Precision Measurement of Microscope Images

by

Christopher Paul Kirk

Submitted in accordance with the requirements
for the degree of Doctor of Philosophy.

The University of Leeds
Department of Electrical and Electronic Engineering

September 1985

Abstract.

The problems of measuring the dimensions of small geometries using an optical microscope are investigated, with particular attention to the measurement of the critical linewidths on semiconductor integrated circuit wafers and photomasks.

Conventional scalar diffraction models are used to investigate the imaging process of the optical microscope and these are extended to include image transducers and measurement devices. Particular attention is paid to the use of a video camera as an image transducer and an image shearing module as a measurement device.

The performance of different measurement techniques is investigated both theoretically and experimentally and systematic errors in different measurement techniques are identified. The limits of scalar diffraction theory are investigated experimentally.

The design of digital array filters is investigated in order to develop algorithms for automating the location of the feature to be measured, focusing and the measurement process itself. A novel method of automating the image shearing measurement technique using array filters is presented.

The models for the optical imaging, transducer response and automation algorithms are used to develop an automated image shearing based measurement system for measuring the gaps in magnetic recording heads. The design of the system is described and experimental performance tests demonstrate good agreement with the predictions of the theoretical system models.

The problem of modelling the images formed by thick layer objects is considered and a waveguide model is developed. Experimental and theoretical tests of the model show that the image profiles of shaped, multi-layer objects can be successfully predicted.

The model is used to investigate the imaging of thick layer objects in order to study the performance of different linewidth measurement techniques. A novel method of improving the repeatability of measurements on thick layers is presented, based on a contrast correction technique.

Abbreviations and Symbols.

Abbreviations.

CSS	Coincidence setting shear.
FIR	Finite impulse response.
ISM	Image shearing module.
NA	Numerical aperture.
NA_o	Objective numerical aperture.
NA_c	Condenser numerical aperture.
NBS	National Bureau of Standards (USA).
NPL	National Physical Laboratory (UK).
PSF	Point spread function.
SEM	Scanning electron microscope.

Symbols.

b	spatial frequency.
c	velocity of light.
i	$\sqrt{-1}$
k	wave number ($2\pi/\lambda$)
n_o	refractive index of air.
n_s	refractive index of substrate.
n_x	refractive index of material x.
R	relative reflectivity.
s	coherence parameter.
z	optical axis.
γ	photometric non-linearity.
ε_o	permittivity of free space.
ε_r	relative permittivity.
λ	wavelength.
μ_o	permeability of free space.
σ	resolution width or standard deviation.
ω	angular frequency.

All other symbols are defined as they appear in the text or are assumed to have their normally accepted meaning.

Acknowledgements.

To my Parents.

Contents

CHAPTER ONE

Introduction.

1.1 The need for optical micrometrology.

For many years, the optical microscope has been used extensively for measuring the dimensions of small objects. Probably the first serious microscopy work was "Micrography" published by Robert Hooke in 1665, and later in 1676 Antony van Leeuwenhoek discovered bacteria using an optical microscope.

In order to fully classify the objects under the microscope it was necessary to be able to measure their dimensions. Coarse measurements could be made by placing a graduated scale in the image plane of the objective. Measurements were made by moving the object to lie on the scale and then reading off its length. Despite its relative simplicity, it is possible to make quite accurate measurements on simple artefacts and measurements to within 1μm of the true value are possible.

An improvement to the setting precision was achieved by using one or more adjustable cross-hairs in the image plane. A micrometer adjustment of a fine cross-hair enables some simple objects to be measured to within about 0.1μm of the true length.

The precision with which measurements can be made improves as the magnification increases, however it was soon realised that there was a limit to this. The resolution of the optical microscope was found to be limited by diffraction and this was modelled by Fraunhofer using the Huygens-Fresnel wave theory of light. In 1834, Sir George Airy published the solution to the problem of diffraction at a circular aperture and demonstrated that the image of a point source was a diffuse disc of light which came to be known as the Airy disc. By the end of the 19th century, the performance of objective lenses had begun to approach the theoretical diffraction limit.

In 1873, Ernst Abbé published a model for the optical microscope based on two stage Fresnel diffraction. This became the foundation of modern scalar diffraction theory and is used extensively in the work presented here.

For a long time there have been biological and metallurgical applications requiring measurements of optical microscope images, but it is only in the last decade that there has been considerable commercial pressure for making precise measurements on small features. The recent demand for optical microscope measurement systems has come from the semiconductor industry. During the manufacture of integrated circuit photomasks and wafers, it is necessary to be able to measure the dimensions of features in the circuit pattern. Many of these features are now approaching 1µm or less in width and manufacturers wish to control dimensions to within tolerances of less than 0.1µm. In order to do this, the measurements must be precise to only a few hundredths of a micrometre. This precision is well beyond the resolution limit which arises from Abbé's diffraction model for the optical microscope.

A scanning electron microscope (SEM) has much better resolution than the optical microscope. Unfortunately, SEM's are too slow to use in a production line environment as they require the chamber to be evacuated before an image can be obtained. Most semiconductor features include dielectric materials which become charged as the electron beam sweeps across the specimen. This causes the image to saturate very quickly and makes measurement difficult. Charging effects can be greatly reduced by gold coating the specimen, but this cannot be done if the specimen is to be made up into a device. Image interpration is difficult with an SEM as the image profile is very sensitive to the electrical drive conditions, specimen material and tilt angle.

Improved resolution can be obtained with an optical microscope by using a shorter illumination wavelength. Unfortunately, as the illumination wavelength decreases and becomes comparable with the thickness of the feature, the object begins to act as a waveguide and produces very complex image profiles. Thus although the resolution has improved, the image has become a complex pattern of waveguide modes which is difficult to interpret in order to obtain a measurement.

It is not resolution which limits the performance of an optical measurement system, but image interpretation. To make a precise measurement, features must be located in the image profile which define the line edges. The features used are often dark fringes at

the line edges, steep slopes in the image profile or just a particular intensity threshold. The measurement precision is then determined by how well the system can locate these features. If these edge defining features in the image can be related to the physical edge of the object, then an accurate measurement is possible. The optical microscope has proved quite satisfactory for measuring large dimensions but for linewidths approaching 1μm and where very precise measurements are required, it is necessary to improve the understanding of the process of image formation.

There are other areas where precise dimensional measurement of small geometries is required. These include measuring the size of the gaps in magnetic recording heads and the geometries of optical fibres.

1.2 Precision, accuracy, repeatability and resolution.

There is much confusion over the use of the terms precision, accuracy, repeatability and resolution. These terms are used frequently throughout this thesis and so their meaning will be explained by way of the following example.

Consider a general optical system which is forming an image of two incoherent point sources of light and it is required to measure the distance between these two points. The optical system will have a limited aperture and various defects. This results in the image of each point source spreading out. Each point source will produce an image known as the point spread function (PSF) and as the two points are brought closer together, the two PSF's will merge together. The point where the two images appear to merge into one is called the resolution limit and hence it is the smallest distance between the two points where the images appear as two and not one.

Different optical systems and illumination conditions produce different shaped point spread functions. Hence comparison between systems becomes difficult. There have been many different resolution criteria proposed, but probably the most well known is the Rayleigh resolution limit. According to the Rayleigh resolution criterion, it is possible to resolve two points which are separated by a distance greater than the distance from the centre of the point spread function to the position of the first zero. This definition was based on the ability of the human eye to see two peaks. If the image were scanned

with a photodetector then from the image profile, it would be possible to resolve the two points even when they were much closer together, in which case other resolution criteria may be more appropriate.

In general the definition of the term "resolution limit" must be stated when assessing a system. The "resolution limit" is an attempt to describe an entire function (the point spread function) which may exist in more than one dimension, with only one number. From an instrumentation point of view this single number is frequently insufficient to fully characterise a system. Rayleigh's definition falls down when applied to instruments such as video cameras where the response is Gaussian and thus has no zero. According to this definition, it would not be possible to resolve two points no matter how far apart they were.

Now consider the images of these two points, with the points themselves sufficiently far apart as to produce negligible interference between each other in the image plane. Choose any intensity threshold in the image and locate the points where the image has this intensity (figure 1.2.1). Now consider only points A and D, and measure the distance between them. Move the two point sources further apart by a small distance δx and remeasure the distance between the intensity levels A and D. The measured distance will have increased by an amount δx, which may be considerably smaller than the resolution limit. Precision is defined as the smallest change (δx) which can be detected by the measurement system.
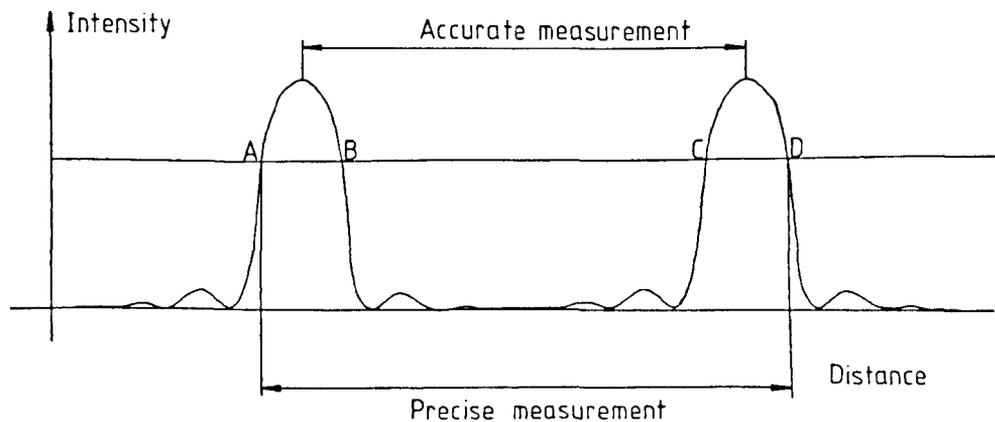
Figure 1.2.1. The accurate measurement is the distance between the two peaks, but a precise measurement can be obtained by measuring between any two points.

Although this measurement is precise to much better than the resolution limit, it is not very accurate. Accuracy (or more strictly, inaccuracy) is defined as how close the measured value is to the true value. In this case, the true value is obtained by measuring the distance between the tops of the peaks.

If the measurement is repeated on the same object, then the measured value may change. Measurement variations can be caused by factors such as temperature changes, age, alignment, different operators etc. The repeatability of a measurement is defined as the spread in values which are obtained under particular conditions and it is usually measured in terms of standard deviations from the mean. When quoting a value for repeatability it is necessary to quote the conditions under which it applies (eg. temperature range, time period, misalignment etc.).

1.3 Measurement standards.

An isolated measurement system can be used to measure a batch of objects and rank them in an order of comparative size. This enables short term process variations to be identified. If a second batch of objects is measured later, then it may not be possible to compare the results with the first batch as the system calibration may have drifted. Similarly the measurements from two isolated systems cannot be compared.

This problem can be overcome by measuring some of the objects on both systems and then keeping them to use as calibration standards. This approach allows the production at one site to be monitored but unfortunately it does not help with relating the measurements to accurate values. In some cases, some of the objects can be measured by another method, for instance in a scanning electron microscope and used as known calibration standards.

This would produce an in-house standard but does not enable measurements to be compared with those from other sites. In order to solve this problem, calibration standards are necessary. If each production site has a standard and they are all calibrated centrally, then there should be good agreement between sites. This approach has been adopted successfully by the National Bureau of Standards and the National Physical Laboratory, who both produce photomask linewidth

calibration plates. These plates are made from the same chrome on glass materials as production masks and contain a range of accurately known linewidths. The masks are calibrated optically but have good agreement with diffraction theory and results from a scanning electron microscope.

A pitch calibration standard is not sufficient for most measurement systems as it is necessary to know whereabouts on the image of the edge, the true physical edge actually lies. Using a line from a calibration standard, it is possible to locate the position (usually in terms of intensity thresholds) in the image profile, corresponding to the true edge. This can then be used as the edge definition when measuring unknown lines.

In some measurement systems, it is not possible to control the edge location definition and measurements have to be made between two fixed thresholds. These systems can be calibrated by measuring the calibration standard to determine the offset between what the system measures and the true value. This offset can then be used to correct the measurements made on unknown linewidths.

A calibration standard must have similar imaging properties to the object to be measured. If the object and calibration standard produce different image profiles of a line edge, then it is no longer possible to find the true edge from the calibration standard and transfer this to the unknown. This presents significant problems when making measurements on semiconductor wafers as the images vary significantly not just between processes, but between batches, wafers and even across a wafer. The image is a function of object thickness, refractive index, illumination wavelength etc. and so a linewidth measurement standard must be able to take these variations into account.

Nyyssonen (1978) realised that thin layers (<200nm thick) could be represented by objects with a given phase and contrast. Careful design of a linewidth measurement standard enabled good approximations to most phase and reflectivity conditions to be achieved with a limited number of cases.

Unfortunately, the imaging of thick layer objects is far more complex and as yet no attempts have been made to produce a comprehensive thick

film linewidth measurement standard. The models capable of describing the images of thick film line objects have been limited and consequently there has been little understanding of the imaging process. In this thesis a general model is developed for imaging thick film objects.

1.4 Thesis overview.

The aim of this work has been to increase the understanding and capabilities of optical critical dimension measurement. The project has been directed by the nature of the problem to be tackled, rather than to develop technology, models or techniques and then seek applications for them.

The problem of measuring the critical dimensions of small objects has been split into two parts; those objects which are thin and can be modelled using existing scalar diffraction theory, and thick objects which require the development of new theoretical models.

