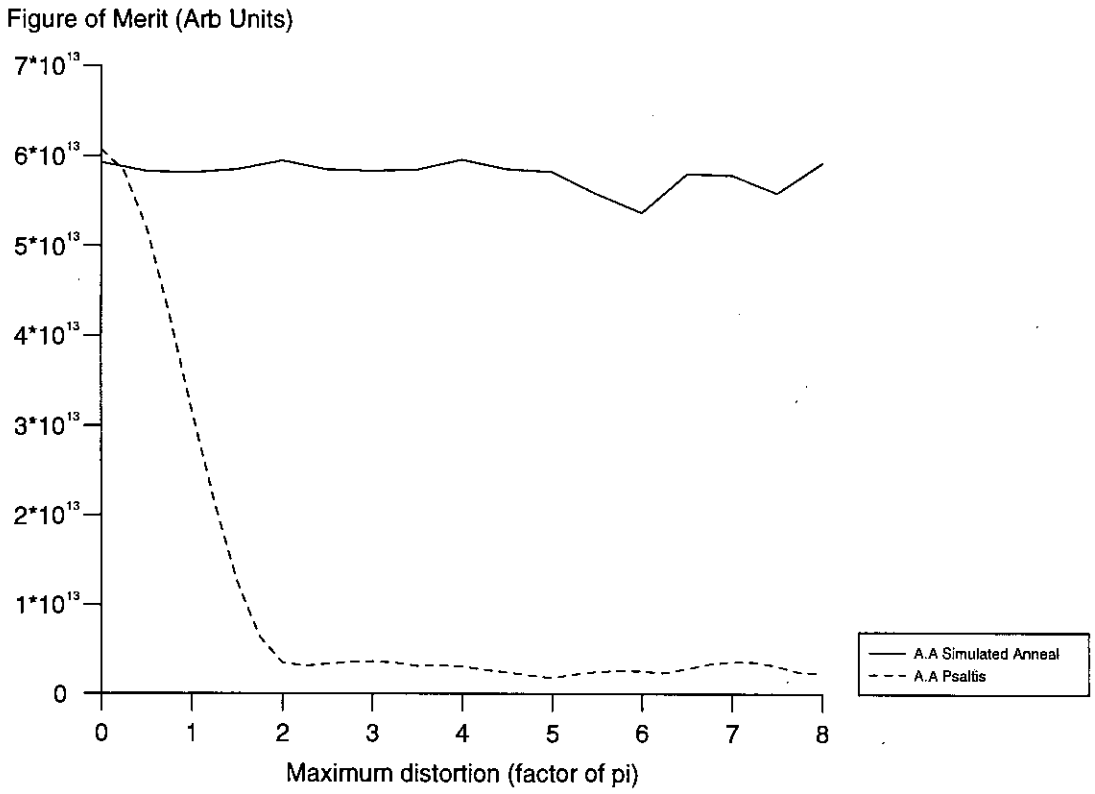


Figure 5.18. Performance of simulated annealling algorithm over Psaltis algorithm (A.A)

As sample output from these distorted systems, Figures 5.19 and 5.20 respectively show the simulated annealling output plane and the Psaltis output plane for a maximum distortion of n=4.

There are striking differences between the two figures: the simulated annealling algorithm has produced a sharp well-defined correlation peak whereas the static algorithm has degenerated into noise only, as the Fourier plane filter is completely mismatched to the input object Fourier plane phase spectrum due to the superimposed distortion..

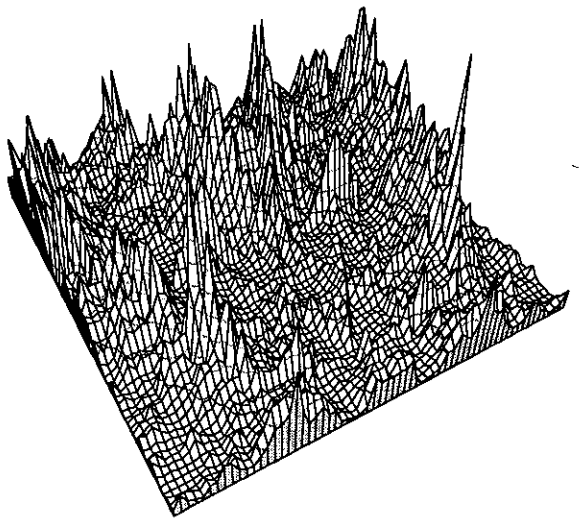


Figure 5.19 Autocorrelation of letter R with n=4 distortion in the input plane (simulated annealling).

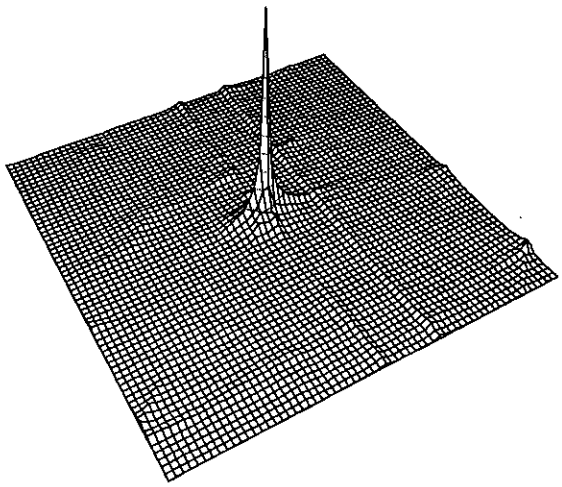


Figure 5.20 Autocorrelation of letter R with n=4 distortion in the input plane (Psaltis)

Repeating the above simulations with the linear wedge replaced by the quadratic, also generates interesting graphs: Figure 5.21 shows the effect of increasing n on the auto-correlation peak intensity (all letters), Figure 5.22 the effect on the cross correlations (R as target) and Figure 5.23 the ratio of the autocorrelation peak intensity to the cross-correlation intensity for the given objects (again using the R as the target and cross-correlating with the O, N and A). The results hold for the remaining targets also. A representative graph comparing the Psaltis static algorithm with the simulated annealling algorithm is presented as Figure 5.24 and shows once more how an adaptive technique surpasses the static form. The corresponding graphs

for the O, N and A are not presented: although results are available, this case is representative of the rest.

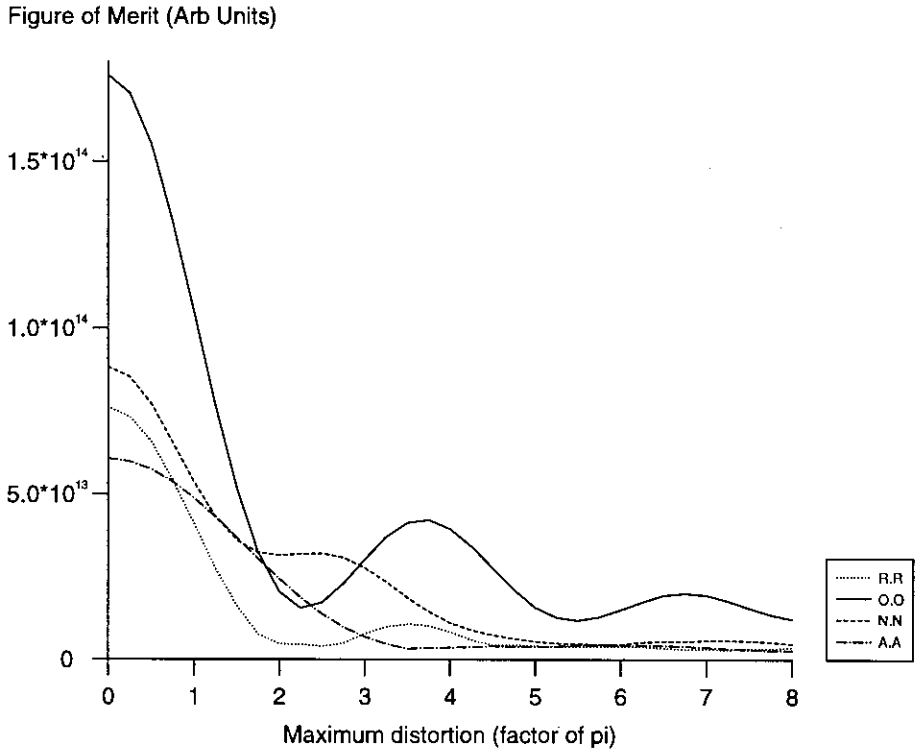


Figure 5.21. Central peak intensity in the output plane versus maximum phase distortion (n) in the input plane (letters R, O, N and A, auto-correlations, quadratic wedge).

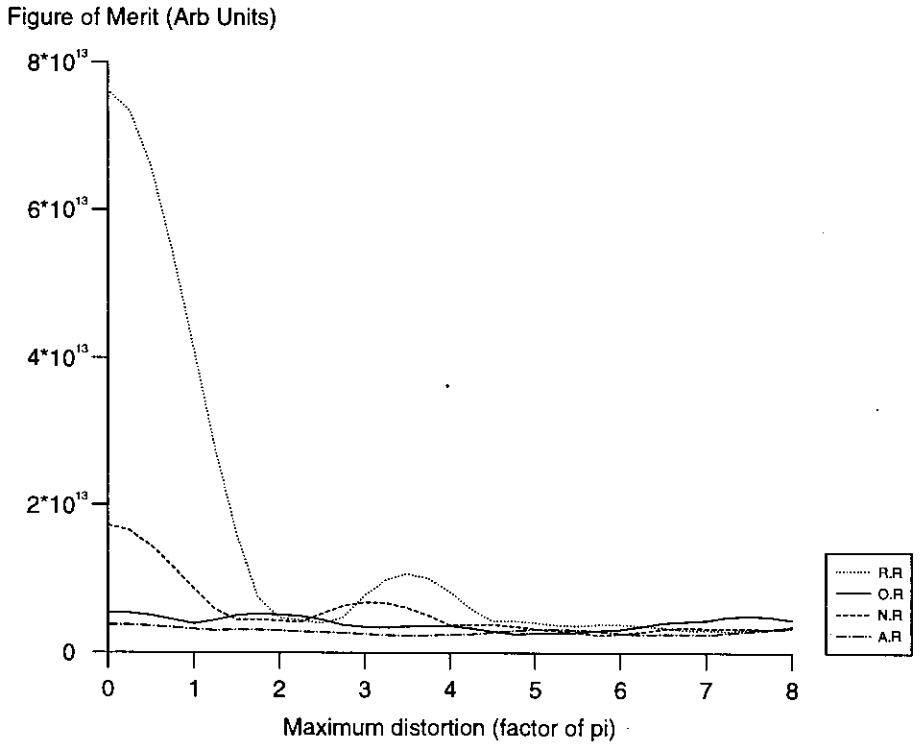


Figure 5.22. Cross-correlation central peak intensity in the output plane versus maximum phase distortion in the input plane (letter R, quadratic wedge).