Scalar diffraction theory is developed in chapter two into a form which is appropriate to the problem of linewidth measurement on planar objects. These imaging models are used to investigate the optical image profiles formed by line objects. The effects of different materials, illumination conditions and focus errors are studied with particular reference to semiconductor wafers and photomasks. The models are then used in chapter three to investigate and analyse the performance of existing measurement systems. Models are developed for measurement transducers and techniques. The system models are used to identify performance limitations and to determine design requirements for optimum systems. Some practical experiments are described which show the limitations of scalar diffraction theory.

Chapter four tackles the problems of measurement automation and models are developed for automation algorithms. These include methods of detecting critical focus, locating the feature to be measured and automating the measurement itself. The development of a measurement system using threshold detection is discussed.

The purpose behind modelling the measurement process and automation techniques is to develop an optimised measurement system. Chapter five is a case study where these models are applied to the design of an automated system for measuring the gaps in the magnetic circuits of

recording heads for computer discs and video recorders. It is shown how the models can be used to analyse the system operation and design in order to improve the performance.

At this point, the work has reached the limits of scalar diffraction theory and so in chapter 6 the measurement of thick objects is considered. The problem of thick layer imaging is defined in relation to semiconductor structures and a waveguide model is developed. The model is tested rigourously and then used in chapter seven to investigate the imaging of thick layer objects. These studies are then used to develop improved measurement techniques for thick layers.

CHAPTER TWO

Modelling the Optical Images of Planar Objects.

Classical models for the imaging of an optical microscope have been based on Fraunhofer diffraction theory for planar objects (Born and Wolf (1980)). It is assumed that the object may be represented by a two dimensional complex function f(x,y) which describes the amplitude and phase of the field across the surface of the specimen. The relative simplicity of this approach compared with the complexity of treating the object as a three dimensional structure makes it very desirable to reduce objects to two dimensional field patterns.

2.1 The physical structure of quasi-planar objects.

There are no known objects which are truly optically two dimensional. All materials have finite real and imaginary components to their refractive indices and a general object may be described fully by its refractive index distribution $n(x,y,z)$. In practice there are many objects which are sufficiently thin that they may be approximately modelled by two dimensional field distributions. This approximation holds for objects with patterned layers which are thin compared with the illumination wavelength.

Figure 2.1.1 shows the cases of transmission and reflection for a completely general thin film object. At the top surface of the specimen defined by $z=0$, there will be a complex field with amplitude $a(x,y)$ and phase $\phi(x,y)$. Although this is valid in two dimensions, the field may be simplified to $a(x)$ and $\phi(x)$ in the case of one dimensional line objects.
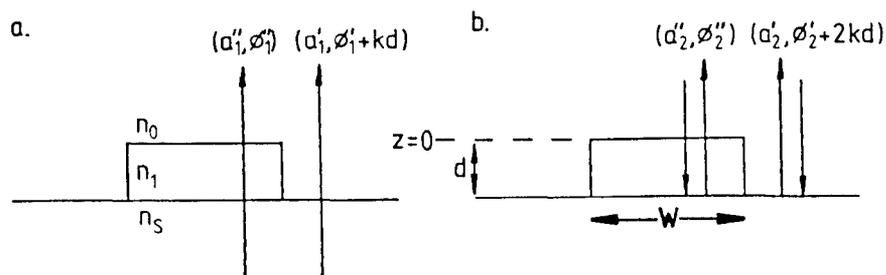
Figure 2.1.1. Phase and amplitude of the light transmitted and reflected by a thin film object.

The amplitude and phase arising from the two light paths may be computed using the Fresnel equations (Hecht and Zajac (1974)). When

calculating the relative phase over the two paths it must be remembered that the path which does not go through the line object has an additional path length in air.

Usually it is only relative amplitude and phase distribution which is of interest and so the field for the transmitted case may be written in normalised form.

$$a(x) = \begin{cases} a_1''/a_1' & |x| < w/2 \\ 1 & |x| > w/2 \end{cases} \qquad (2.1.1)$$

$$\phi(x) = \begin{cases} \phi_1' + kd - \phi_1'' & |x| < w/2 \\ 0 & |x| > w/2 \end{cases} \qquad (2.1.2)$$

w = linewidth.

The field for the incident illumination case may be written in a similar form. Thus the complex amplitude profile at z=0 may be written as,

$$f(x) = \begin{cases} a \cdot e^{i\phi} & |x| < w/2 \\ 1 + 0i & |x| > w/2 \end{cases} \qquad (2.1.3)$$

It is important to note that the a(x) and ϕ(x) terms computed using the Fresnel equations are dependent upon the illumination angle (Born and Wolf (1980)). If the incident illumination deviates significantly from the optical axis then the relative reflectivity will change. This effect becomes more severe as the patterned layer gets thicker.

The object description of equation (2.1.3) is applicable to all thin film objects. The relative amplitude of the line object is readily determined as it is simply the square root of the relative intensity of the line image. The phase however is harder to determine and in general it will be necessary to know the refractive index and thickness of the line object.

Practical semiconductor wafer structures such as thin (<200nm) oxide layers on silicon, produce images with relative intensities which vary from about 25% to 100%. Over this range of thicknesses, the phase difference between the two paths may lie anywhere from 0 to 2π.

With chrome on glass photomasks the problem is less serious as the chrome coating transmits very little light. Thus since the amplitude (a) is small, ϕ need not be known as accurately. Most commercially available photomasks have nominal chrome thicknesses of 0.1 μm. This

results in an optical density in green light of between 2.0 and 3.0 and thus the relative intensity is less than 1%.

2.2 Physical model of the optical microscope.

The optical microscope may be operated in either transmission or reflection, however both modes may be modelled in the same manner. The image produced by an object is determined by the illumination conditions as well as the objective. Thus in transmission, the microscope model must include the condenser as well as the objective. In reflection, the objective acts as the condenser and for simplicity only the transmission mode will be discussed, but the model is equally applicable to reflection.
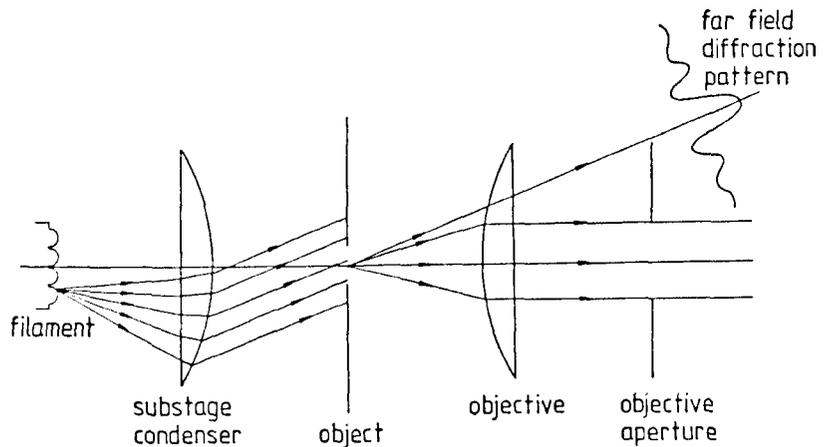
Figure 2.2.1. Simplified physical model of the optical microscope.

Köhler illumination is the most common form of illumination in critical dimension microscopy and is the mode which will be assumed in this model. It is achieved by focusing the image of the lamp filament onto the back focal plane of the condenser. The effect of this is to collimate the light emanating from any one point on the surface of the filament and so the lamp filament is completely out of focus in the object plane.

According to figure 2.2.1, the lamp filament lies in the back focal plane of the condenser lens and the acceptance angle of the objective is determined by an aperture in the back focal plane. The object lies at the principle focus of both the condenser and the objective. Thus the illumination is controlled by the brightness distribution in the back focal plane of the condenser and the captured light used to form the image is determined by the objective aperture.

The most common illumination source is a tungsten halogen lamp which is an intrinsically incoherent source. Alternatively, incoherent sources may be produced by illuminating a ground glass screen with any source and this has the advantage that a laser may be used. The light emanating from any point in the filament plane is assumed to be incoherent with the light from any other point. This is reasonable as the illumination from a filament is thermally generated and the filament is several millimetres long. Thus each point in the plane of the filament may be considered separately.

The filament lies in the back focal plane of the condenser and so the light emanating from one point will be collimated by the condenser lens to produce a planar wavefront which illuminates the specimen. The different points in the filament plane will all produce wavefronts which are incident upon the specimen at different angles. Figure 2.2.1 shows one angle and how the object diffracts this light to produce a far field diffraction pattern. This pattern lies in the back focal plane of the objective but only the portion which lies within the objective aperture will contribute to the image.

There are several assumptions regarding the optical system which are necessary in order to keep the model simple.

1. Planar object.
2. Single wavelength illumination.
3. Isoplanatic and diffraction limited optics.
4. No glare.
5. Evenly distributed illumination.
6. Scalar diffraction theory applies.

Planar models for objects were discussed in section 2.1 and it was pointed out that for many structures the transmission function is itself a function of illumination angle and consequently the far field diffraction pattern varies with illumination angle. For some objects the relationship between the transmission function and the illumination angle is known and may be included in the model. However in many cases it is not known and for simplicity it will be assumed to be constant with angle.

All diffraction effects are wavelength dependent and so the image must be computed for a series of wavelengths across the spectrum of the

source. Alternatively a filter may be used to restrict the illumination bandwidth and thus enable the assumption of monochromatic illumination to be made. The latter approach is preferrable as it greatly simplifies the computation. This is because the object transmission function, far field diffraction pattern and the aberrations are all wavelength dependent.

The optical components of the microscope will contain aberrations which result in wavefront distortions. The most serious source of error is spherical aberration in the objective and a lens should be used which has been corrected for operation at the chosen wavelength. In general the exact nature of the aberration for a given lens is not known and hence it is necessary to start from the assumption of aberration free optics.

The optical surfaces in a lens not only refract light but they scatter it as well. The result is that not all light from one point in the object plane reaches its conjugate point in the image plane. Some of this light is scattered and ends up in other areas of the image plane. The light is incoherent with the image and becomes more out of focus, the further it deviates from its correct point in the image plane.

The glare in the image plane may be resolved into two components, light from the neighbourhood of the feature of interest and light from outside this neighbourhood. The former component will be out of focus in the image plane, but it will vary in intensity across the feature. The latter component arises from features sufficiently far away from the feature of interest that it is completely defocused in the image plane. Thus the glare consists of a component which varies slowly across a small feature and a second component which is a constant background level.

If the object has a very high contrast then it is possible to measure the background glare component by measuring the light intensity of a large dark region. However the slowly varying component cannot be measured as this will change away from the feature of interest. The glare may be assumed to be constant across a small feature, but not constant across the field of view.

Each point in the back focal plane of the condenser will produce a coherent image. The intensity of this image is proportional to the

brightness of the point source and thus the final image will be modulated by the filament structure in the condenser back focal plane. If a ground glass screen is used then it is reasonable to assume that the illumination is evenly distributed and this greatly simplifies the model.

The object is assumed to be described by a transmission function which is independent of polarisation angle. This provides considerable simplification as the polarisation effects will be a function of the angle of incidence.

The validity of these assumptions and the constraints they impose on the specification of the optics will be investigated later once the model has been established.

2.3 Analytical model of the optical microscope.

The choice of model for the optical system will depend on the application for which it is required. The importance and validity of the assumptions made in section 2.2 will determine the nature of the model.

Hopkins (1950) considered the effect of the substage condenser on the resolution of the optical microscope. This naturally determines image edge shape which is important for precise dimension measurement. Later, Hopkins (1953) presented a complete analysis for partially coherent images of both periodic and non-periodic patterns. The analysis assumed that scalar diffraction theory held and that the object could be described in terms of a planar complex transmission function.

Richards and Wolf (1959) investigated the electromagnetic field around the focal plane when the illumination is linearly polarised. They demonstrated that there was a deviation from the scalar theory at high numerical apertures.

Watrasiewicz (1965a) investigated the discrepancies between scalar theory and experimental image profiles and found that at high numerical apertures, the spherical aberration in an objective lens became appreciable. He also noted from experimental results that the edge profile was a function of the plane of polarisation of the illumination. In a second paper (1965b) he presented an analysis for

imaging edges in partially coherent light and concluded that good agreement with scalar theory was obtained for angles of convergence up to 30°.

Ichioka and Suzuki (1976a) investigated general phase and amplitude gratings with partially coherent illumination. They restricted their attention to diffraction limited aberration free optics but in a later paper they considered the effect of defocus.

The development of comprehensive models for partially coherent imaging was limited in the early years by the lack of computing power available. Hopkins (1953) described diffraction limited imaging in terms of transmission cross-coefficients. Ichioka and Suzuki (1976a) used this approach which greatly speeded up the computation of image profiles. The technique was later used by Kintner (1977) specifically for computing image profiles for linewidth measurement.

Nyyssonen (1977) compared theoretical and experimental image profiles and obtained close agreement even at high angles of convergence. Previously, agreement had only been obtained for low numerical apertures and this improvement may be attributable to the fact that Nyyssonen took great care to ensure good edge quality on the line objects and also selected objectives with very low spherical aberration. Nyyssonen (1978,1979,1982a) also investigated the imaging of thin transparent layers in reflection and reported good agreement for small illumination apertures. It would be unreasonable to expect good agreement at high angles of incidence as the relative reflectivity of the object would be a function of angle.

In order to carry out investigations into the imaging of thin layer objects it is necessary to develop a model for the optical system. In section 2.3.1 a general analytical model for the imaging of line objects is presented and in section 2.3.2 this is developed further into a more practical discrete realisation based on periodic object structures.

2.3.1 Imaging planar objects with a finite support.

In this section a model will be derived to describe the imaging of an optical microscope. Although this is largely based upon the previously published work discussed in section 2.3, the model will be adapted for application to critical dimension measurement. This means

close attention to aspects such as the definition of the object, focusing errors and wavelength effects. On the other hand, since the work is concerned primarily with measuring the width of lines, the model will only consider one dimensional objects.

The optical microscope is a two dimensional optical system and where circular apertures are used this cannot be ignored. In order to simplify the analysis only a one dimensional model will be derived here but this is later extended to two dimensions in order to accommodate practical systems.
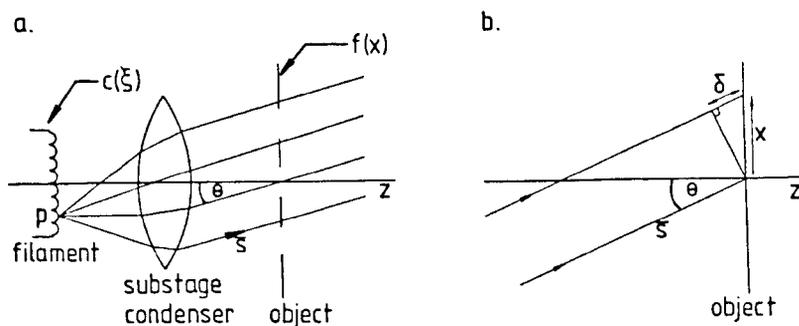
Figure 2.3.1. a) Point source illumination of the focal plane.
b) Phase delay across the incident wavefront.

Consider light emanating from a point (p) on the filament surface (figure 2.3.1a). This will pass through the condenser to produce a collimated beam and thus illuminate the specimen with a planar wavefront. The direction of the wavefront is given by the vector $\bar{s}$ which makes an angle θ with the optical axis $\bar{z}$. This produces a phase gradient across the focal plane. The phase at a point x relative to the origin (figure 2.3.1b) will thus be given by,

$$\phi(x) = x \frac{2\pi}{\lambda} \sin\theta \qquad (2.3.1)$$

The range of θ is determined by the numerical aperture of the condenser and so it is more convenient to use the notation,

$$\xi = \sin\theta \qquad \text{and} \qquad k = \frac{2\pi}{\lambda} \qquad (2.3.2)$$

where,

$$-NA_c < \xi < NA_c \qquad (2.3.3)$$

Thus the incident field at the focal plane will be given by ,

$$s(x) = e^{ikx\xi} \qquad (2.3.4)$$

and the field at the top surface of the specimen by,

$$f'(x) = f(x).e^{ikx\xi} \qquad (2.3.5)$$

In writing equation 2.3.5, it is implicitly assumed that $f(x)$ is not a function of θ and as explained in section 2.1 this is not always true.
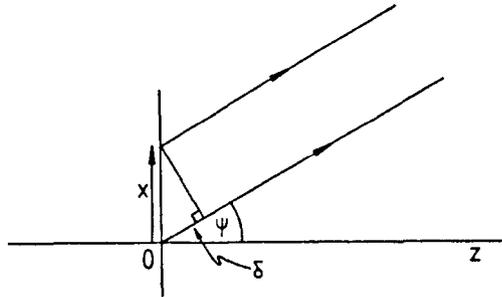
Figure 2.3.2. Phase delay between two points in the focal plane.

Having determined the field distribution in the focal plane, it is now necessary to find the far field diffraction pattern produced by the specimen. Consider a wavefront leaving the object plane at an arbitrary angle ψ (figure 2.3.2). At a distance (r) from the origin, the wave equation as a function of distance and time is,

$$G_0(r,t) = f'(0).\exp(i\omega t - ikvt) \qquad (2.3.6)$$

and for a wave originating from a point which is a distance (x) from the origin,

$$G_x(r,t) = f'(x).\exp(i\omega t - ikvt - ixk.\sin\psi) \qquad (2.3.7)$$

Consider a point which lies at an angle ψ from the origin and an infinite distance from the focal plane. The summation of all the waves from the neighbourhood of the origin will thus be,

$$g(\psi) = \int_{-\infty}^{\infty} f'(x).\exp(ixk.\sin\psi) \, dx \qquad (2.3.8)$$

The variation with (t) has been left out and so $g(\psi)$ represents the amplitude only. This distribution represents the far field diffraction pattern of the object and it lies around a circle described by ψ. Clearly this is the wavefront which will be accepted by the objective and so it is more convenient to write,

$$\xi^- = \sin\psi \qquad\qquad (2.3.9)$$

and thus diffraction angles which satisfy the condition,

$$-NA_0 < \xi^- < NA_0 \qquad\qquad (2.3.10)$$

will be accepted by the objective and used to form the image. Thus equation 2.3.8 may be written as,

$$g(\xi^-) = \int_{-\infty}^{\infty} f^-(x) \cdot \exp(ixk\xi^-) \, dx \qquad\qquad (2.3.11)$$

In this form the far field diffraction pattern is recognised as being the Fourier transform of the field distribution $f^-(x)$. ξ^- is of course only observable over the range described by equation 2.3.10.

Now $f^-(x)$ is a function of the illuminating angle as well as the object transmission function. Substituting equation 2.3.5 into equation 2.3.11 gives,

$$g(\xi^-) = \int_{-\infty}^{\infty} f(x) \cdot \exp(ikx(\xi^-+\xi)) \, dx \qquad\qquad (2.3.12)$$

Alternatively, this may be written as,

$$g(\xi^-, \xi) = \int_{-\infty}^{\infty} f(x) \cdot \exp(ikx\xi^-) \, dx \qquad\qquad (2.3.13)$$

and this leads to the result,

$$g(\xi^-,\xi) = g(\xi^-+\xi,0) \qquad\qquad (2.3.14)$$

This is an important and useful result as it means that the far field diffraction pattern does not change in shape as the illumination angle changes, it merely shifts relative to the optical axis. This simplifies computation considerably because the far field diffraction pattern need only be computed once and not for every point on the filament surface. Thus for any given illumination angle, the image formed by the objective may be described by the function,

$$g^-(\xi^-,\xi) = \begin{cases} g(\xi^-+\xi) & -NA_0 < (\xi^-+\xi) < NA_0 \\ 0 & \text{otherwise} \end{cases} \qquad (2.3.15)$$

Thus the image formed by the point source (p) is given by,

$$I(x,\xi) = \left| \int_{-\infty}^{\infty} g^-(\xi^-,\xi) \cdot \exp(i\xi^-xk) \, d\xi^- \right|^2 \qquad (2.3.16)$$

The overall image profile may be found by summing the images formed by all the points on the filament and hence the final image is given by,

$$I_t(x) = \int_{-1}^{1} c(\xi).I(x,\xi) \, d\xi \qquad (2.3.17)$$

Where $c(\xi)$ represents the brightness distribution of the filament across the condenser aperture. Since $c(\xi)$ is not normally known as it depends on the exact filament shape and position, it is convenient to define the effective source as,

$$c(\xi) = \begin{cases} 1 & |\xi| < NA_c \\ 0 & |\xi| > NA_c \end{cases} \qquad (2.3.18)$$

This is quite valid if a ground glass screen is used and in practice it is still a quite acceptable approximation without. $c(\xi)$ need not be just a rectangular aperture as defined in equation 2.3.18, but it may be asymmetric about the optical axis and may include stops.

At this stage it is worth considering the effect of aberrations and in particular defocus. Lens aberrations will distort the wavefront and would result in a phase modulation of the expression in equation 2.3.11.

$$g(\xi') = e^{i\rho(\xi')} \int_{-\infty}^{\infty} f'(x).e^{ixk\xi'} \, dx \qquad (2.3.19)$$

where $\rho(\xi')$ is the aberration function of the objective. The treatment of sherical aberration, primary coma and astigmatism have been dealt with in the literature (Steel(1957),Born and Wolf(1980),Nyyssonen(1977)). The effect of a focus error is to distort the wavefront (Stark (1982)) such that equation 2.3.11 becomes,

$$g(\xi') = \exp(\pi i \tfrac{\delta}{\lambda}\xi'^2) \int_{-\infty}^{\infty} f'(x).\exp(ixk\xi') \, dx \qquad (2.3.20)$$

where δ = focus error.

The image described by equation 2.3.17 is valid for one wavelength only and so for a finite illumination bandwidth, it is necessary to compute the image for different wavelengths.

$$I(x) = \int_{0}^{\infty} S(\lambda).I_t(x,\lambda) \, d\lambda \qquad (2.3.21)$$

where $S(\lambda)$ = spectral response of the system.

So far only the one dimensional optical system has been considered but in practice this is not always a reasonable approximation to a two dimensional system. In equation 2.3.14 it was shown that the far field diffraction pattern moves across the objective back aperture as the illumination angle changes. This is true regardless of the direction in which the source moves.
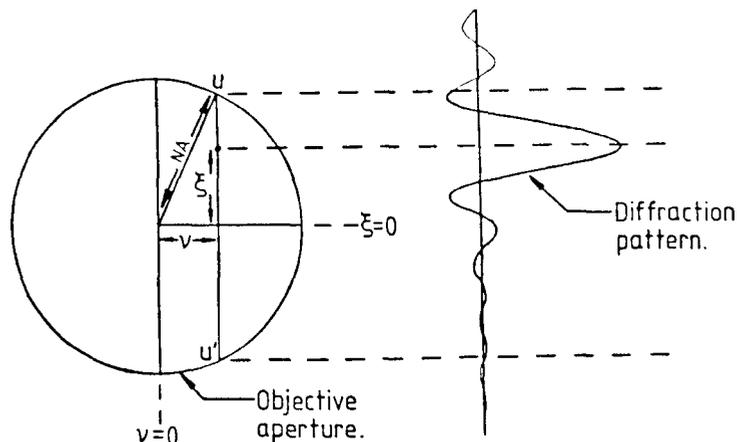
Figure 2.3.3. The effective aperture width in a circular aperture.

Figure 2.3.3 shows how the far field diffraction pattern lies in the back focal plane of the objective. The coordinates ξ and ν represent the position of the point source (p) on the filament surface. The point (p) may lie anywhere within an area defined by the condenser aperture and need not necessarily lie within the objective aperture. The far field diffraction pattern is now centered at the coordinates (ξ,ν) and the portion which passes through the objective to form an image is determined by the effective aperture between the points u and u´. The effective aperture limits are thus given by,

$$u = \sqrt{NA_o^2 - \nu^2} \quad - \xi \qquad (2.3.22a)$$

$$u´ = \sqrt{NA_o^2 - \nu^2} \quad + \xi \qquad (2.3.22b)$$

$\nu = \sin\phi´$, where, $\phi´$ = angle between $\bar{s}$ and the y=0 plane.

Thus for any point (p), the section of the far field diffraction pattern which contributes to the image can be determined and equation 2.3.17 becomes,

$$I_t(x) = \int_s c(\xi,\nu).I(x,\xi,\nu) \, ds \qquad (2.3.23)$$

The integral is now across the surface of the filament.

2.3.2 Periodic object approximation.

Computing the optical image according to the method described in section 2.3.1 requires the solution of several integral equations. However the calculations may be simplified considerably if the object is assumed to be periodic. This is a valid approximation provided that the period is large relative to the feature of interest.

If the general object defined by equation 2.1.3 is replicated with spatial frequency (b) and offset from x=0 by Δ, then the far field diffraction pattern will be a Fourier series of the form,

$$
g(q) = \begin{cases} wbae^{i\phi} + (1-bw) & q=0 \\[2ex] \dfrac{a.e^{i\phi}-1}{q\pi}.\exp(2\pi\Delta qb).\sin(q\pi wb) & q\neq0 \end{cases}
\qquad (2.3.24)
$$

where q is the diffraction order.

Referring to figure 2.3.2 it can be seen that for two points separated by a distance 1/b (the period of the structure) there will be a bright fringe at an angle ψ provided that it satisfies the relation.

$$
b\lambda q = \sin\psi
\qquad (2.3.25)
$$

According to equation 2.3.14 the diffracted orders are shifted by the illumination angle. Let (r) be the illumination order then equation 2.3.25 becomes,

$$
b\lambda(q-r) = \sin\psi \qquad r = \frac{\sin\theta}{b\lambda}
\qquad (2.3.26)
$$

The range of $\sin\psi$ is limited by the effective objective aperture width and the range of $\sin\theta$ is determined by the condenser. If it is assumed that both the condenser and the objective apertures are circular then the limit on the number of orders imaged by the objective will be given by,

$$
-\sqrt{NA_o^2 - \sin^2\theta} - \sqrt{NA_c^2 - \sin^2\theta} < qb\lambda < \sqrt{NA_o^2 - \sin^2\theta} - \sqrt{NA_c^2 - \sin^2\theta}
$$

$$
\qquad (2.3.27)
$$

q can only take on integer values and so for a given (r), the waves

which make up the solid angle given by,

$$b\lambda(q-r) < \sin\psi < b\lambda(q+1-r) \qquad (2.3.28)$$

will all produce the same image as the same number of orders will pass through the objective aperture. Consequently it is possible without loss of accuracy to restrict the calculation to a finite number of orders and a finite number of points across the filament.

Similarly (r) may be restricted to a finite number of values as the image only needs to be calculated when (r) has changed sufficiently to cause the inclusion or loss of an entire order. In practice it is not necessary to consider this many values for (r), as changes of less than 0.1 in $\sin\theta'$ have negligible effect on the image of a planar object. For most microscope systems it is not necessary to use more than about 20 values for r across the back aperture and so for convenience the condenser may be sampled in equal steps in both directions.