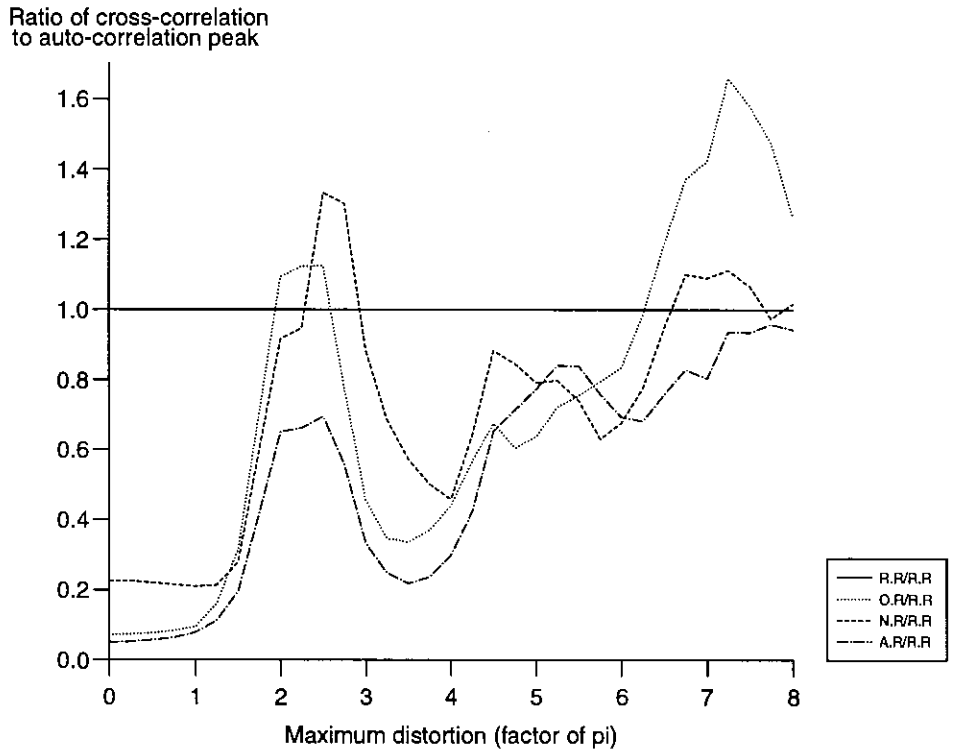


Figure 5.23. Ratio of cross-correlation peak intensity to auto-correlation peak intensity versus maximum phase distortion (n) in the input plane (letter R, quadratic wedge).

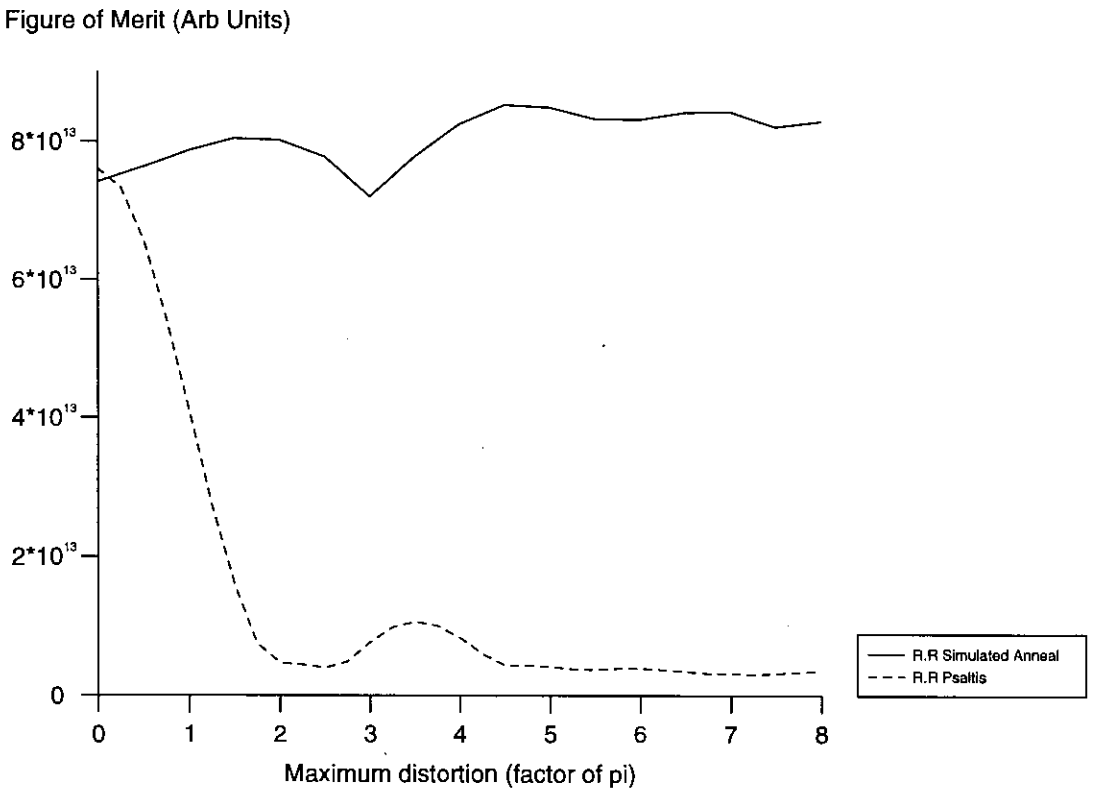


Figure 5.24. Comparison of autocorrelation performance of Psaltis and simulated annealling algorithms versus maximum phase distortion (n) in the input plane (letter R, quadratic wedge).

Phase distortions in the Fourier plane.

Simulations were carried out on the same objects but this time imposing the linear or quadratic phase wedges onto the Fourier plane filter. The binary filter acts, therefore, not as binary device but a (very) complex filter although still with uniform transmission. Sample graphs are shown in the following figures for the case of a linear phase wedge in the Fourier plane (Figure 5.25 - autocorrelation of data letters using Psaltis algorithm, Figure 5.26 - cross-correlation of the four letters with the filter corresponding to the letter R, Figure 5.27 displaying the ratios of cross-correlation to auto-correlation and Figure 5.28 comparing Psaltis performance to that of the simulated annealling algorithm). Figures for the quadratic phase distortion case are presented as 5.29 to 5.32 and these correspond to the same objects (respectively) as for figures 5.25 to 5.28.

Again, a periodic structure is noted. This time, as the physical effect is more predictable (a phase wedge in the Fourier plane producing a shift in the position of the output peak) it is more justifiable to assume that the correlation peak is being shared over two pixels. Additional evidence is clear in the 2π period which corresponds to a full pixel shift in the output plane. This is not an artificial effect - for a rigid detector it is expected that this effect will be observed on real optical systems as the correlation peak gradually shifts position.

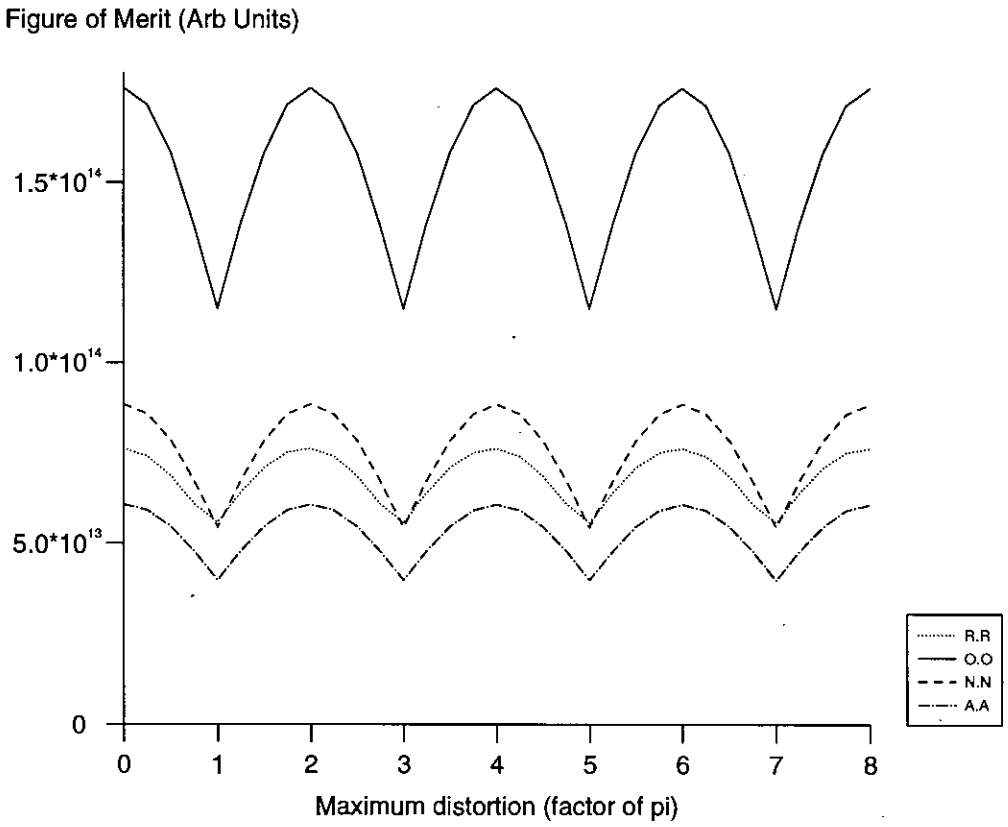


Figure 5.25. Linear phase wedge in Fourier plane. Autocorrelations using Psaltis algorithm

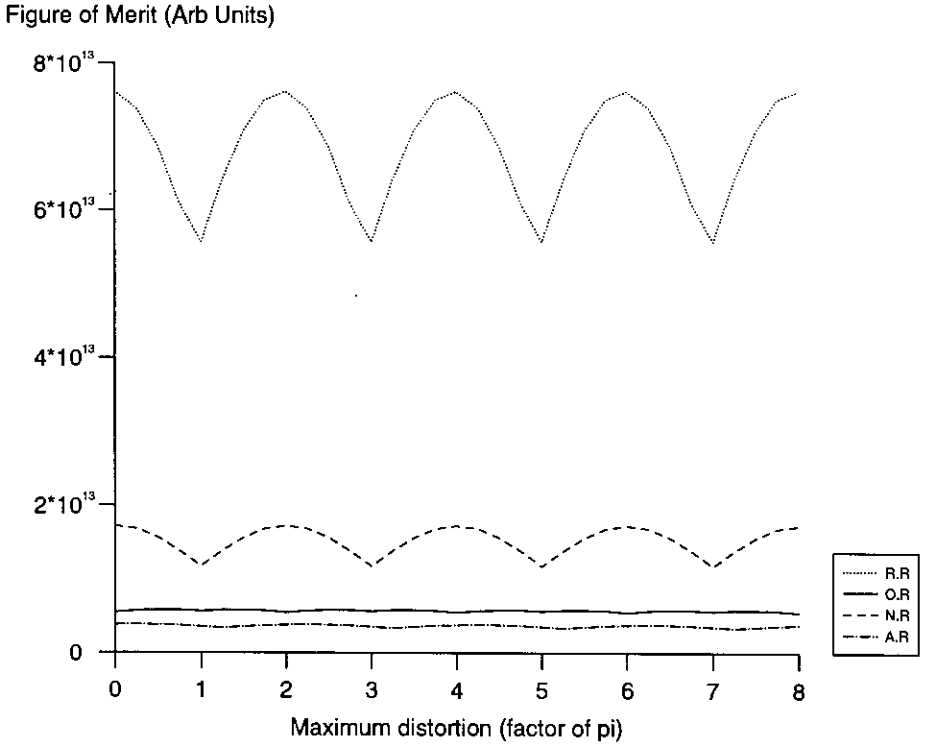


Figure 5.26. Linear phase wedge in Fourier plane. Cross-correlations using Psaltis algorithm. Letter R as target.

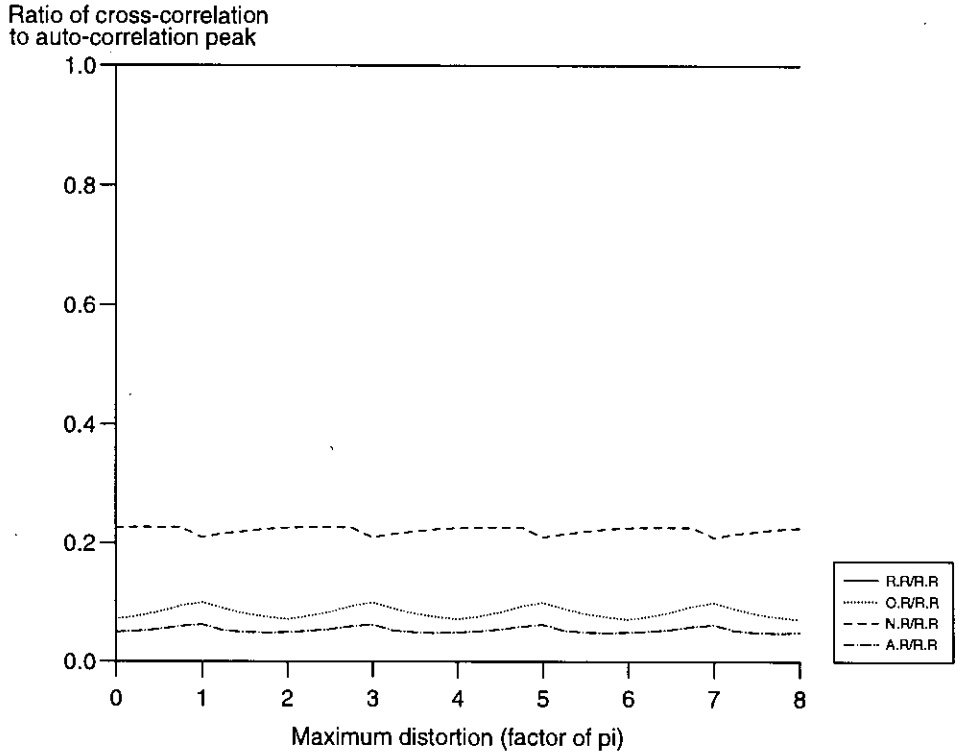


Figure 5.27. Linear phase wedge in Fourier plane. Ratios of auto-correlation to cross-correlations. Letter R as target.

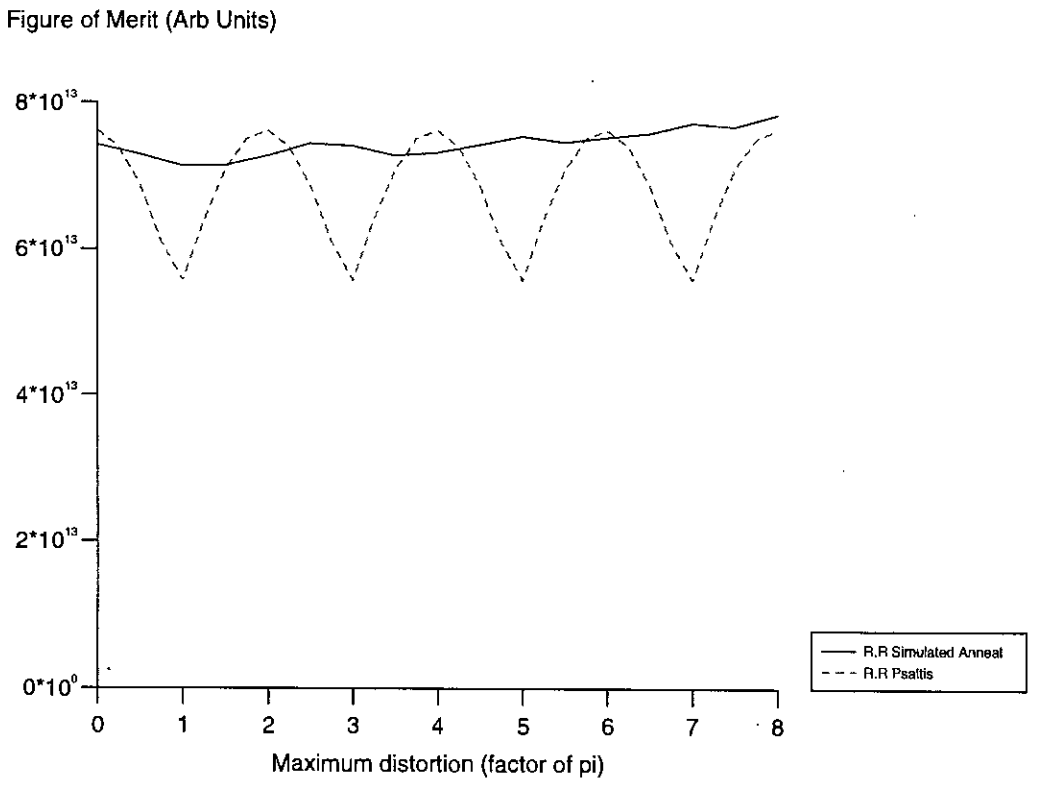


Figure 5.28. Linear phase wedge in Fourier plane. Autocorrelations of R comparing Psaltis algorithm with simulated annealling algorithm.

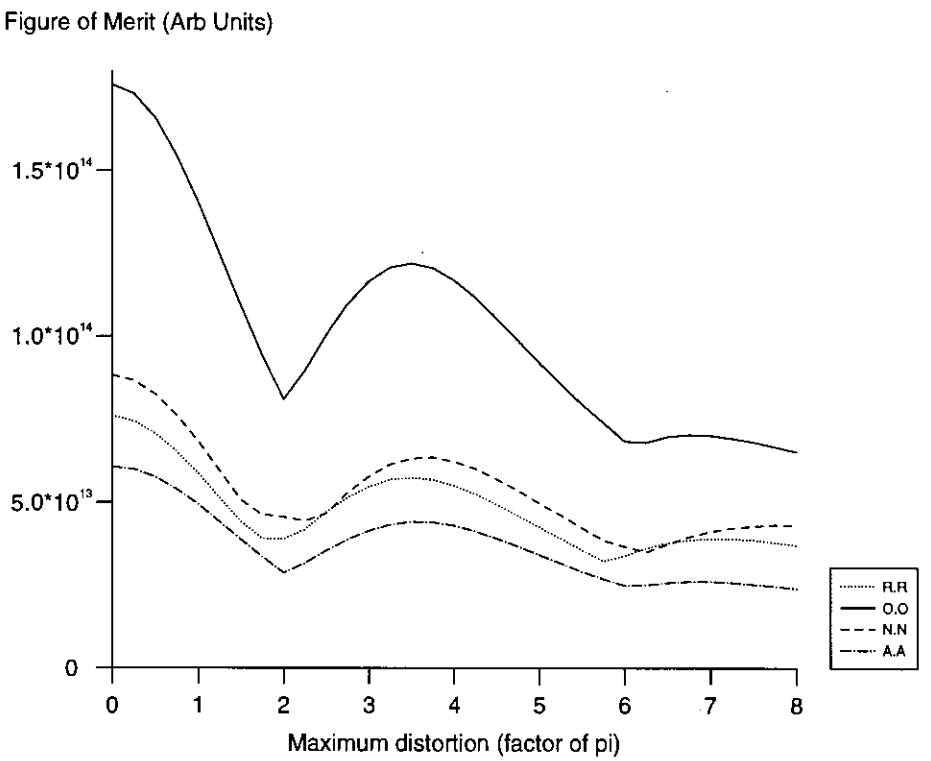


Figure 5.29. Quadratic phase wedge in Fourier plane. Autocorrelations using Psaltis algorithm