Using this discrete model the image may be written as,

$$I(x) = \sum_{n=-N}^{N} \sum_{m=-M}^{M} \left| \sum_{q=Q_1}^{Q_2} g(q).\exp(2\pi iqbx) \right|^2 \qquad (2.3.29)$$

where Q_1 and Q_2 are given by,

$$Q_1 = \frac{-1}{b\lambda} \sqrt{NA_o^2 - (b\lambda n)^2} - M \qquad (2.3.30a)$$

$$Q_2 = \frac{1}{b\lambda} \sqrt{NA_o^2 - (b\lambda n)^2} - M \qquad (2.3.30b)$$

$$M = \frac{1}{b\lambda} \sqrt{NA_c^2 - (b\lambda n)^2} \qquad N = \frac{NA_c}{b\lambda} \qquad (2.3.31a,b)$$

A defocus term may be included by applying the same discrete model to equation 2.3.19 and this produces a phase term for each order in the back aperture.

$$F'(q) = F(q).\exp(\pi i\delta\lambda b^2(n^2+m^2)) \qquad (2.3.32)$$

Where F(q) is the amplitude of the qth diffraction order.

Equation 2.3.29 is quite general but in some cases it contains a certain amount of redundancy. n is allowed to vary over both halves of the aperture but provided that the apertures are symmetrical and the object is not a function of y, then it will not matter from which

side of the ξ axis the object is illuminated. Thus equation 2.3.29 may be computed from n=0 instead of from n=-N. However it should be remembered that the central order must be halved as it lies in both halves of the aperture. Unfortunately the aperture cannot be reduced to a quadrant as, in general, symmetry does not exist about the ν axis.

2.4 Theoretical investigation of thin film imaging.

The discrete model described in section 2.3.2 was used to investigate the imaging of planar objects. In particular, the effect of operating conditions and object structures on the location of the line edge were examined as this is the key to linewidth measurement.

2.4.1 Coherent illumination.

The image profile of an object is dependent upon the illumination angle. Figure 2.4.1 shows the images of a coherently illuminated slit when illuminated from three different angles (Duffieux (1983)). The objective has a numerical aperture of 0.45 and the three conditions represent the source on-axis (sinθ=0.0), just inside the objective aperture (sinθ=0.4) and just outside the objective aperture (sinθ=0.5). This last condition is often referred to as dark ground illumination as only the changes in the field profile are imaged.
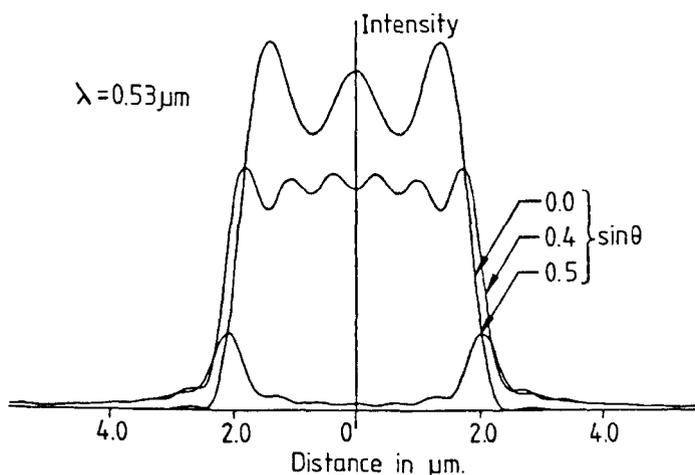
Figure 2.4.1. Image profiles of a coherently illuminated gap for three different illumination angles. Wavelength = 530nm.

As the illumination moves off axis, the far field diffraction pattern is truncated at a lower spatial frequency on one side and a higher frequency on the other. This increases the high frequency range at

the expense of the lower frequencies in the image. Since the higher frequencies carry less energy, moving the illumination off axis reduces the intensity of the image. This is shown by the relative intensities of the image profiles in figure 2.4.1.

When the illumination moves off axis, higher frequencies are imaged and so the the resolution improves. This can be seen by comparing the edge widths of the profiles for sinθ=0 and sinθ=0.4. Also the sinθ=0.4 condition creates a better approximation to a rectangle for the same reason.
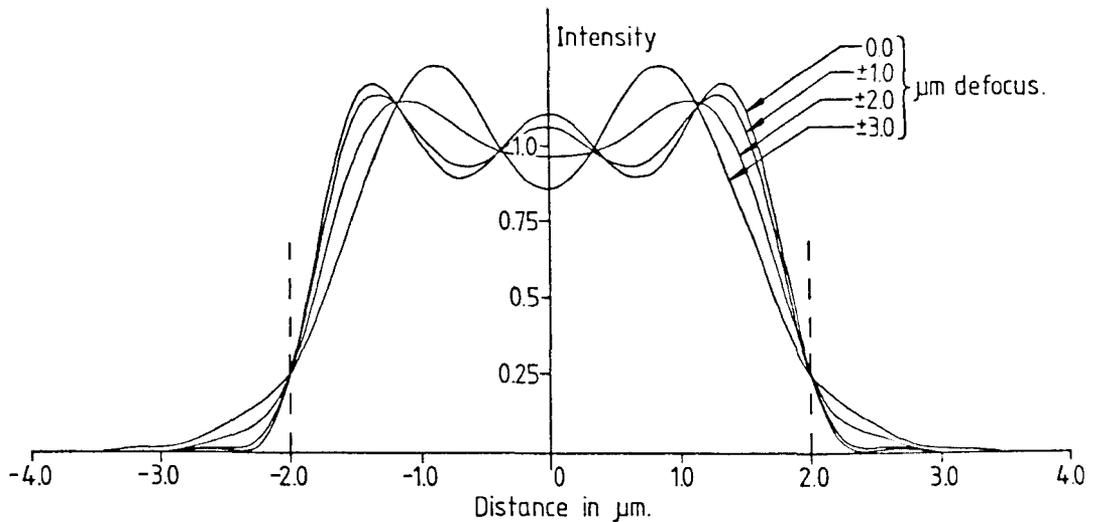
Figure 2.4.2. Through focus image profiles of an on-axis coherently illuminated slit. Wavelength = 530nm, 0.45 NA objective.

The model is capable of predicting the effect of focus errors. Figure 2.4.2 is a through focus series for the case of on-axis coherent illumination. According to equation 2.3.32 the phase modulation is the same for positive or negative spatial frequencies. Thus, provided that the far field diffraction pattern lies centrally within the objective aperture and is real and symmetrical, then the image profile will be the same for focus errors in either direction. It can be seen that the true edge lies at the 25% intensity threshold and that this is true even for significant focus errors. It can be shown that for an isolated edge with infinite contrast, the true edge will always lie at the 25% intensity threshold (Nyyssonen (1982)).

When the illumination is off axis the location of the true edge is no longer insensitive to focus. Figure 2.4.3 is a through focus series under the same conditions as figure 2.4.2 but with the illumination
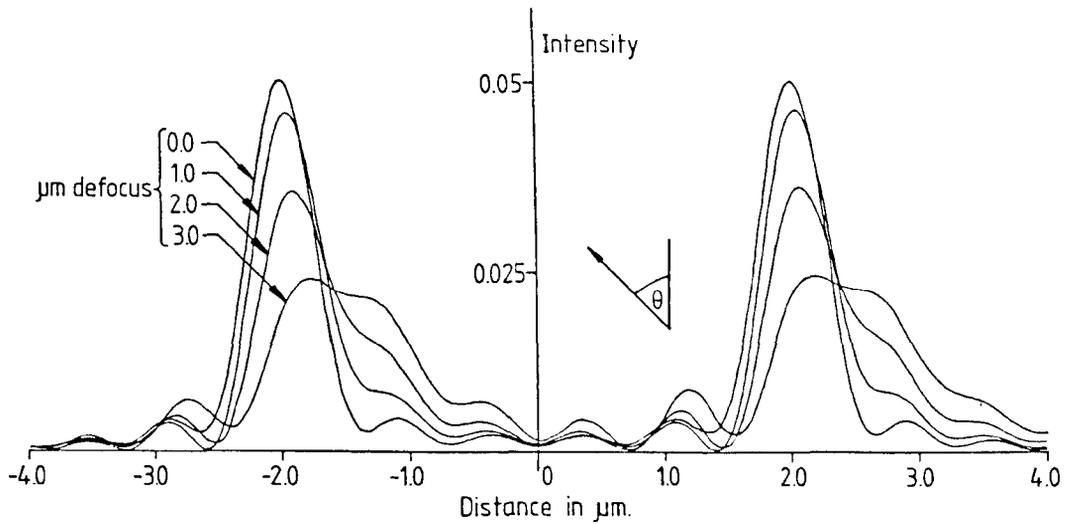
Figure 2.4.3. Through focus series for a slit illuminated from
a coherent source outside the objective aperture.

moved out to beyond the objective aperture such that $\sin\theta=0.7$. As the
object moves out of focus the bright fringes at the line edges blur
and move sideways. If the focus error is in the opposite direction or
if the illumination is from the opposite side then the fringes move
the other way.

The asymmetric focus response arises due to the asymmetric
illumination. If the illumination is now made symmetrical by
illuminating from both sides simultaneously then the image does not
move sideways but just becomes less sharp (figure 2.4.4).
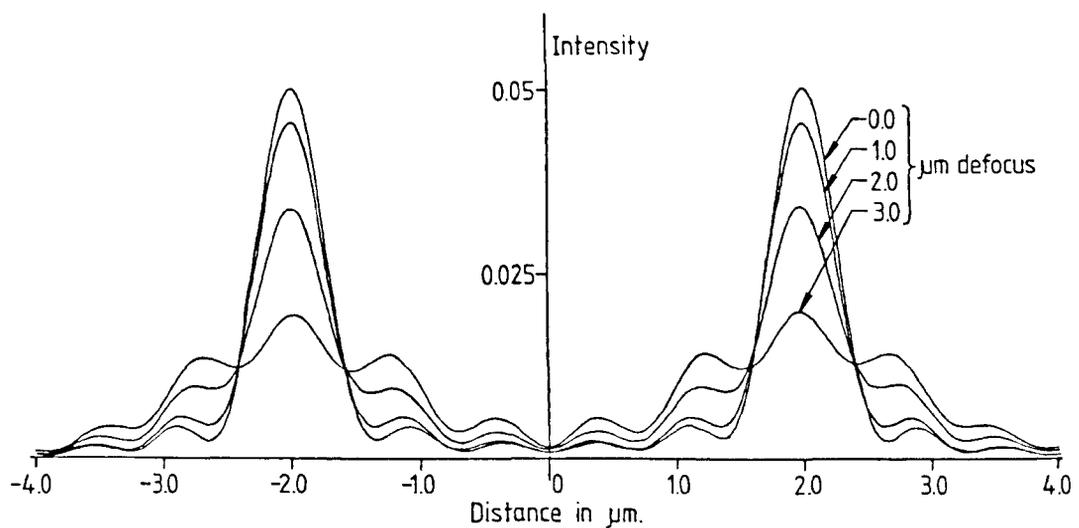
Figure 2.4.4. Through focus image profiles of a slit illuminated from
two point sources located outside and on opposite sides
of the objective aperture

2.4.2 Partially coherent illumination.

If the source is extended to cover a finite angle then the illumination is said to be partially coherent. The degree of spatial coherence (s) is usually defined as the ratio of the numerical aperture of the condenser to the objective. In many areas of microscopy a value of s=2/3 is regarded as an acceptable compromise. Highly coherent illumination (s=0) results in low light levels and severe edge ringing, whereas highly incoherent illumination is difficult to achieve and introduces problems with high angles of incidence.
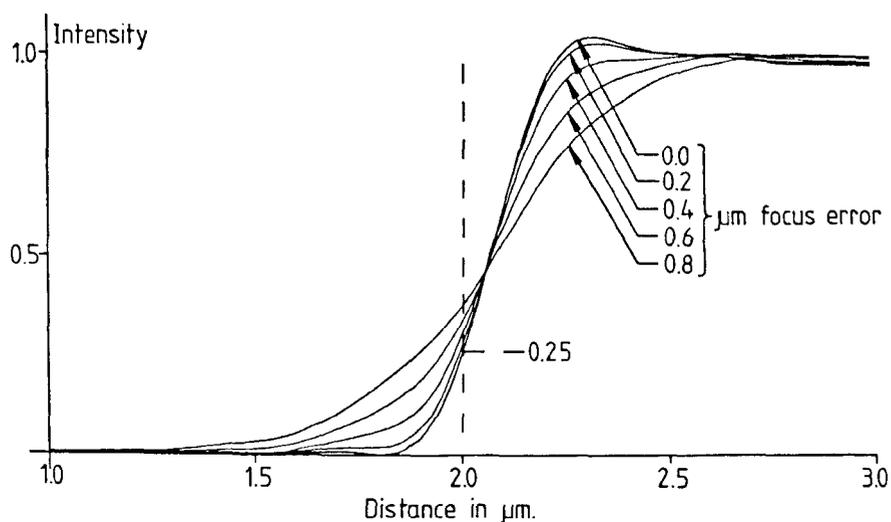
Figure 2.4.5.   Through focus image profiles for an opaque line under
              partially coherent illumination s=2/3. 0.9 NA objective.

Figure 2.4.5 is a through focus series for the image of the edge of a 4μm wide opaque line on a clear background. The optics are assummed to have circular apertures and the coherence parameter (s) has been set at 2/3. The argument used for the isofocal point for coherent illumination no longer applies and the true edge threshold can now be seen to be sensitive to focus.