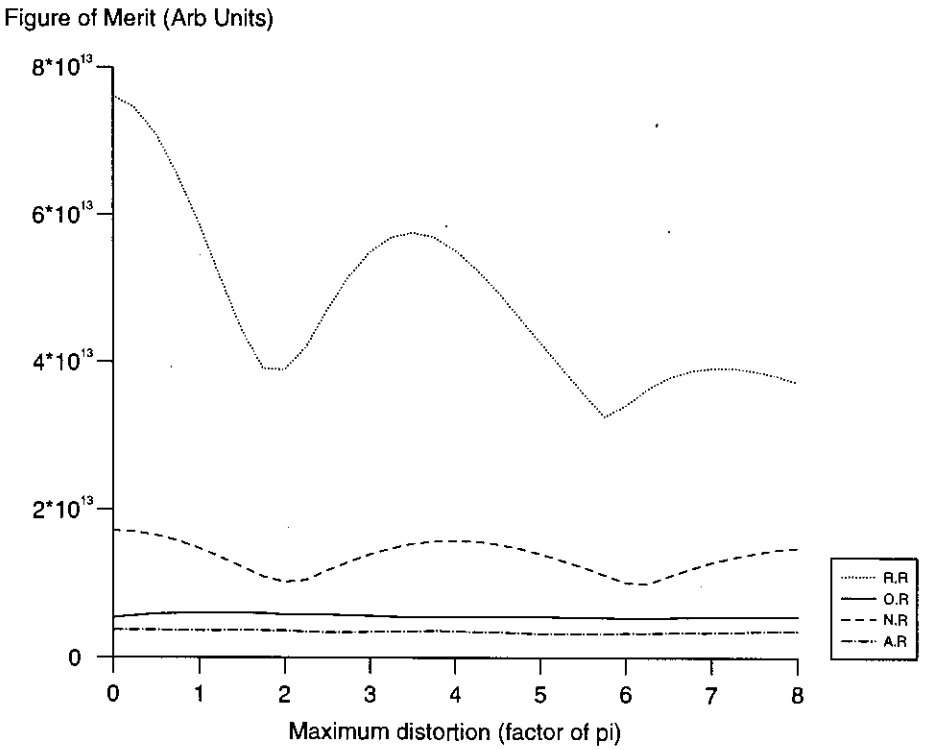


Figure 5.30. Quadratic phase wedge in Fourier plane. Cross-correlations using Psaltis algorithm. Letter R as target.

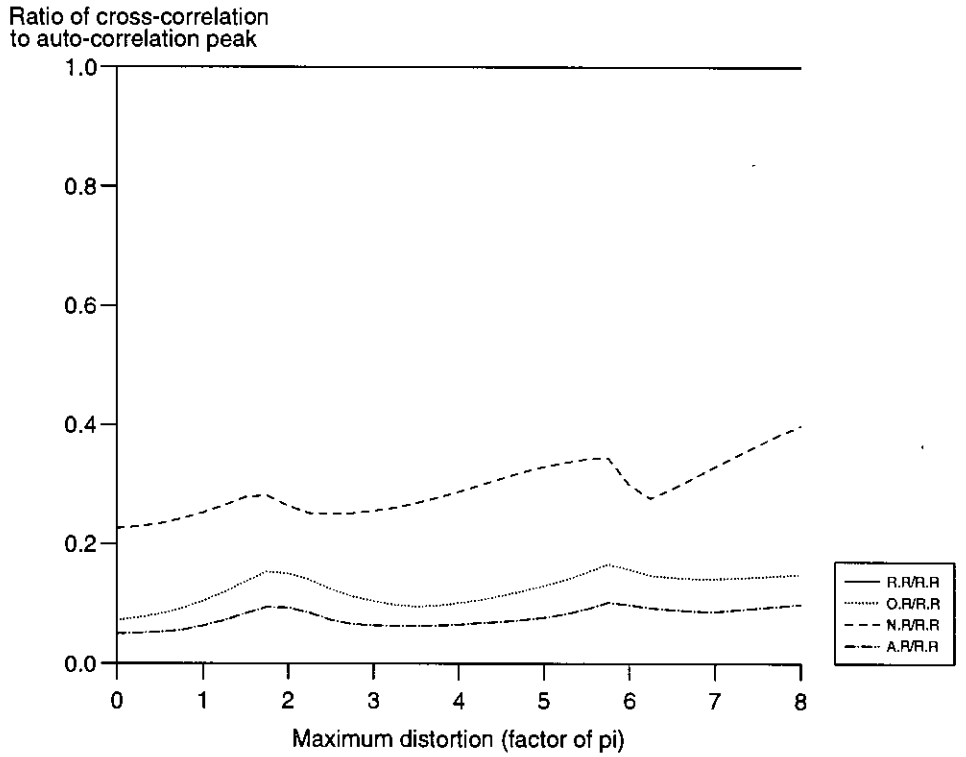


Figure 5.31. Quadratic phase wedge in Fourier plane. Ratios of cross-correlations to auto-correlation using Psaltis algorithm. Letter R as target.

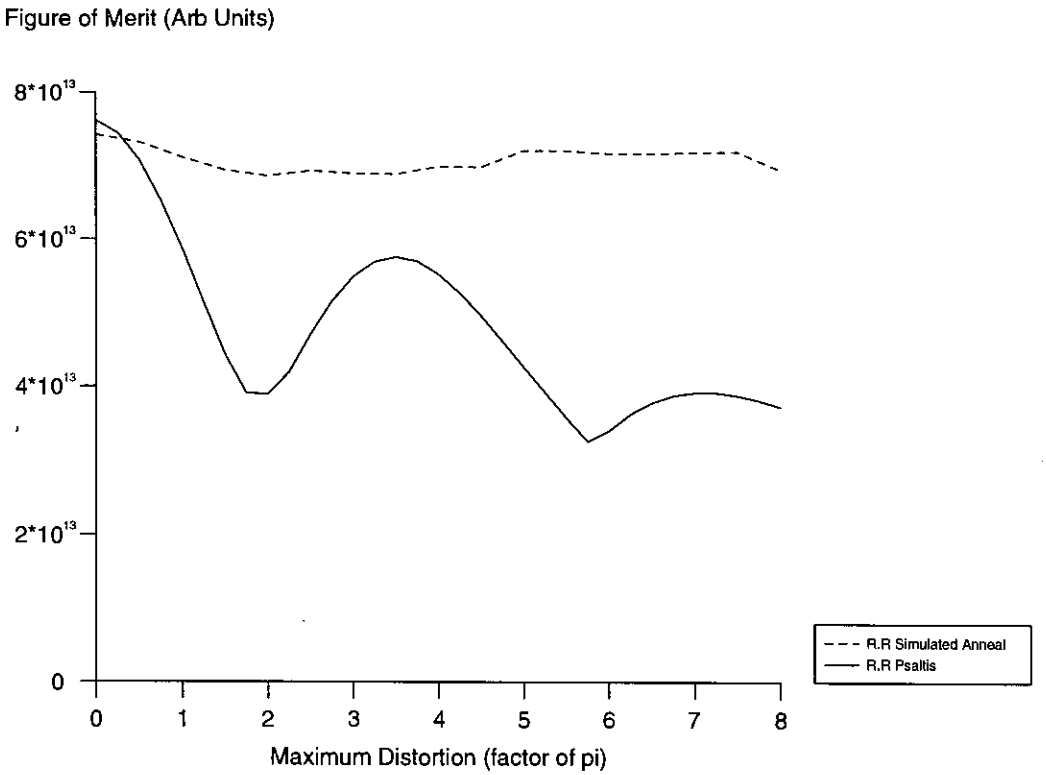


Figure 5.32. Quadratic phase wedge in Fourier plane. Autocorrelations of R comparing Psaltis algorithm with simulated annealling algorithm.

Interesting points to note from the linear phase distortions are the periodic pattern. This is due to the fact that a phase wedge in this plane corresponds to a shift in the energy distribution in the output plane. The dip in peak corelation energy is therefore at point where the peak has been shifted until it is 'shared' between two pixels in the output plane. As can be noted, a maximum phase distortion of 2π denotes a shift of a whole pixel in the output plane although the intensity of the peak remains high. The simulated annealling filter calculated for each phase distortion case again is shown to force the peak to be central and the figure used to create the graph in Figures 5.28 and 5.32 is the central (spot) peak intensity for the simulated annealing case.

For the quadratic phase wedge, the form of the graphs (Figures 5.29, 5.30) again appears to be periodic but is obviously much more complex than for the simple linear wedge. The peak again is shifted although not by such a simple relationship and the overall peak intensity reduces with higher distortions. Figure 5.34 shows that the annealling algorithm has once more removed most of the effects of the phase distortion, emphasizing its versatiliy in the phase of such optical imperfections and variations.

For the final case, the Fourier filter was taken as a binary phase filter binarising to $\phi=a\pi$ or 0 where a is real (standard binarising technique is therefore realised for a=1). This is therefore a uniform amplitude filter which is complex rather than entirely real. The graph in figure 5.33 below illustrates the effect of such a distortion, again using the Psaltis algorithm and the letter R as compared with the equivalent simulated annealling simulations. As can be seen, there is no appreciable difference in the performance of these algorithms under these conditions. Indeed, on comparison of the output SLM patterns generated through the annealling process it

could be seen that they were all extremely similar (i.e. this distortion had no solution through reprogramming the filter) and in fact only small improvements are detectable at the a=1.75, a=0.25 levels and none at all in between. It may therefore be concluded that (within the resolution limits of the series of simulations)the optimal binary phase-only filter MUST have a phase depth of π radians (i.e. on and off states of -1 and 1 respectively).