As the (s) value increases, the true edge threshold and the isofocal point move up from 25% to 50%. The isofocal point moves towards 50% very rapidly as (s) increases but the the true edge threshold moves much more slowly. Figure 2.4.6 shows the case of matched apertures (s=1). The true edge threshold is now 34% which again is not the isofocal point.
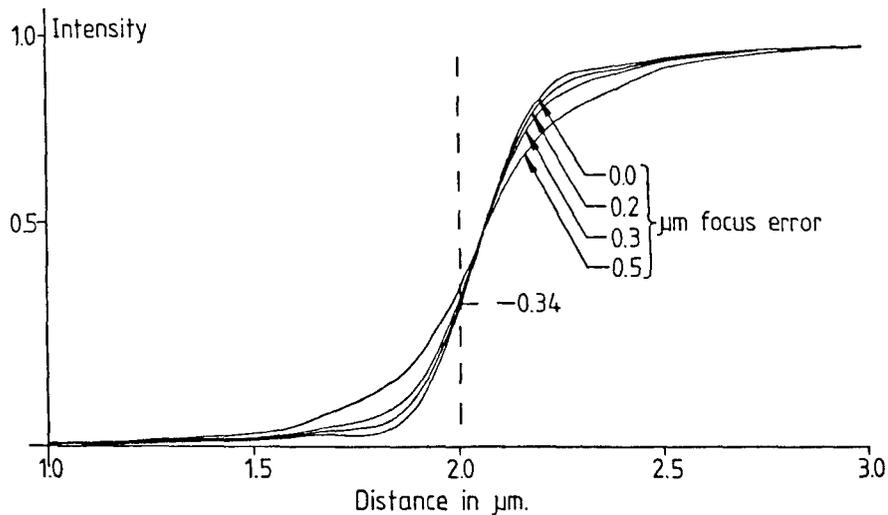
Figure 2.4.6. Through focus image profiles for an opaque line under partially coherent illumination s=1. 0.9 NA objective.

2.4.3 Small line widths.

The limited resolution of the optical microscope need not be a limiting factor in determining the accuracy with which edges can be located. Provided that the object structure and optical conditions are known then accurate edge detection is possible.

The resolution limit only begins to become a problem as the linewidth decreases. On narrow lines the diffraction fringes from neighbouring edges interfere with each other and so the edge intensity threshold becomes a function of linewidth.

Increasing the resolution of the optical microscope is not easy. A 0.9 NA objective is normally used and so oil immersion would be required in order to achieve any significant improvement in resolution. This is unacceptable in the semiconductor industry due to contamination. Alternatively a shorter wavelength could be used but this introduces difficulties in optical design and causes the validity of the thin layer approximation to break down quicker.

The choice of illuminating conditions influences the amount of adjacent edge interference. Figure 2.4.7 shows how the true edge intensity threshold of a clear slit varies as a function of linewidth for the two cases of coherent and partially coherent (s=2/3) illumination. In both cases a 0.9 NA objective has been used. The curves show that if partially coherent illumination is used then the
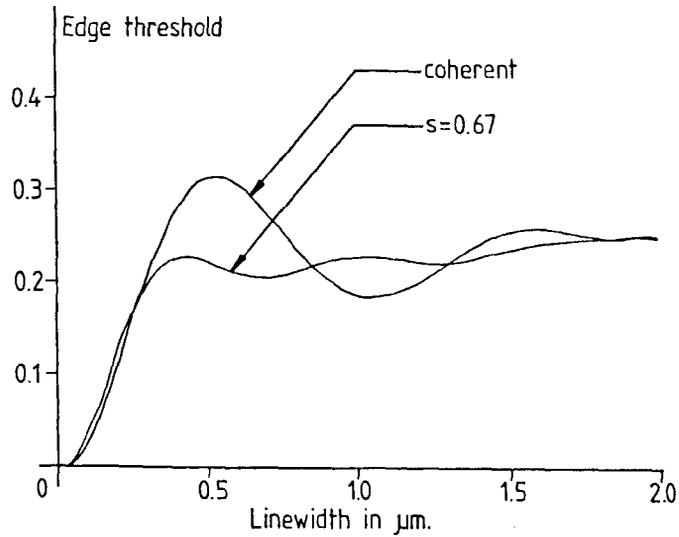
Figure 2.4.7. The effect of linewidth on the intensity threshold
of the edges of a narrow slit (bright field).

true edge threshold is much less sensitive to linewidth and only small
errors are incurred.
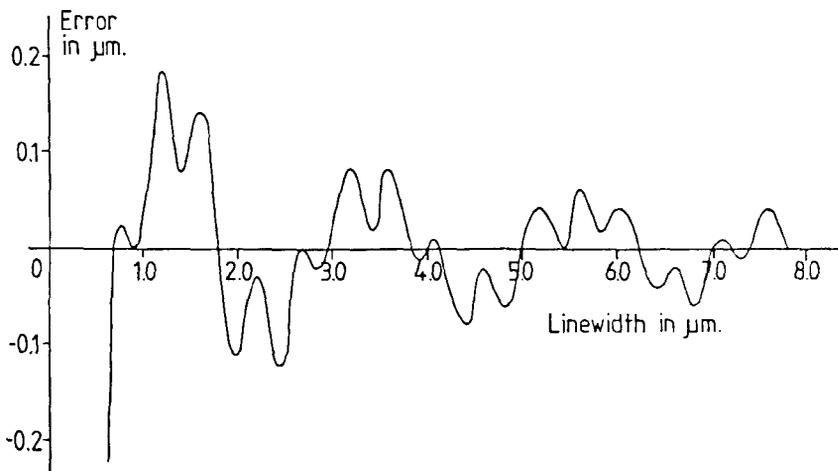
Figure 2.4.8. The effect of linewidth on the peak to peak measurement
error of a coherently illuminated slit (dark field).

The dark ground illumination technique produces bright fringes
associated with the line edges and on very wide lines the true edge
lies at the centre of the fringe. Unfortunately interference between
adjacent edges occurs even with relatively wide lines and so the peak
to peak measurement is not very accurate. Figure 2.4.8 shows how the
peak to peak measurement error varies with linewidth for a coherently
illuminated (sinθ=0.7) clear gap imaged with a 0.45 NA objective.

Lines narrower than 0.5μm cannot be measured as the two fringes merge into one. It should be noted that the case presented here is for a 0.45 NA objective and in practice a slightly higher numerical aperture could be used. This would have the effect of reducing the magnitude and period of the error and enabling slightly narrower lines to be measured. Even so for a measurement technique, the errors are large and for linewidths in the region of 1.6μm to 1.9μm the error is decreasing so fast that the measured linewidth remains stationary at 1.8μm, making it impossible to use an error curve for correction.

The shape of this curve is a function of wavelength and position in the condenser aperture. Thus if a ring illuminator is used over a wide spectral band then the error profile will be smoothed out, improving the accuracy.

2.4.4 The effect of contrast and phase on edge location.

So far the only objects considered have been ideal clear or opaque lines. In practice there will always be a finite contrast and for low contrast objects such as wafers viewed in reflection, this changes the edge image profile considerably. Figure 2.4.9 shows sets of image profiles for three different contrast levels. A 0.85 NA objective has been used with a 0.14 NA illuminating aperture. It is important to restrict the illuminating aperture for thin films as the relative reflectance is a function of angle.

When the contrast is infinite (figure 2.4.9a) the phase is not important as the dark region produces no light. This is useful for photomask linewidth measurement where the contrast is high and so the phase may be neglected.

Some objects exhibit no contrast (figure 2.4.9b). Such an object is easy to fabricate by cutting a step into bulk material. The edge is then only identifiable by a phase change. If the step height is chosen such that the phase shift between the two regions is close to a multiple of 2π, then the edge will disappear and the feature will be indistinguishable from the background. The two phase conditions in figure 2.4.9b show that the depth of the dark fringe depends on the phase but from arguments of symmetry it is always centred on the edge.

When the contrast is finite (figure 2.4.9c) then the bottom of the dark fringe no longer lies at the line edge. The position of the dark

a.

b.

c.
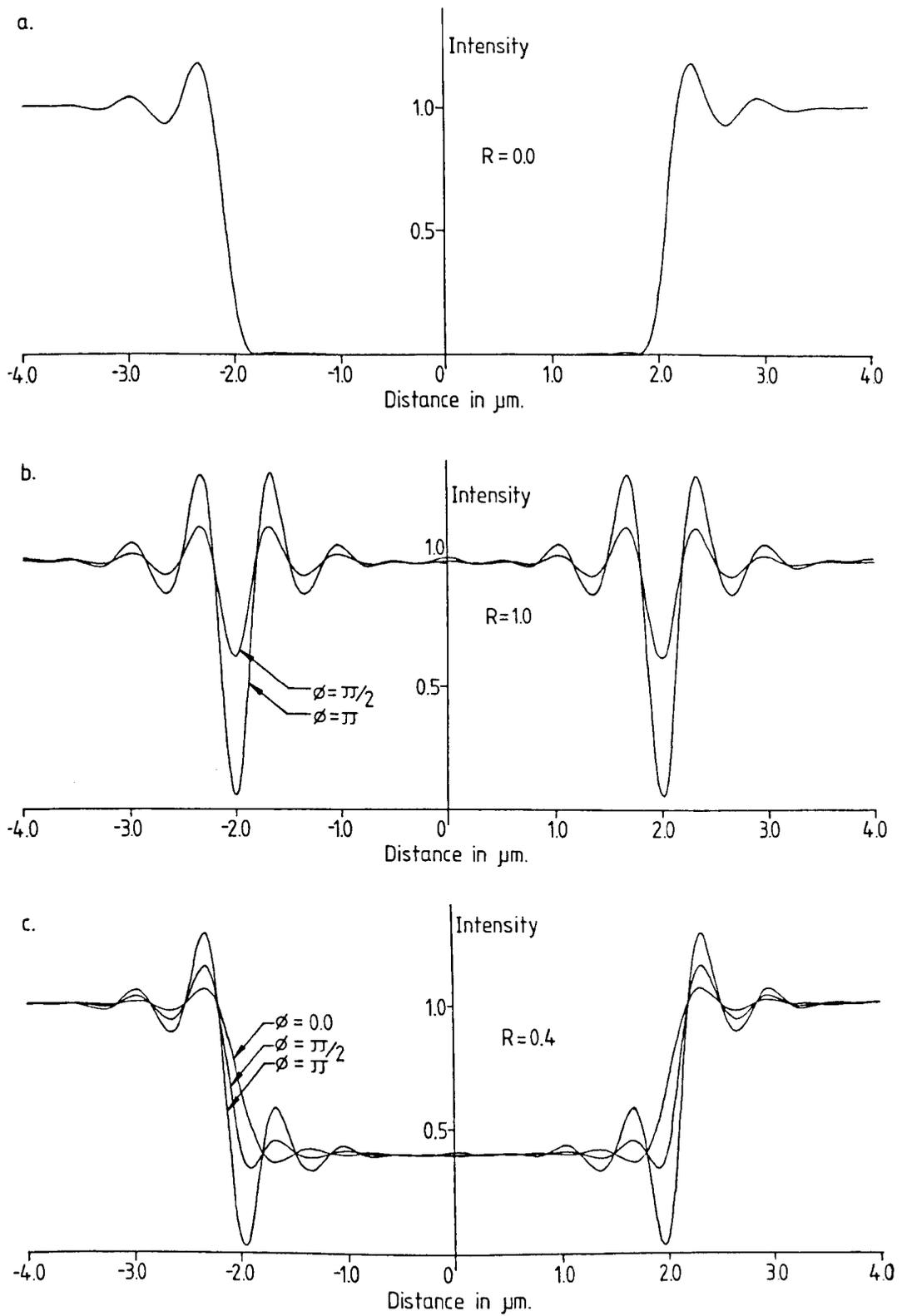
Figure 2.4.9. The effect of reflectivity and phase on the image profile. a) infinite contrast, b) zero and c) R = 0.4.

fringe or indeed whether it exists at all is dependent on the phase. Similarly the edge slope and true edge intensity threshold are phase dependent.

For a general line object where there is a tolerance on thickness, both the phase and relative reflectance will vary. Figure 2.4.10 shows the theoretical image intensity profiles of a silicon dioxide line on a silicon substrate for different oxide thicknesses. Only as the line gets thicker is there a well defined dark fringe and the edge slopes change considerably with thickness. Under these circumstances a different edge location criterion is required for each case.
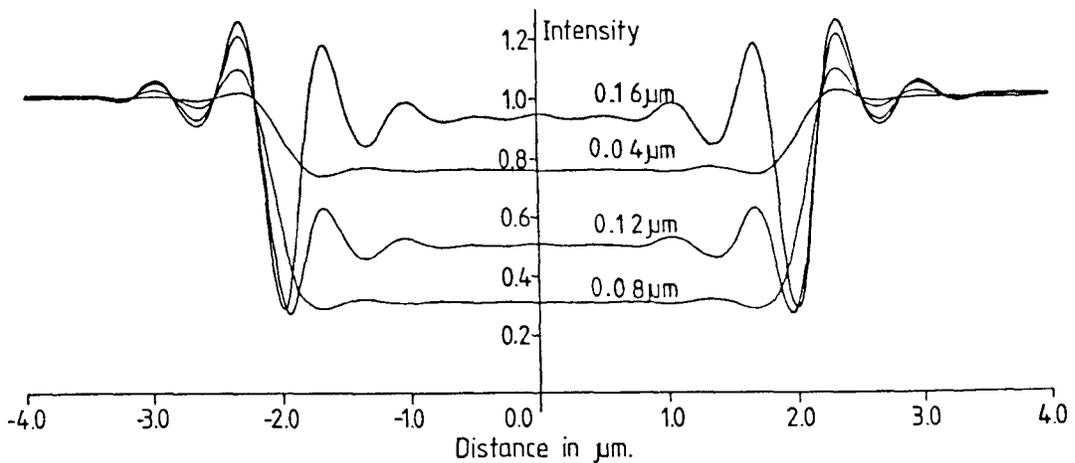
Figure 2.4.10. The effect of thickness on the image profile of an oxide line (refractive index = 1.46+0) on silicon (refractive index = 4.1+0.06).

2.5 Linewidth measurement on thin layer objects.

In order to make an accurate linewidth measurement it is necessary to be able to locate the true edge directly or to locate a feature which lies a known distance from the true edge. Locating the true edge directly is not always advisable as it is quite possible that the image profile in the region of the true edge is very sensitive to parameters such as refractive index and wavelength and so a considerable error may be introduced into the measurement.

In general, linewidth measurements are required for process control and it is usually sufficient that the functional relationship between the measurement width and the true width remains constant. This gives measurement reproducibility across specimens but not necessarily accurate measurements. In order to do this, it is necessary to

identify measurement techniques which are insensitive to changes in the optical image profile caused by changes in the object (other than linewidth).

2.5.1 Focus detection.

The optical image of a line object is quite sensitive to focus errors. Defining critical focus for a planar object is relatively simple. Since the object is assumed to be flat, it is said to be in focus when it lies at the principal focus of the objective and any movement away from this plane represents a focus error.

Intuitively an object is said to be in focus when it appears "sharpest". This rather subjective criterion is reached by adjusting focus to obtain the steepest edge slopes or sharpest fringes in the image profile. There is of course no reason why this condition should locate the focal plane.

Many focusing techniques are based on monitoring sections of the spatial frequency spectrum as the object moves through focus. If the object is replicated with spatial frequency (b), then it will form an image (I(x)) which can be represented by a Fourier series ($F_s(q)$).

$$I(x) = \left| \sum_{q=Q_1}^{Q_2} F_s(q).\exp(2\pi iqxb) \right| \qquad (2.5.1)$$

Where F(q) are the Fourier coefficients of the harmonics (q) in the image profile and Q_1 and Q_2 define the limits of the frequencies present in the image.

Most image profile based focusing techniques can be analysed in terms of how they select sections of the series F(q). There is an upper limit to the spatial frequencies in the image. If the object is illuminated from an angle θ from the optical axis, then the incident field (s(x)) is,

$$s(x) = -\exp\left(2\pi ix\frac{\sin\theta}{\lambda}\right) \qquad (2.5.2)$$

The field across the object (f´(x)) is determined by the object (f(x)) and the field (s(x)).

$$f´(x) = f(x).s(x) \qquad (2.5.3)$$

Assuming that f(x) is periodic with spatial frequency (b) and can be expressed as the Fourier series F(q), then,

$$f'(x) = \sum_{q=-\infty}^{\infty} -F(q).\exp(2\pi iqbx).\exp\left(2\pi ix\frac{\sin\theta}{\lambda}\right) \qquad (2.5.4)$$

or

$$f'(x) = \sum_{q=-\infty}^{\infty} F(q+r).\exp(2\pi iqxb) \qquad (2.5.5)$$

where,

$$r = \frac{\sin\theta}{b\lambda} \qquad (2.5.6)$$

The objective will only pass spatial frequencies which lie in the range,

$$\frac{-NA_o}{\lambda} < qb < \frac{NA_o}{\lambda} \qquad (2.5.7)$$

Thus the image due to one point from the illumination source is given by,

$$i(x) = \left| \sum_{q=-Q}^{Q} F(q+r).\exp(2\pi iqxb) \right|^2 = \sum_{q=Q_1}^{Q_2} f_s(q).\exp(2\pi iqxb) \qquad (2.5.8)$$

Where $f_s(q)$ is the Fourier series representation of the image due to one point in the illuminating aperture.

where,

$$Q = \frac{NA_o}{b\lambda} \qquad Q_1 = -2Q + 2r \qquad Q_2 = 2Q + 2r \qquad (2.5.9)$$

$$Q_2 < 2 \left[\frac{NA_o}{b\lambda} + \frac{\sin\theta}{b\lambda} \right] \qquad (2.5.10)$$

Now the limit to $\sin\theta$ is determined by the illumination numerical aperture (NA_c).

$$\hat{Q} < \frac{2}{b} \left[\frac{NA_o + NA_c}{\lambda} \right] \qquad (2.5.11)$$

Where $\hat{Q}$ is the highest harmonic in the image profile and thus $b\hat{Q}$ is the cut-off point in the spatial frequency spectrum of the image profile.

If $NA_o = 0.9$, $NA_c = 0.2$ and $\lambda = 0.53\mu m$ then the highest spatial frequency present in the image will correspond to a period of $0.25\mu m$. Thus there is no point in using a focus detection technique which is sensitive to shorter periods than this.

Having determined that there is only a finite frequency spectrum which is of interest when focusing, the behaviour of these frequencies through focus was investigated for four different line objects. The object was taken to have a period of 12μm and a width of 4μm. A 0.85 NA objective, 0.14 NA illuminating aperture and 0.53μm illuminating wavelength were assumed. The four objects were,

a) Dark line with infinite contrast (chrome photomask). R=0, ϕ undefined.

b) R = 1.0, $\phi = \pi$. }   { Thin layer of material such
c) R = 0.4, $\phi = \pi/2$ }   { as oxide or nitride.

d) 0.6μm thick layer of dielectric (refractive index = 1.46) on silicon (refractive index = 4.1+0.06i).

The image of the thick layer object was generated using the waveguide model described in chapter 6. For each object a through focus series of image profiles was generated and the images were transformed into their Fourier series. The amplitude of several of the harmonics as a function of focus has been plotted in figure 2.5.1. As would be expected, the sensitivity to focus increases with harmonic order but there are some peculiar anomalies in figure 2.5.1. The lower order harmonics of objects (b) and (c) do not go through a maximum around focus but some actually go through a minimum. It is only at the very high frequencies where the focal plane can be successfully located on thin layers. However on a thick layer object, none of the harmonics can be relied upon to locate the focal plane. Focusing techniques such as "steepest slope" will be unreliable as they use information from all the harmonics and so the lower harmonics may introduce a focus error.

2.5.2 Linewidth measurement techniques.

According to the Fresnel equations (Hecht and Zajac (1974)), the reflectivity and phase of a thin film on a substrate is a function of thickness, refractive index, illumination wavelength and angle of incidence. If the illumination is contained within a narrow cone angle about the optical axis, then it may be assumed that the illumination is entirely normally incident for the purpose of calculating the reflectivity. If wide angle illumination is used then the reflectivity will not be constant for all points in the illuminating aperture. At wide angles, the reflectivity becomes
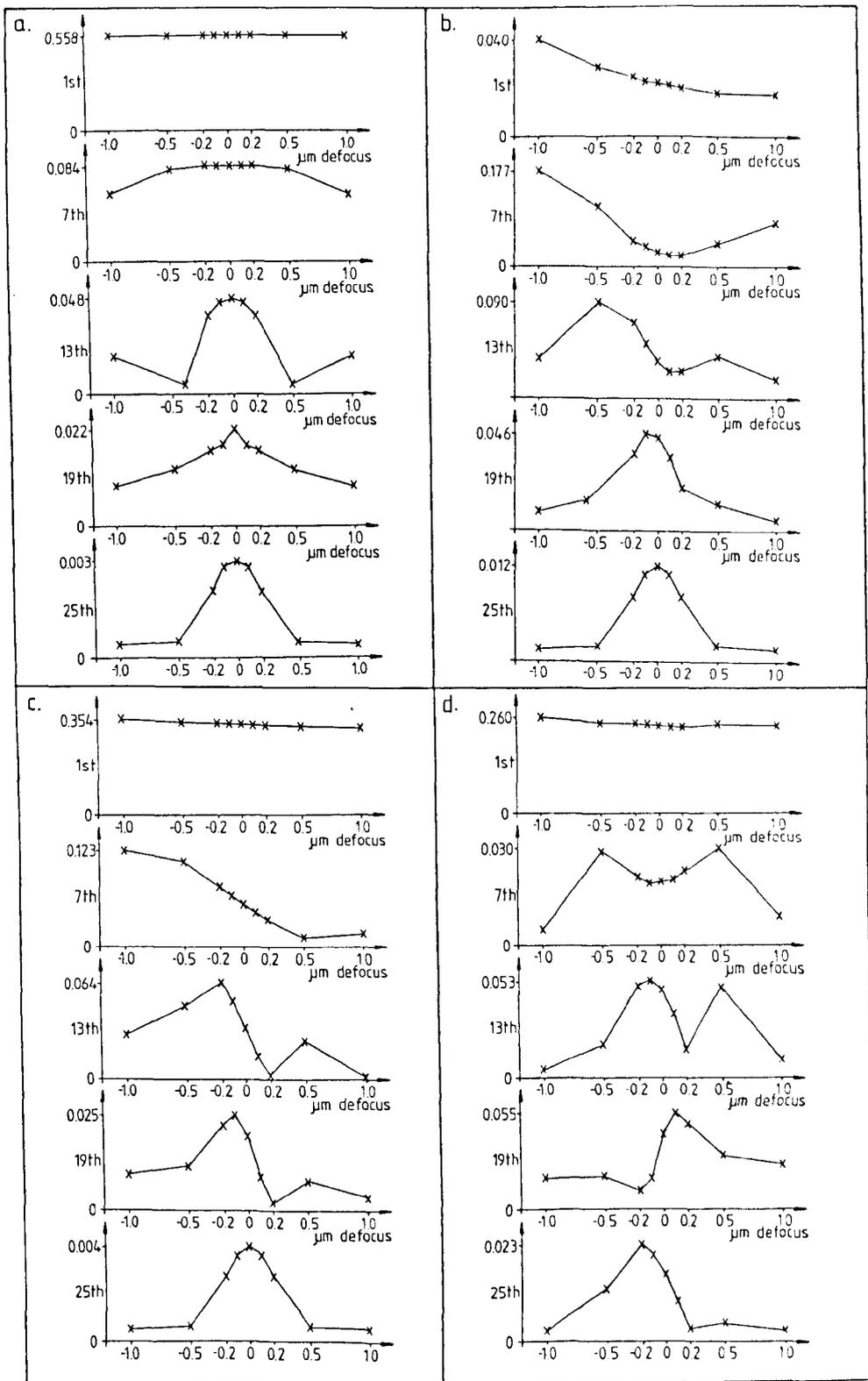
Figure 2.5.1. Through focus amplitude responses of the harmonics in the image profiles of objects (a),(b),(c) and (d).

increasingly sensitive to thickness, wavelength and refractive index. For this reason low numerical aperture illumination is preferred.

The features printed onto wafers will vary slightly in thickness and refractive index and so the reflectivity will vary. This causes lines of the same width to give different optical linewidth measurements. Figure 2.5.2 shows how the linewidth measurement error varies with contrast for an object with no phase change between the line and the substrate when measured at different intensity thresholds. The edge threshold has been defined in two ways. In figure 2.5.2a, the 0% threshold has been defined as the black level and the 100% threshold as the intensity of a large clear area. In figure 2.5.2b however the 0% level has been defined as the intensity of the reflecting line object. It can be seen that this second definition is far less sensitive to changes in reflectivity than the first. Also the curves show that the 50% threshold is slightly less sensitive than the 25% level. The reflectivity is easy to measure when performing a linewidth measurement and so first order corrections to measurements can be made to compensate for changes in relative reflectivity.
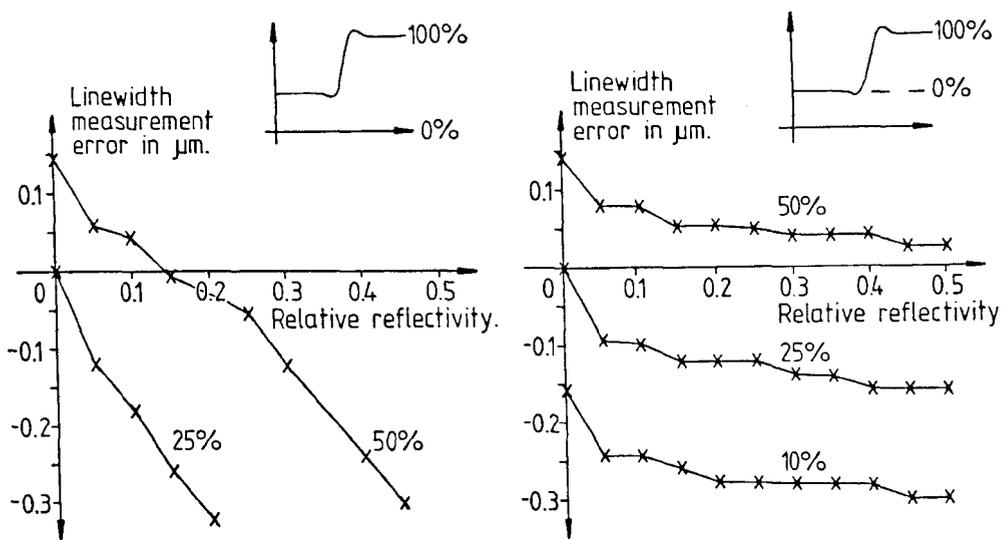
Figure 2.5.2. Linewidth measurement error as a function of contrast. a) black level taken as 0% and b) intensity of the line taken as the 0% level.