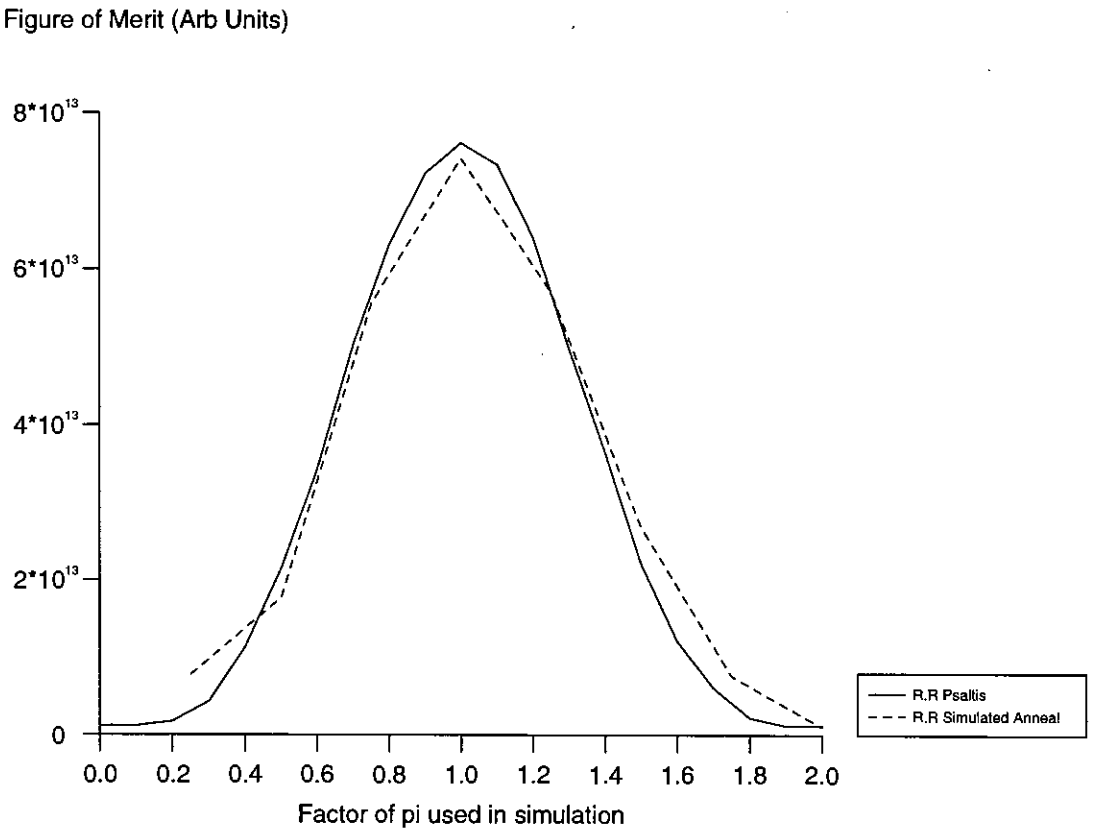


Figure 5.33. Autocorrelation of letter R with Fourier plane filter and additive phase factor.

Discussion and conclusions

It has been shown that the stochastic algorithm has far more flexibility as a generator of Fourier plane filters than the standard static forms. It is possible to

perform quite complex shaping of the energy distributions in the correlation plane and thereby improve the discriminatory powers of the pattern recognition device.

In addition, phase distortions of the types likely to be encountered in experimental work (both wedge and quadratic forms) have been forced on both object and Fourier planes, with the simulated annealling algorithm producing consistent central sharp peaks while the static algorithms have degenerated into mere noise or, in the case of the linear phase distortion in the Fourier plane, a shifted correlation peak (thereby giving misleading information as to the relative positions of the target filter and the object). The one case where no noticeable improvement is to be made is that of a constant phase addition to a single state of the SLM pixels. In this case, the generated pattern on the SLM remains essentially constant throughout. Calculations showed the difference between filters calculated for varying phase additions to be less than one percent different pixels.

A functioning electro-optical implementation of this architecture could not be manufactured using available materials within the proscribed period of this thesis for reasons described earlier but a putative design of such a correlator is illustrated in figure 5.34 below.

Essentially, the device is based on the classic matched filter correlator. The filter SLM (here a transmissive device) filters the Fourier plane with a programmable binary pattern and driven by suitable drive circuitry. The output (correlation) plane is coincident with an array of photodetectors whose orientation, position and geometry may be tailored according to the application of the correlator and whose output is processed by the output electronics. This corresponds to the ability in the computer simulation of choosing a particular set of pixels and

performing some calculation on the outputs of those pixels. The figure of merit is calculated and passed back to the SLM drive circuitry whereon a simple second calculation is made, comparing that value to the one calculated at the previous iteration whereupon a (randomly chosen) SLM pixel may be flipped or not according to improvements made or the 'temperature' of the anneal.

This type of arrangement is perfectly realisable with current microprocessor and SLM technology and with a device whose pixels may be addressed individually, the frame rate (and therefore iteration rate) may be extremely fast. Taking a frame rate (where only a single pixel is altered) of 10 kHz we may assume (using the SLM as the only limiting factor) that a typical iteration schedule for a 64x64 pixel array may be conducted in 4096 x 0.0001 or about half a second. This is clearly unrealistic - the limiting factor in any such device is likely to be the external thresholding and computational electronics but still this compares favourably with several hours to perform the same iteration on a computer workstation.

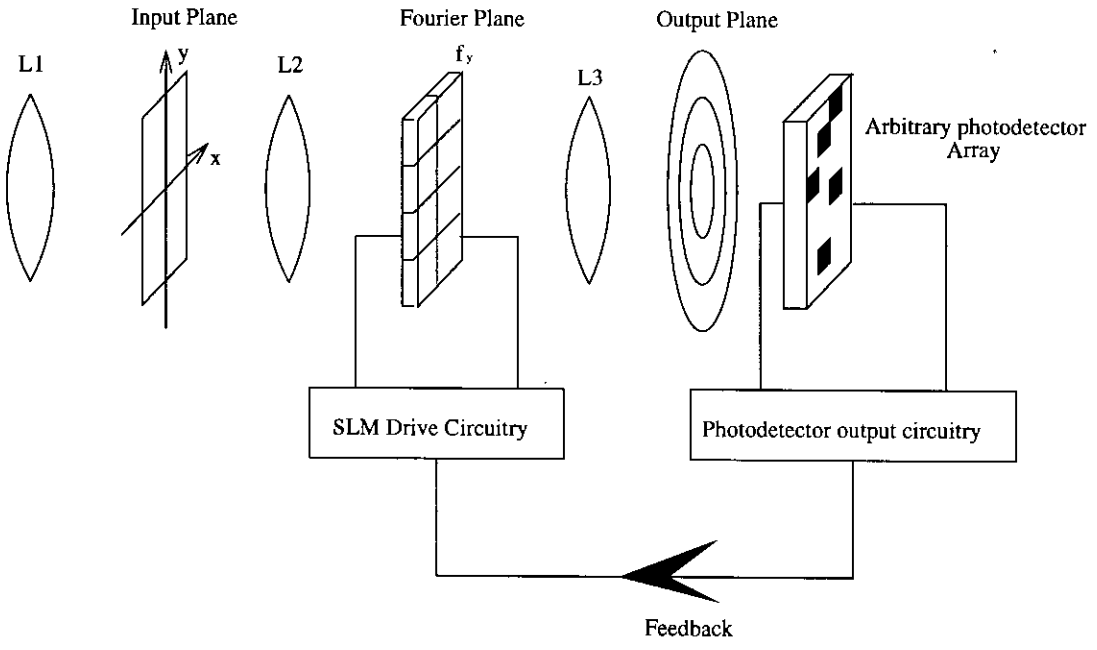


Figure 5.34 Putative electro-optical simulated annealling correlator.

Chapter 6

Conclusions and Future Work

Introduction

The contents and results presented in this thesis are summarised and presented in terms of the potential for future research. Although it would have been preferable for the sake of completeness to have presented relevant optical results alongside the theoretical results of computer simulations, it is felt that, in the light of many recent papers, that the conclusions made are no less valid. Many researchers have consistently proved that such simulations agree extremely well with subsequent optical implementations, even when using low space-bandwidth optical filters (e.g. Potter, (1990)), and so these results may be taken in the same context as those double checked through optical means.

The SLM devices studied were potentially of good enough quality and resolution to demonstrate the effects discussed in this thesis but the design faults in both the 128 by 128 SLM and an inability to locate a working 50 by 50 device together with the unfortunate timing of the SERC removal of base funding for the Edinburgh University microfabrication facility forced the abandonment of the practical thrust of this work. A putative architecture for the realization of the

concepts in this work was presented for future investigations, emphasising the exploitation of the near instantaneous computational powers of an optical system when conducting Fourier transforms (and hence correlations) of two-dimensional objects. The advantages of such an optical arrangement over computer techniques are immense as regards the iteration times and such devices are practical with current technologies.

All Optical Correlation

The number of groups researching optical correlation techniques is continually growing. This is one application which quite clearly both outperforms electronic implementations and also fully exploits the major capabilities of optical calculation through the instantaneous Fourier transforming property of a lens. Much of the research has been into static algorithms which have been shown to cater for specific classes of objects but these do not allow for flexible interpretation of the system by some "learning" algorithm.

Of these novel algorithms, one in particular stands out in terms of ease of use and its potential for an all optical implementation and that is the simulated annealing algorithm. It has been shown that this caters very favourably in simple optimisation cases with the existing algorithms but also that, through its flexibility it may be used for very complex tailoring of the correlation plane if that is what is required.

Some of the problems involved in research in this field are due to problems in obtaining high class filter devices to place in optical systems. The high quality devices are expensive, and the low cost end of the market produces optical

modulators with poor contrast and additive phase distortions, reducing the efficiency of the correlator and affecting the potential experimental results.

Simulated Annealing

Many examples have been generated and simulated through computer programs. In all cases, the algorithm functions as well as or better than, the contending algorithms, consistently producing high quality output and high object differentiation and no case has been found where the algorithm does not perform as expected.

The algorithm may be extended arbitrarily to more degrees of freedom (for instance the optimisation of ternary or quaternary level filters is a real application) and will automatically cater for all pixel shapes and sizes - indeed, those filters with complex pixel functions may have been underperforming through a lack of such a suitable optimising method.

Some examples of the tailoring of the correlation plane have been presented which aim to show the potential for such methods to improve existing correlator architectures and discriminatory systems especially when considering specific cases or target types. It is extremely interesting to see that the more exotic merit figures tested have shown the most potential for object discrimination, simply by exploiting the differences between object and target automatically. For future investigation, it would be interesting to examine whether by generating such a specific filter the system becomes completely intolerant to distortions of the object, whether through small rotations or scale variance (although this would be a disadvantage in, say, human face recognition where a smile may distort the image over much, it would be a positive filter in testing for similar rigid object shapes which may be constrained in

a known orientation). Although is should be possible to extend the simulated anneal to test for varied sizes of object all at once, this would be to the general detriment of the correlator performance.

The main finding of this work is the ability to "anneal" out imperfections in a typical Fourier plane filter. Chapter 5 shows this extremely clearly through simulations of a SLM with large phase distortions superimposed. Whilst the autocorrelations generated through other algorithms deteriorate rapidly to a background noise, or perhaps worse, shift the correlation peak away from the centre of the plane (for a centred object in the object plane) the simulated anneal removes most of the effects introduced and forces the correlation peak to be central, resulting in a correlation peak of a similar level and position to that expected from a perfect filter device. Phase distortions of several wavelengths may be annealed allowing substandard spatial light modulators or liquid crystal televisions to be utilized in an otherwise demanding architecture.

Phase distortions of two types were examined, the algorithm performing extremely well under the conditions of either object plane or Fourier plane distortion and either a linear phase distortion or a quadratic one. A third case (that of inaccurate representation of a 0 and π binary phase only device) was also simulated. It was interesting to note there that no particular improvements were discernable in the output of the simulated annealling simulation over the static algorithms. The conclusion here is that this problem is here to stay - the filter patter is optimised and remains (almost) unchanged when altering the 'on' phase and reannealling.

Future Work

The main goal of future work in this area should be to implement the architecture in an all optical system. The advantages as stated suggest that this would potentially improve the optical results immensely, allowing for imperfect devices to produce near perfect binary phase only correlations. Demonstrating the training aspect of the algorithm is seen as the strongest point in its favour and exploiting the superior speed of the Fourier transforming optics would aid the process immensely. It may have potential uses as an adaptive recognition system which can retrain itself constantly to a gradually changing object or environment.

There now exist devices which may modulate light in greyscale rather than binary scale and it would be interesting to investigate how these new devices may be optimised - the algorithm is flexible enough and easily extendable to cope with the extra degrees of freedom which would be encountered.

A third interesting possibility is to marry this field with that of the smart SLM design field (as discussed in chapter 2). The particular figure of merit used in the 'depletion zone' simulation will fit in well with intelligent photodetecting circuitry boards which could easily provide sufficient mathematical processing power to remove the need for a computer in the system. This then hints at potential reductions in size of the entire self-teaching correlator system, making it more practical and commercial.

BIBLIOGRAPHY

Armitage D, Thackera J I, Eades W D (1987,I): Ferroelectric liquid crystal devices. SPIE, v.825; 32-40

Armitage D, Thackera J I , Eades W D, Stiller M A and Anderson W W (1987,II): Fast nematic liquid crystal spatial light modulator. SPIE v.824 "Advances in Nonlinear polymers and inorganic crystals, liquid crystals and laser medic"; 34-.

Awwal A A S, Karim M A and Jahan S R (1990): Improved correlation discrimination using an amplitude-modulated phase-only filter. Applied Optics 29 (2); 233-236.

Barnes T H, Matsuda K and Ooyems N (1988): Reduction of false correlations with binary phase-only filters. Applied Optics 27 (18); 3785-3790.

Bartelt H and Horner J L (1985): Improving binary phase correlation filters using iterative techniques. Applied Optics 24 (18); 2894-2897.

Bracewell R N (1986): "The Fourier Transform and it's applications". Pub. McGraw-Hill, New York (2nd edition).

Brooks R E (1984): Micromechanical light modulators for data transfer and processing. SPIE, v.465 "Spatial Light Modulation and Applications"; 46-54.

Burns D C (1991): Personal Communication.

Burns D C (1994): Ph.D. thesis, University of Edinburgh.

Burns D C, Underwood I, Murray A F and Vass D G (1994): An Optoelectronic Neural network with Temporally Multiplexed Grey-scale weights.

Burns D C, Underwood I, Vass D G and O'Hara A (1994): Electronically addressed ferroelectric liquid crystal over silicon spatial light modulators for optical computing. Int Conf. on Optical Computing (technical digest); pp27-28

Campes J, Turon F, Yfnel M J and Vallintjans S (1990): Phase-only filters codified with Birckhardt's method. Applied Optics, 29 (35); 5232-5234.

Casasent D and Xia S-F (1986): Phase correction of light modulators. Optics Letters, 11 (6); 398-400.

Chiou A E, Yeh P (1990): Symmetry filters using optical correlation and convolution. Optical Engineering, 29 (9); 1065-1072.

Cognard J (1982): Alignment of nematic liquid crystals and their mixtures. Mol.Crystals Liquid Crystal Supplemental Series 1.

Cottrell D M, Davis J A, Schamschula M P and Lilly R A (1987): Multiplexing capabilities of the binary phase-only filter. Applied Optics 26 (5); 934-937.

Cottrell D M, Lilly R A, Davis J A and Day T (1987): Optical correlator performance of binary phase-only filters using Fourier and Hartley transforms. Applied Optics 26 3755.

Cutrona L J, Leith E N, Palermo C J and Porcello L J (1960): Optical data processing and filtering systems. IRE Transactions on information theory; 386-400.

Davis J A, Cottrell D M, Bach G W and Lilly R A (1989): Phase-encoded binary filters for optical pattern recognition. Applied Optics 28 (2); 258-261.

Davis J A, Waring M A, Bach G W, Lilly R A and Cottrell D M (1989): Compact optical correlator design. Applied Optics 28 (1); 10-11.

Davis J A, Connely S W, Bach G W, Lilly R A and Cottrell D M (1989): Programmable optical interconnections with large fan-out capability using the magneto-optic spatial light modulator. Optics Letters 14 (1); 102-104.

Davis J A, Heissenberger G M, Lilly R A, Cottrell D M and Brownell M F (1987): High efficiency optical reconstruction of binary phase-only filters using the Hughes LCLV. Applied Optics 26 (5); 929-933.

Dickey F M, Stalker F and Mason J J (1988): Bandwidth considerations for binary phase-only filters. Applied Optics 27 (18); 3811-3818.

Downie J D and Reid M B (1990): Mapping considerations for optimal binary correlation filters. Applied Optics 29 (35); 5235-5241.

Downie J D (1991): Design of optimal binary phase and amplitude filters for maximisation of correlation peak sharpness. Optics Letters 16 (7); 508-510.

Efron U, Owechko Y, Wu S T, Lacoe R C, Welkowsky M S, Bates T D and Hess L D (1985): The liquid crystal spatial light modulators: recent studies. SPIE, v.567 "Advances in materials for active optics"; 99-105.

Efron U, Wu S T, and Bates T D (1986): Nematic liquid crystals for spatial light modulators: recent studies. Journal of the Optical Society of America B 3 (2); 247-252.

Farn M W and Goodman J W (1988): Optimal binary phase-only matched filters. Applied Optics 27 (21); 4431-4437.

Farn M W and Goodman J W (1989): Optimal maximum correlation filter for arbitrarily constrained devices. Applied Optics 28 (15); 3363-3366.

Fielding K H and Horner J L (1990): 1-f binary joint-transform correlator. Optical Engineering 29 (9); 1081-1087.

Fisher A D and Lee J N (1986): The current status at two-dimensional spatial light modulator technology. SPIE v. 634 "Special Institute on hybrid and optical computing".

Flavin M A and Horner J L (1989): Amplitude encoded phase-only filters. Applied Optics 28 (9), 1692-1696.

Flannery D L, Biernacki A M, Loomis J S and Cartwright S L (1986): Real time coherent correlator using binary magnetooptic spatial light modulators at input and Farrier planes. Applied Optics 25 (4); 466.

Flannery D L, Loomis J S, Milkovich M E (1988, I): The use of binary magnetooptic spatial light modulators in pattern recognition processors. Spatial light modulators and applications, 1988 technical digest series 8; 221-224.

Flannery D L, Loomis J S, Milkovich M E (1988, II): Application of binary phase-only correlation to machine vision. Optical Engineering 27 (4); 309-320.

Flannery D L, Loomis J S and Milkovich M E and Keller P E (1988): Design elements of binary phase-only correlation filters. Applied Optics 27 (20); 4231-4235.

Gaskill J (1978): "Linear systems, Fourier Transforms and Optics". Pub Wiley, New York.

Geman S and Geman D (1984): Stochastic relaxation, Gibbs distributions and the Bayesian Restoration of images. IEEE Trans.-Pattern Analysis and Machine Intelligence PAMI - 6 (6); 721-741.

Goodman J W (1968): "Introduction to Fourier Optics". Pub. McGraw-Hill.

Goodman J W, Silvestri A M (1970): Some effects of Fourier domain phase quantization. IBM Journal of Research and Development, Sept; 478-484.

Gregory D A (1986): Real-time pattern recognition using a modified liquid crystal television in a coherent optical correlator. Applied Optics 25 (4); 467-469.

Gregory D A, Loudin J A and Yu F T S (1989): Illumination dependence of the joint transform correlation. Applied Optics 28 (15); 3288-3290.

Grinberg J, Jakobsen A, Bleha W, Miller L, Fraas L, Boswell D and Myer G (1975): A new real-time non-coherent to coherent light image converter - the hybrid field effect liquid crystal light valve. Optical Engineering 14; 217-225.

Haas G, Wöhler H, Fritsch M and Mlynski D A (1988): Polarizer model for liquid-crystal devices. Journal of the Optical Society of America A 5 (9); 1571-1575.

Hecht E and Zajac A (1974): "Optics". Pub. Addison-Wesley, Reading, Mass.

Heddle S B, Vass D G and Sillitto R M (1992): Reduction of aliasing in correlation using a pixellated spatial light modulator. SPIE 1772; 116-127

Heddle S B (1993): Optical correlation using pixellated spatial light modulators. Ph.D. Thesis University of Edinburgh.

Heddle S B and Sillitto R M (1993): Reduction and removal of replication in an optical processor by randomization of pixel positions in the Fourier plane filter. J. Modern Optics 40 (2); 299-313.

Herman S (1969): The effect of vignetting on the impulse response of coherent optical processors. Proc IEEE, March ;346-347 (letter).

Hornbeck L J (1988): A Bistable Deformable mirror device. Spatial light modulators and applications, 1988 technical digest series 8; 107-110.

Horner J L (1982): Light utilisation in optical correlators. Applied Optics 21; 4511

Horner J L (1988): Is phase correction required in SLM-based optical correlators? Applied Optics 27 (3); 436-438.

Horner J L and Bartelt H O (1985): Two-bit correlation. Applied Optics 24 (18); 2889-2893.

Horner J L and Gianino P D (1984): Phase-only matched filtering. Applied Optics 23 (6); 812-817.

Horner J L and Gianino P D (1984): Applying the phase-only filter concept to the SDF correlation filter. Proc. Soc. Photo-Opt. Instrum. Eng. 519; 70

Horner J L and Leger J R (1985): Pattern recognition with binary phase-only filters. Applied Optics 24 (5); 609-611.

Hossack, W J, Vass D G and Underwood I (1991): Fourier processing with binary Spatial Light Modulators. SPIE 1564; 697-702.

Hutzler P J S (1977): Spatial frequency filtering and its application to microscopy. Applied Optics 16 (8); 2264-2272.

Javidi B and Kuo C-J (1988): Joint transform image correlation using a binary spatial light modulator at the Fourier plane. Applied Optics 27 (4); 663-665.

Johnson K M, McKnight D J and Underwood I (1993): Smart spatial light modulators using liquid crystals on silicon. IEEE Journal of Quantum Electronics 29 (2); 699-710.