As the thickness of a line changes, the relative phase of the reflected light changes along with the contrast. When the phase of the light from the line object differs from that from the substrate by an amount close to $(2n+1)\pi$, then a dark interference fringe is observed close to the line edge. In the case of a zero contrast object, the bottom of this dark fringe will always lie on the true edge regardless of the phase. When the contrast is finite, the bottom

of the fringe will lie close to the line edge. The bottom of this fringe can be used as an edge location criterion for linewidth measurement. Figure 2.5.3 shows how the measurement error varies as a function of phase when using this technique. The line object used in figure 2.5.3 has a relative reflectivity of 0.5. This technique is usable over the range of 0.5π to 1.5π where there is a definite dark fringe at the line edge. Over this range the measurement error will vary by only a few hundredths of a micron. Where there is little phase shift, the dark fringe is much less visible and the error increases rapidly.
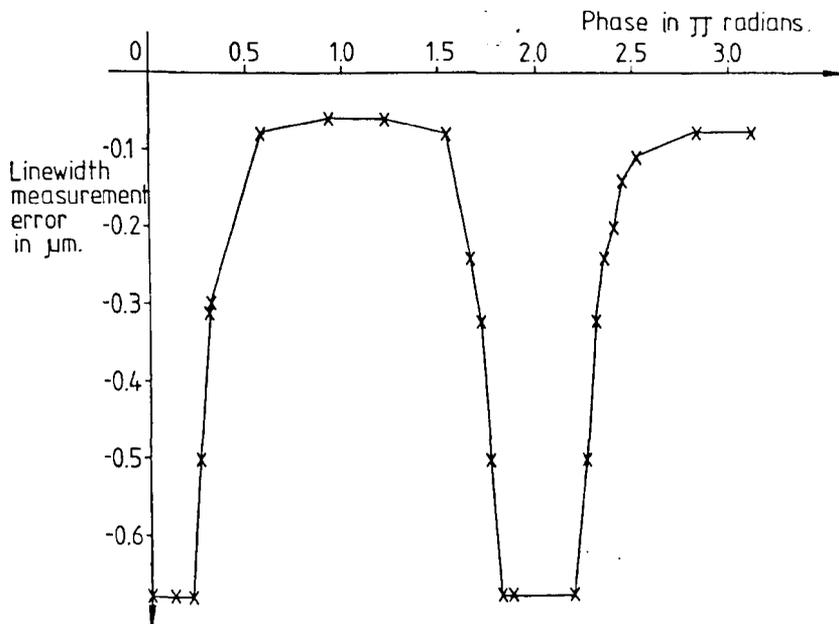
Figure 2.5.3. Linewidth measurement error as a function of phase; using the dark fringe as the criterion for edge location. R=0.5.

Figure 2.5.4 was generated using the same set of image profiles as figure 2.5.3 but the edge is now defined by an intensity threshold. The higher the threshold, the less sensitive it is to phase and although this would result in a linewidth measurement error of a few tenths of a micron, the variation between specimens will be less.

The main cause of variation in phase and reflectivity is small changes in thickness. A line of material with refractive index (n_1) and thickness (t) on a substrate of refractive index (n_s) and surrounded by air (n_o) will generate a reflected wave with amplitude given by,

$$r = \frac{n_1(n_o - n_s)\cos k_o t + (n_o n_s - n_1{}^2) i \sin k_o t}{n_1(n_o + n_s)\cos k_o t + (n_o n_s + n_1{}^2) i \sin k_o t} \qquad (2.5.12)$$
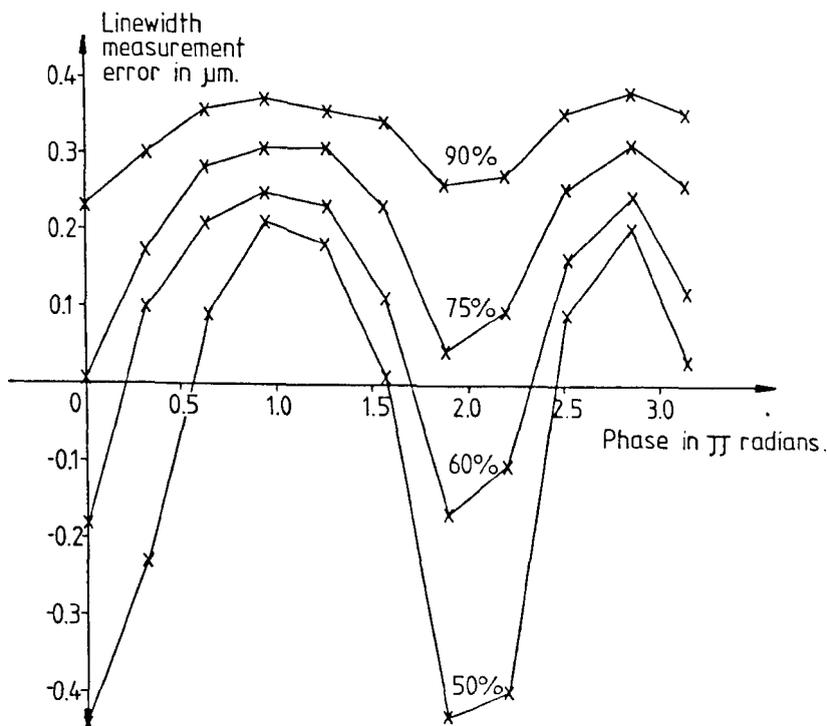
Figure 2.5.4. Linewidth measurement error as a function of
phase; using intensity thresholds to define
the edge of the line. R=0.5.

where,
$$k_o = \frac{2\pi}{\lambda}$$

The reflectivity is determined by the ratio of thickness to wavelength
and so small changes in thickness can be compensated for by small
changes in wavelength. The compensation for thickness will not
produce exactly the same image profile as the diffraction effects and
chromatic aberrations are wavelength dependent. For small changes in
thickness however it is true that provided the ratio of thickness to
wavelength is kept constant, the image profile will change only
slightly.

Consider a white light source between the limits of λ_1 and λ_2. The
image profile for a line of thickness (t) will be given by,

$$I_s(x,t) = \int_{\lambda_1}^{\lambda_2} I(x,\lambda,t) \, d\lambda \qquad (2.5.13)$$

Where $I(x,\lambda,t)$ is the image profile for one wavelength.
If (t) is now allowed to change by a factor of (c) such that the
thickness becomes (ct) then the image profile will be,

$$I_s(x,ct) = \int_{\lambda_1}^{\lambda_2} I(x,\lambda,ct) \, d\lambda \qquad (2.5.14)$$

Since (t) and (λ) are in direct proportion for a given profile then the limits may be changed and written in terms of equation 2.5.13.

$$I_s(x,ct) = \int_{\frac{\lambda_1}{c}}^{\frac{\lambda_2}{c}} I(x,\lambda,t) \, d\lambda = I_s(x,t) - \int_{\lambda_1}^{\frac{\lambda_1}{c}} I(x,\lambda,t) \, d\lambda + \int_{\lambda_2}^{\frac{\lambda_2}{c}} I(x,\lambda,t) \, d\lambda$$

$$(2.5.15)$$

Provided that (c) is close to unity and λ_1 and λ_2 are not too close together (ie. the bandwidth is not too narrow), then the two terms on the far right hand side of equation 2.5.15 will be small and so the image profile will not change significantly for a small change in thickness. This in turn will mean that the linewidth measurement is much less sensitive to changes in thickness. The use of broad band illumination will help to improve repeatability between specimens but it makes it harder to relate the linewidth measurements to accurate values should this be required.

CHAPTER THREE

Analysis of Measurement Techniques for Planar Objects.

3.1 Review of measurement techniques.

The different types of linewidth measurement systems available have been reviewed in the literature (Nyyssonen (1980), Moore and Kirk (1984)) and a considerable amount of comparative measurement work has been done (Swyt and Rosberry (1977), Jerke and Wendell (1982), Croarkin and Varner (1982), Jerke et al. (1982)). The performance of many commercially available systems is determined as much by engineering limitations as it is by the physical principles behind the technique. For this reason only techniques will be reviewed here and not particular systems.

3.1.1 Image scanning.

The image scanning technique is the principle behind by far the most linewidth measurement systems currently in use (figure 3.1.1a). An area of the specimen which is large compared with the point spread function of the optical system is illuminated and the intensity distribution in the image plane is captured and used for measurement. There are several different methods used for capturing the image and these include;

i.   Scanning the image with a pinhole or slit in front of a photo-detector.

ii. Keeping the detector and slit stationary and scanning the object.

iii. Using a video camera or charge coupled device (CCD) array of photodiodes to generate a scanned signal of the image.

The method chosen for capturing the image will depend on the application for which the system is required. Video techniques are the most convenient but suffer from poor resolution and severe distortions. Scanning the image is slower and more sensitive to vibration but enables a more accurate image profile to be obtained. The effects of aberrations in the optical system can be minimised by keeping the detector aperture on axis and scanning the object. Although this technique produces the least distorted image profile, it
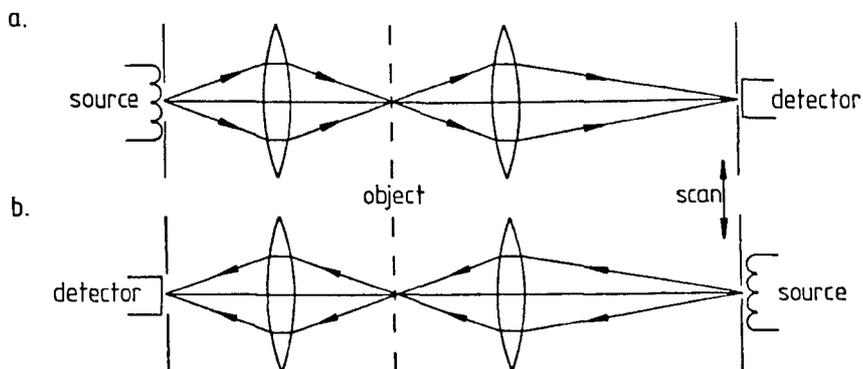
Figure 3.1.1. a) Conventional image scanning microscope.

b) Inverted or object scanning microscope.

is very sensitive to vibration and requires very accurate measurement of the object displacement.

A slight variation of the image scanning technique was presented by Chisholm (1984). A conventional bright field image scanning measurement technique was used but instead of using the reflected bright field, the microscope was operated in fluorescence mode. This has the advantage of producing nearly incoherent illumination and the effects of nearby non-fluorescing sites are negligible. It is of course restricted to specimens where the feature of interest fluoresces.

Most image scanning measurement systems operate in bright field, but other techniques such as placing a central stop in the objective aperture (Young (1983)) or dark field illumination (Kirk et al. (1984)) have been tried.

3.1.2 Object scanning.

According to the analysis presented in section 2.3, it is possible to invert the optical system and still observe the same image (figure 3.1.1b). This is known as the "principle of equivalence" and has been discussed in the literature (Kermisch (1977), Sheppard (1980)). The equivalence is based on the assumption that the far field diffraction patterns produced by two illumination angles are merely displaced versions of each other (equation (2.3.14)). Nyyssonen (1982c) investigated this principle of equivalence and demonstrated that the images produced by the systems shown in figures 3.1.1a and 3.1.1b were not the same when the object had a finite thickness.

The object scanning technique requires knowledge of the angular dependence of the image profile and so the object can no longer be represented by just its phase and amplitude distribution.

3.1.3 Far field scanning.

In a conventional bright field measurement system the image exists as a diffraction pattern in the back focal plane. All the information which appears in the final image profile is present in this plane and may be used for measurement.

Guillaume et al.(1984) presented another method based on computing the linewidth from the separations between the zeros in the Fourier transform of the image profile.

Accuracy and repeatability are improved when the line is wide and a large number of zeros are produced. As the linewidth decreases there are fewer zeros in the Fourier transform and for small linewidths there are no zeros. The technique is only applicable to planar objects and does not readily extend to thick films.

3.1.4 Image shearing.

The technique of image shearing is an extension of the image scanning measurement approach. Dyson (1960) described a method of measurement which involved splitting the image into two and displacing the two images until they just touch at the edges. By measuring the displacement distance a measure of the object width can be obtained (figure 3.1.2).
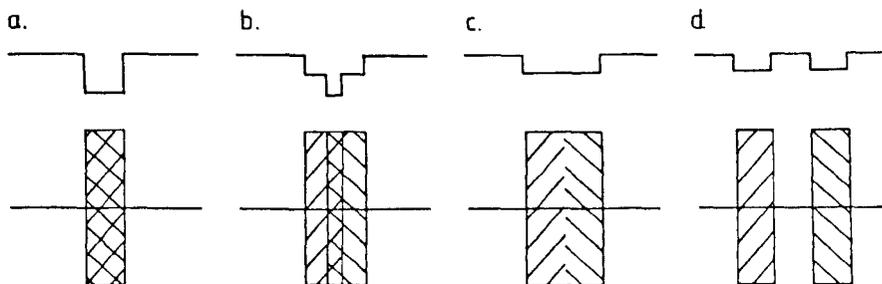
Figure 3.1.2.   Linewidth measurement using image shearing.
a) Image of a line object, b) under shear,
c) exact shear and d) over shear.

Unfortunately where there are fringes at the edges of the feature, the subjective edge setting becomes difficult. Smith (1984) described the historical development of image shearing and the development of coincidence setting shear (CSS) (Smith (1978)). In this technique a strip of the image is displaced relative to the rest of the field of view. This enables features such as dark fringes to be used as verniers and set against each other (figure 3.1.3).
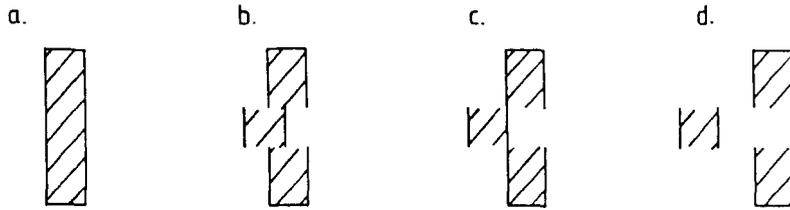
Figure 3.1.3. Linewidth measurement using coincidence setting shear.
a) Image of a line object, b) under shear,
c) exact shear and d) over shear.