Johnson R V, Hecht D L, Sprague R A, Flores L N, Steinmetz D L and Turner W D (1983): Characteristics of the linear array total internal reflection (TIR)electrooptic spatial light modulator for optical information processing. Optical Engineering 22 (6); 665-674.

Kalestynski A and Smolinska B (1977): Spatial frequency sampling by phase modules as a method of generating multiple images. Applied Optics 16 (8); 2261-2263.

Kermisch D (1970): Image reconstruction from phase information only. Journal of the Optical Society of America 60 (1); 15-17.

Kim M S, Feldman M R, Guest C C (1989): Optimum encoding of binary phase-only filters with a simulated annealing algorithm. Optics Letters 14 (11); 545-547.

Kim M S, Guest C C (1990): Simulated annealing algorithm for binary phase-only filters in pattern classification. Applied Optics 29 (8); 1203-1208.

Kim M S, Guest C C (1990): Choice of phase in binary phase-only filters for optical pattern recognition. Applied Optics 29 (14); 1997.

Kim H M, Jeong J W, Kang M H and Jeong S I (1988): Phase correction of a spatial light modulator displaying a binary phase-only filter. Applied Optics 27 (20); 4167-4168.

Kirsch J C, Gregory D A, Thie M W and Jones B K (1992): Modulation Characteristics of the Epson liquid crystal television. Optical Engineering 31; 963-970.

Kirkpatrick S, Gelati C D (Jr), Vecchi M P (1983): Optimization by simulated annealling. Science 220 (4598); 671-680.

Kozaitis S P and Foor W E (1992): Optical correlation using reduced-resolution filters. Optical Engineering 31; 1929-1935.

Kumar B V K V (1989): Signal to noise ratio loss in correlators using real filters. Applied Optics 28 (15); 3287-3288.

Kumar B V K V, Bahri Z (1989): Phase-only filters with improved signal to noise ratio. Applied Optics 28 (2); 250-257.

Latham S G and Owen M P (1986): A silicon liquid-crystal spatial light modulator. GEC Journal of Research 4 (3); 219-222,

Leclerc L, Shang Y and Arsenault H H (1989): Rotation invariant phase-only and binary phase-only correlation. Applied Optics 28 (6); 1251-1256.

Lin H-K, Kung S Y and Davis J A (1986): Real-time optical associative retrieval technique. Optical Engineering 25 (7); 853-856.

Lindell S D and Flannery D L (1990): Experimental investigation of transform ratio ternary phase-amplitude filters for improved correlation discrimination. Optical Engineering 29 (9); 1044-1051.

Liu H-K and Chao L-H (1989): Liquid Crystal television spatial light modulators. Applied Optics 28; 4772-4780.

Lugi A V (1966): Operational notation for the analysis and synthesis of optical data-processing systems. Proc IEEE 54 (8); 1055-1063.

McKnight D J (1989): An electronically addressed spatial light modulator. Ph.D. Thesis - University of Edinburgh.

McKnight D J, Vass D G, Sillitto R M (1989): Development of a spatial light modulator: a randomly addressed liquid-crystal over Nmos array. Applied Optics 28 (22); 4757-4762.

Mahlab U, Shamir J (1989): Phase-only entropy-optimised filter generated by simulated annealling. Optics Letters 14 (21); 1168-1170.

Miller D A B (1990): Quantum-well self electro-optic effect devices. Optical and Quantum Electronics 22 (5); S61-S98.

Miller P C (1993): Optimum reduced resolution phase-only filters for extended target recognition. Optical Engineering 32 (11); 2890-2989.

Molley P A and Stalker K T (1990): Accousto-optic signal processing for real-time image processing. Optical Engineering 29 (9); 1073-1080.

Nomura T, Nagase T, Itoh K and Ichioka Y (1990): Designing a binary fourier phase-only correlation filter using a simulated annealing algorithm. Journal of Applied Physics, Part 2, 29 (11); L2074-L2076.

O'Hara A, Hannah J R, Underwood I, Vass D G and Holwill R J (1993): Mirror quality and efficiency improvements of reflective spatial light modulators by the use of dielectric coatings and chemical-mechanical polishing. Applied Optics 32 (28); 5549-5556.

Potter D J, Ranshaw M J, Al-Chalabi A O, Fancey N E Sillitto R M and Vass D G (1991): Optical correlation using a phase-only liquid crystal over silicon spatial light modulator. SPIE 1564 Optical Information processing Systems and Architectures III; 363-372.

Potter D J (1992): Phase-only optical information processing. Ph.D. Thesis University of Edinburgh.

Pratt W K, Kane J, Andrews H C (1969): Hadamard transform image coding. Proc. IEEE 57 (1); 58-69.

Psaltis D, Paek E G and Venkatesh S S (1984): Optical image correlation with a binary spatial light modulator. Optical Engineering 23 (6); 698-704.

Ranshaw M J (1988) Phase modulating spatial light modulators. Ph.D. Thesis - University of Edinburgh.

Rogers S K, Cline J D, Kabinsky M, Mills J P (1990): New binarisation technique for joint transform correlation. Optical Engineering 29 (9); 1088-1093.

Ross W E, Psaltis D and Anderson R H (1985): Two-dimensional magneto-optic spatial light modulator for signal processing. Optical Engineering 22; 485-490.

Rosvold G O (1990): Fast measurements of phase using a PC-based frame grabber and phase stepping technique. Applied Optics 29 (2); 237-241.

Samus S (1994): Personal communication.

Smith R (1995): Personal communication.

Spanier J and Oldham K B (1987): "An Atlas of Functions". Pub. Hemisphere Publishing Corporation.

Tam E C, Yu F T S, Tanano A, Gregory D A and Juday R D (1990): Data association multiple target tracking using a phase-mostly liquid crystal television. Optical Engineering 29 (9); 1114-1121.

Turner R, Johnson K M, Jared D A (1991): High speed compact correlator design using VLSI/FLC SLMs. Unpublished.

Underwood I (1987) An nMOS addressed liquid crystal spatial light modulator. Ph.D. Thesis - University of Edinburgh, 1987.

Underwood I, Vass D G V, Sillitto R M, Bradford G, Fancey N and Al Chalabi A O (1991): A high performance spatial light modulator. Proc. SPIE 1562,107-115.

Underwood I, Vass D G, O'Hara A, Burns D C, McOwan P W and Gourlay J (1994): Improving the performance of liquid-crystal-over-silicon spatial light modulators: issues and acheivements. Applied Optics 33 (14); 2768-2774

Vander Lugt A (1964) : Signal detection by complex spatial filtering. IEEE Trans. Inf. Theory IT-10; 139-145.

Vander Lugt (1974) : Coherent Optical Processing. Proc IEEE 62; 1300-1319.

Walsh T R and Giles M K (1990): Statistical filtering of time-sequenced peak correlation responses for distortion invariant recognition of multiple input objects. Optical Engineering 29; 1052.

Weaver C S and Goodman J W (1966): Technique for optically convolving two functions. Applied Optics 5; 1248-1249

Weigelt J (1987): Binary logic by spatial filtering. Optical Engineering 26; 28-33.

Welkowsky M S, Efron U, Byles W, Goodwin N W (1987): Status of the Hughes charge-coupled-device-addressed LCLV. Optical Engineering 26 (5); 414-417.

Wherrett B S, Tooley F A P (editors) (1989): Optical Computing, Proceedings of the 34th Scottish Universities Summer School in Physics. SUSSP Publications, Edinburgh.

Wilson J and Hawkes J F B (1983): "Optoelectronics: An introduction". Pub. Prentice Hall International Inc.

Wilson R (1993): "Mirror, mirror on the wall". Electronics Weekly Nov 1993, p16.

Wu S T (1985): Towards high speed modulation by nematic liquid crystals. SPIE v.567 "Advances in materials for active optics".

Yu F T S, Lu X and Cao M (1984): Application of a magnetooptic SLM to white-light optical processing. Applied Optics 23 (22); 4100-4104.

Yu F T S, Song Q W, Cheng Y S and Gregory D A (1990): Comparison of detection efficiencies for VanderLugt and joint transform correlators. Applied Optics 29 (2); 225-232.

Appendix 1

Lens as a Fourier Transforming Instrument

1a Derivation

The Fourier transform, $\mathcal{F}$ of a complex function, f of two independant variables, x and y, is given by:

$$\mathcal{F}(f(x,y)) = F(f_x, f_y)$$

$$= \int_{-\infty}^{+\infty}\int f(x,y)\ \exp[-j2\pi(f_x x + f_y y)]\ dx\ dy \qquad \text{equation A1.1}$$

The inverse Fourier transform, f(x, y) of $\mathcal{F}(f_x, f_y)$ is given by:

$$f(x,\ y) = \int_{-\infty}^{+\infty}\int F(f_x,\ f_y)\ \exp[+j2\pi(f_x x + f_y y)]\ df_x\ df_y \qquad \text{equation A1.2}$$

Start (1982) and Goodman (1968) provide an excellent introduction to Fourier optics and include a rigorous derivation of the Fourier transforming property of a single lens. A simple summary of Goodmans work is detailed below.

Consider the situation of Figure A1.1 below.

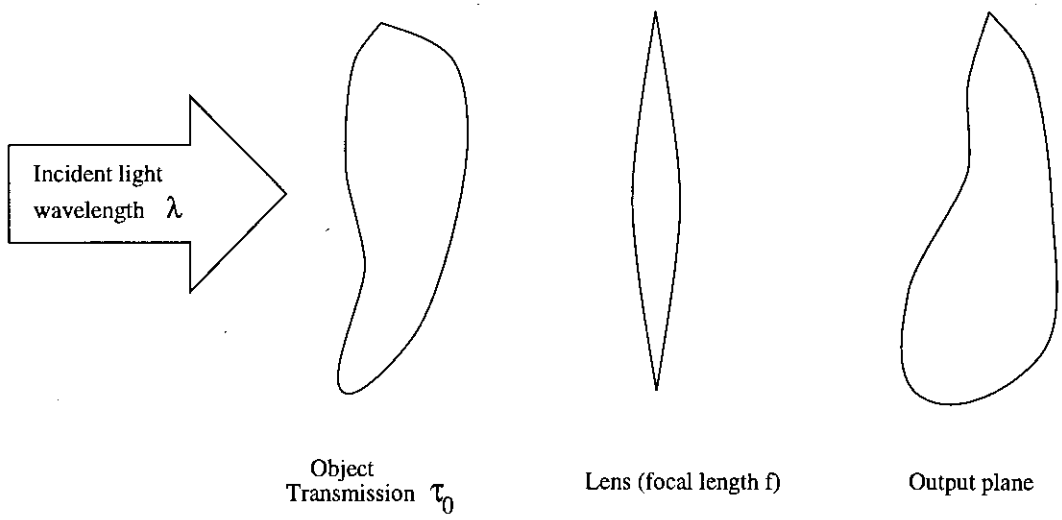


Fig A1.1 - Generic single lens optical system.

A plane wave of amplitude $\mathcal{A}$ is incident on an object and so, the light amplitude transmitted by the object is given by $\mathcal{A}\tau_0$ where $\tau_0\ (x_0,\ y_0)$ is the transmission function of the object. From this we can define the following:

The Fourier spectrum of the light transmitted by the object is:

$$\mathcal{F}(\mathcal{A}\tau_0) = F_0(f_x, f_y)$$

Let k represent the wavenumber $\frac{2\pi}{\lambda}$ where λ is the wavelength of the light. Defining $H(f_x, f_y)$ as a transfer function describing the effect of the propagation of a wavefront over a distance z (for Fresnel diffraction) - and in this case, the distance z is d, defined as the distance between the object and the lens - then the Fourier spectrum of the light incident at the plane of the lens is:

$$\begin{aligned}\mathcal{F}(\mathcal{U}_L) &= F_L(f_x, f_y)\\ &= H(f_x, f_y).\mathcal{F}(\mathcal{A}\tau_0)\\ &= \exp(jkd).\exp(-j\pi\lambda d\,(f_x^2 + f_y^2)).F_0(f_x, f_y)\\ &\propto \exp(-j\pi\lambda d.(f_x^2 + f_y^2)).F_0(f_x, f_y),\ \text{as } \exp(jkd) \text{ represents a}\end{aligned}$$

constant phase factor.

The field just behind the lens may be represented as $\mathcal{F}(\mathcal{U}_L).\tau_L$, where

$$\tau_L(x, y) = \exp(jkn\Delta_0).\exp\left[-j.\left(\frac{k}{2f}\right)(x^2 + y^2)\right],$$

using Goodman's derivation of the phase changes undergone by a plane wave as it traverses an ideal lens, with n being the refractive index of the lens and Δ_0 being the maximum thickness of the lens. Now, $\mathcal{U}'_L(x,y)$ can be defined as:

$$\mathcal{U}'_L(x,y) = \exp(jkn\Delta_0).\exp\left(-j\left(\frac{k}{2f}\right)(x^2+y^2)\right).\mathcal{U}_L(x,y)$$

Fresnel's diffraction formula for light propagating from (x_1,y_1,z_1) to (x_2,y_2,z_2) is given by:

$$\mathcal{U}(x_2,y_2) = \left(\frac{\exp(jk\Delta_z)}{j\lambda\Delta_z}\right).\exp\left(\left(\frac{jk}{2\Delta_z}\right)(x_2^2+y_2^2)\right)$$
$$\times\int_{-\infty}^{\infty}\int\mathcal{U}_1(x_1,y_1).\exp\left(\left(\frac{jk}{2\Delta_z}\right)(x_1^2+y_1^2)\right).\exp\left(-j\left(\frac{2\pi}{\lambda\Delta_z}\right)(x_1x_2+y_1y_2)\right)dx_1dy_1$$

where $\Delta_z = z_2 - z_1$, and for light propagating from the lens to its back focal plane, $\Delta_z = f$ where f is the focal length of the lens in question. From this the field at the output plane, $\mathcal{U}_f(x_f,y_f)$ is given by:

$$\mathcal{U}_f(x_f,y_f)=\frac{\exp(jkf)}{j\lambda f}.\exp\left(\left(\frac{jk}{2f}\right)\left(x_f^2+y_f^2\right)\right)$$

$$\times\int_{-\infty}^{\infty}\int\mathcal{U}'(x,y)_L.\exp\left(\frac{jk}{2f}\left(x^2+y^2\right)\right).\exp\left(-j\frac{2\pi}{\lambda f}\left(xx_f+yy_f\right)\right)dxdy$$

$$=\left(\frac{\exp\left(\frac{jk}{2f}\right)\left(x_f^2+y_f^2\right)}{j\lambda f}\right)\int_{-\infty}^{\infty}\int\mathcal{U}_L(x,y).\exp\left(-\frac{j2\pi}{\lambda f}\left(xx_f+yy_f\right)\right)dxdy$$

$$=\frac{\exp\left(\frac{jk}{2f}\left(x_f^2+y_f^2\right)\right)}{j\lambda f}F_L\left(\frac{x_f}{\lambda f},\frac{y_f}{\lambda f}\right),\text{ using the substitution from eqn A1.1}$$

$$\text{Now } F_L(f_x,f_y)=F_0(f_x,f_y).\exp\left(-j\pi\lambda d\left(f_x^2+f_y^2\right)\right)$$

$$=\mathcal{F}\left(\mathcal{A}\tau_0\right).\exp\left(-j\pi\lambda d\left(f_x^2+f_y^2\right)\right)$$

$$\text{so } U_f(x_f,y_f)=C\frac{\exp\left(\frac{jk}{2f}\left(x_f^2+y_f^2\right)\left(1-\frac{d}{f}\right)\right)}{j\lambda f}$$

$$\times\int_{-\infty}^{\infty}\int\tau_o(x_o,y_o).\exp\left(-\frac{j2\pi}{\lambda f}\left(x_ox_f+y_oy_f\right)\right)dx_ody_o$$

where C is a constant and d still represents the distance between the object and the lens.

But for the phase factor before the double integral, this is a perfect forward Fourier transform. The phase factor imposes a spherical wavefront on the output plane, but for the special case of d=f, the phase factor goes to 0 and we have a perfect planar Fourier transform, with the variables in the transform scaled such that:

$$f_x=\frac{x_f}{\lambda f}\quad\text{equation A1.3(a)}$$

$$f_y=\frac{y_f}{\lambda f}\quad\text{equation A1.3(b)}$$

This implies that doubling the focal length of the Fourier transforming lens will double the scaling of the Fourier transform (in the Fourier plane) and that this fact may be used to scale the Fourier plane to fit any chosen filter device.

1b Key Relationships

Convolution

The convolution $g(x,y)$ of two functions $f_1(x,y)$ and $f_2(x,y)$ is defined as

$$g(x,y) = \int_{-\infty}^{\infty}\int f_1(\alpha,\beta).f_2(x-\alpha,y-\beta)\ d\alpha d\beta$$

and this is commonly abbreviated to

$$g(x,y) = f_1(x,y)*f_2(x,y)$$

A mathematical property of the convolution formula is:

$$\mathcal{F}(f_1(x,y)*f_2(x,y)) = F_1(f_x,f_y).F_2(f_x,f_y)$$

i.e. theFourier transform of a convolution of functions ie equal to the product of the Fourier transforms of the individual functions ($\mathcal{F}$ denotes the Fourier transforming operator and F_1 and F_2 are the Fourier transforms of $f_1()$ and $f_2()$ respectively.

Correlation

The cross-correlation, $X_{f_1f_2}(x,y)$ is calculated as:

$$X_{f_1f_2}(x,y) = \int_{-\infty}^{\infty}\int f_1(\alpha+x,\beta+y).f_2^*(\alpha,\beta)\ d\alpha d\beta$$

(where * denotes a complex conjugate) and this also has a shorthand notation:

$$X_{f_1f_2}(x,y) = f_1(x,y)\otimes f_2^*(x,y)$$

and the corresponding relationship for the Fourier transform of the correlation is:

$$\mathcal{F}(f_1(x,y)\otimes f_2^*(x,y)) = F_1(f_x,f_y).F_2^*(f_x,f_y)$$

For functions such that $f_1() = f_2()$ the correlation is called the autocorrelation of $f_1()$.

Appendix 2

Optimization by Kangaroo

[sourced from the Internet]

Training a network is a form of numerical optimization, which can be likened to a kangaroo searching for the top of Mt. Everest. Everest is the global optimum, but the top of any other really high mountain such as K2 would be nearly as good. We're talking about maximization now, while neural networks are usually discussed in terms of minimization, but if you multiply everything by -1 it works out the same.

Initial weights are usually chosen randomly, which means that the kangaroo may start out anywhere in Asia. If you know something about the scales of the inputs, you may be able to get the kangaroo to start near the Himalayas. However, if you make a really stupid choice of distributions for the random initial weights, or if you have really bad luck, the kangaroo may start in South America.

With Newton-type (2nd order) algorithms, the Himalayas are covered with a dense fog, and the kangaroo can only see a little way around his location. Judging from the local terrain, the kangaroo make a guess about where the top of the

mountain is, and tries to jump all the way there. In a stabilized Newton algorithm, the kangaroo has an altimeter, and if the jump takes him to a lower point, he backs up to where he was and takes a shorter jump. If the algorithm isn't stabilized, the kangaroo may mistakenly jump to Shanghai and get served for dinner in a Chinese restaurant. (I never claimed this analogy was realistic.) In steepest ascent with line search, the fog is VERY dense, and the kangaroo can only tell which direction leads up. The kangaroo hops in this direction until the terrain starts going down again, then chooses another direction.

In standard backprop or stochastic approximation, the kangaroo is blind and has to feel around on the ground to make a guess about which way is up. He may be fooled by rough terrain unless you use batch training. If the kangaroo ever gets near the peak, he may jump back and forth across the peak without ever landing on the peak. If you use a decaying step size, the kangaroo gets tired and makes smaller and smaller hops, so if he ever gets near the peak he has a better chance of actually landing on it before the Himalayas erode away. In backprop with momentum, the kangaroo has poor traction and can't make sharp turns. I have been unable to devise a kangaroo analogy for cascade correlation.

Notice that in all the methods discussed so far, the kangaroo can hope at best to find the top of a mountain close to where he starts. There's a very high mountain. Various methods are used to try to find the actual global optimum.

In simulated annealing, the kangaroo is drunk and hops around randomly for a long time. However, he gradually sobers up and tends to hop up hill.

In genetic algorithms, there are lots of kangaroos that are parachuted into the Himalayas (if the pilot didn't get lost) at random places. These kangaroos do not know that they are supposed to be looking for the top of Mt. Everest. However, every few years, you shoot the kangaroos at low altitudes and hope the ones that are left will be fruitful and multiply.