3.2 Image transducer models.

In most measurement systems and in particular for those where automation is required, it is necessary to be able to convert the intensity distribution in the image plane into an electrical signal. This involves scanning the image with an aperture and converting the light which passes through into an electrical signal. If the aperture amplitude response is defined by T(x,y) and the image by I(x,y) then the scanned image will be given by,

$$I'(x,y) = \int\int_{-\infty}^{\infty} T(X,Y) . I(x-X,y-Y) \quad dX \, dY \qquad (3.2.1)$$

This represents the light falling on the detector surface. The photo-detector will not necessarily have a linear response and its non-linearity is usually measured by the parameter γ. This results in a response (v) given by,

$$v = c + I^{\gamma} \qquad (3.2.2)$$

c = black level response (dark current).

$$v(x,y) = c + \left[\int\int_{-\infty}^{\infty} T(X,Y) . I(x-X,y-Y) \quad dX \, dY \right]^{\gamma} \qquad (3.2.3)$$

This assumes that the transducer response is isoplanatic which for

some devices such as video cameras is not strictly true.

When the image is scanned there will be some error in tracking an accurate scan path and the deviation from the path may be defined in the x and y directions as $e_x(x,y)$ and $e_y(x,y)$. This is known as geometric distortion.

$$v(x,y) = c + \left[\int\int_{-\infty}^{\infty} T(X,Y).I(x-e_x(x,y)-X,y-e_y(x,y)-Y) \, dXdY \right]^Y$$

(3.2.4)

These three sources of error account for most of the distortion to the image profile caused by the transducer.

3.2.1 Pinhole or slit scanning.

A widely used technique for image conversion is to place a pinhole (or slit if the object is one dimensional) in front of a photo-detector and to scan either the slit or object mechanically. This technique has the advantage that $T(x,y)$ is defined by the physical shape of the aperture and is thus constant and so $T(x,y)$ is isoplanatic. The use of a good quality photomultiplier tube as the photodetector enables extremely good photometric linearity to be obtained and very narrow slits to be used. The main source of error with this technique is the geometric distortion which will include vibration and this is entirely dependent upon mechanical design.

3.2.2 Video cameras.

The raster scanned video camera is an attractive transducer to use in a measurement system. The image is captured in a very short time reducing vibration problems and a large field of view can be captured and displayed without using complex mechanical scanning arrangements. Unfortunately most video cameras suffer from poor resolution and considerable geometric distortion. Worse still, the parameters $e_x(x,y)$ and $e_y(x,y)$ and $T(x,y)$ are functions of time, temperature, light intensity and the electrical drive conditions of the tube, thus making correction difficult.

The point spread function ($T(x,y)$) of most video camera tubes is a radial Gaussian profile (Kirk et al.(1983)).

$$T(x,y) = (2\pi\sigma^2)^{-1} \cdot \exp\left(\frac{-r^2}{2\sigma^2} \right) \qquad (3.2.5)$$

Generally the transducer is only required to measure line objects and so the line spread function of the tube is more useful. Expressing equation 3.2.5 in cartesian coordinates and reducing to one dimension gives,

$$L(x) = \int_{-\infty}^{\infty} (2\pi\sigma^2)^{-1} \cdot \exp\left(\frac{-(x^2+y^2)}{2\sigma^2} \right) \, dy \qquad (3.2.6)$$

$$L(x) = (2\pi\sigma^2)^{-0.5} \cdot \exp\left(\frac{-x^2}{2\sigma^2} \right) \qquad (3.2.7)$$

Hence the radial Gaussian distribution of the point spread function reduces to a linear Gaussian distribution of the same width for the line spread function.

Sometimes it is more convenient to express this response in the frequency domain as a function of the spatial frequency (u).

$$\tilde{L}[u] = \exp\left[\frac{-u^2\sigma^2}{2} \right] \qquad (3.2.8)$$

Figure 3.2.1 shows comparisons of the measured response of a video camera tube line spread function and the response to a step edge with the Gaussian model (σ = 2.2 scan lines). The results show that this model is a good approximation. A newvicon tube has been considered as they have very good photometric linearity.
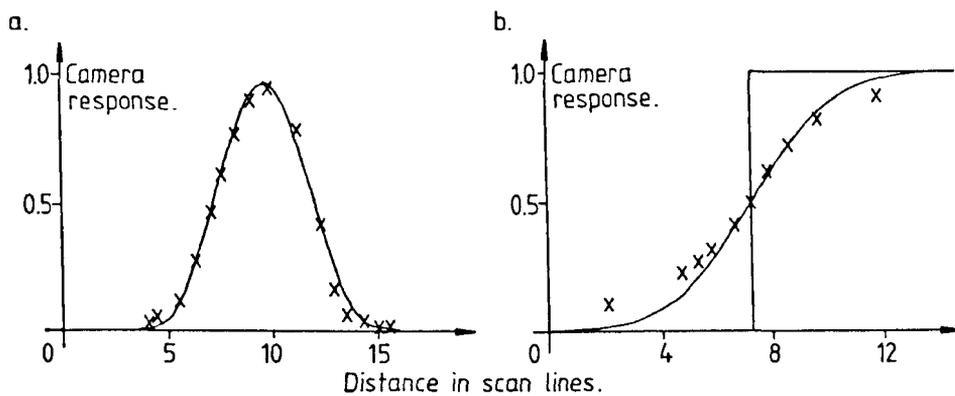
Figure 3.2.1. Comparison of measured video camera responses with Gaussian models for a) line spread function and b) step edge response. National Panasonic (1350AE/B) camera fitted with a 2/3" S4075 newvicon tube. 625 line raster scan with 25Hz frame rate.

The Gaussian model is only an approximation and the exact point spread function is dependent on the image structure and intensity as well as the operating conditions of the tube (Franken (1980)). This makes it difficult to correct for the camera response as it is continually varying. For this reason measurement techniques should be designed to be insensitive to the camera performance.

3.3 Theoretical study of measurement errors.

Although the models described so far are applicable to all images, in this section only the images of high contrast photomasks will be considered.

3.3.1 Illumination bandwidth.

In the results presented so far, the illumination has been assumed to be monochromatic. In practice because of the limited light available from a tungsten halogen source a broadband filter is usually used. Figure 3.3.1 shows the spectral responses of a typical broad band green filter and a typical camera tube. The overall response shows that it is the filter which controls the bandwidth.
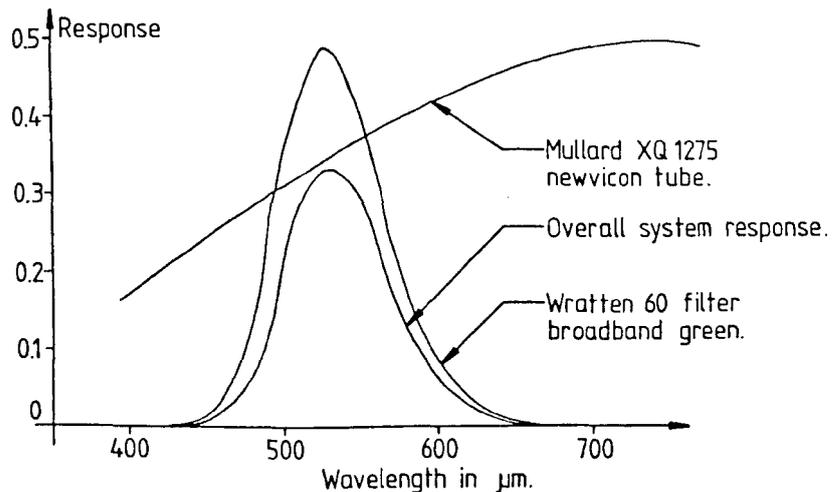
Figure 3.3.1. Spectral response of a broadband green filter and a typical video camera pick up tube (newvicon).

In order to investigate the effect of bandwidth, three illumination conditions were considered for equation 2.3.21. The image profiles for monochromatic ($\lambda = 0.53\mu m$), white light ($0.45\mu m < \lambda < 0.63\mu m$) and the broadband response of figure (3.3.1) were computed for a 0.9 NA objective and 0.6 NA condenser. The location of both the true edge

threshold and the 50% intensity threshold varied by less than 0.005μm for the different bandwidth conditions. Slightly larger variations were detectable at the higher intensities.

Nyyssonen (1977) reported that theoretical coherent image profiles were insensitive to bandwidth but also reported that this was not true in practice. She attributed the disagreement between the white and green light images to chromatic aberration. The results of this work show that theoretical partially coherent images are also quite insensitive to bandwidth.

The objects considered here are of course ideal, with infinite contrast and so it may be concluded that the diffraction effects are not particularly sensitive to illumination bandwidth. However when the object is a film, then the relative reflectance and phase will be wavelength sensitive and so the need to restrict illumination bandwidth arises from the nature of the object and the effects of lens aberrations.

3.3.2 Glare.

Stray light or glare was discussed in section 2.2. The amount of glare in an image is a function of the object geometry and it varies across the field of view. Consequently it is difficult to determine the exact relative reflectance of an object and this introduces an error when computing the edge threshold. Even if the true edge threshold is known then there will be an error in locating it unless the level of glare at the edge is known and compensated for. In the case of a 0.9 NA objective, 0.6 NA condenser and an infinite contrast object, the linewidth measurement will be in error by 0.04μm if there is 5% uncompensated glare.

3.3.3 Transducer aperture width.

The effective width (w´) of a long slit in front of photomultiplier tube in an image scanning system with a small angular misalignment is given by,

$$w´ \approx w + L.\sin\theta \qquad (3.3.1)$$

θ = angle between object lines and slit.
L = slit length.

w = slit width.

Wide slits collect more light but result in lost resolution of the image profile and so the slit width should be made as large as possible without having any significant blurring effect on the image. Nyyssonen (1977) recommended an effective slit width of less than one sixth of the Airy disc diameter (1/3 of the Rayleigh resolution limit). According to the model in equation 3.2.1 the slit width may be up to twice this value before the measurement changes by as much as 0.01μm. If the slit is made 1/6 th of the Airy disc diameter and is 10 times longer than it is wide then the slit and the line objects must be parallel to within ±5°.

In section 3.2.2 it was shown that a video camera point spread function can be modelled by a Gaussian distribution. The effective width of this response is dependent on magnification which in turn is limited by the required field of view. If it is required to provide a field of view of up to 50μm wide then the camera in figure 3.2.1 would produce an effective point spread function width (σ) of 0.16μm. The curves in figure 3.3.2 show how the camera width parameter (σ) effects the image profile (equation 3.2.1). The line object has been taken as having 4% background transmission with no phase shift and this has the effect of raising the edge threshold from 25% to 36%. The poor resolution of the video camera causes the true edge threshold to rise towards 52% (50% of the difference between the bright and dark levels). This 52% threshold is least sensitive to camera resolution.
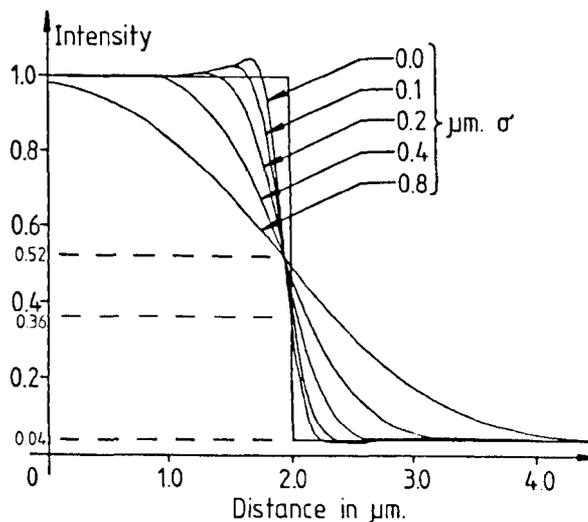
Figure 3.3.2. The effect of the camera point spread function width on the image of a step edge.

If it is required to measure directly at the true edge threshold then, in order to keep the measurement error to less than 0.01μm, σ must not exceed 0.08μm.

3.3.4 Photometric non-linearity.

Many video cameras such as those with vidicon tubes have non-linear photometric responses and in others the γ is deliberately reduced in order to compress the dynamic range. γ values as low as 0.7 are not uncommon and, if not corrected for, will distort the edge profile and introduce an error into the linewidth measurement.

In the case of the 0.9 NA objective and the 0.6 NA condenser with an infinite contrast line object, a γ of 0.7 will move the true edge from 25% to 38%. The threshold which now appears to be 25% will in reality be only 14% and this will have the effect of introducing an error of 0.08μm into the measurement of the linewidth.

In order to keep the linewidth measurement error to within ±0.01μm, γ must lie between 0.95 and 1.05. This is readily achievable for a photomultiplier tube but it is a fairly strict specification for a video camera.

3.3.5 Geometric distortion.

The geometrical distortion of a mechanical scanning system will depend on its design. Many scanning problems can be eliminated by recording the position of the sensor rather than relying on a linear scan.

In a video camera it is not possible to measure the beam position and so geometric distortion can be a problem. It is possible to attempt to correct for geometric distortion by imaging a grid pattern and then constructing an error map for the scan. Unfortunately the geometric distortion is a function of time, temperature, intensity, image structure and the electrical drive conditions.

Manufactures quote figures of around ±2% for geometrical distortion and for a 50μm field of view this would give a non-linearity error of up to 2μm between the top and the bottom of the field. When measuring a 5μm line this could result in a linewidth measurement error of up to 0.2μm. In order to maintain accuracy of ±0.01μm on a 5μm line, the geometric distortion would need to be less than 0.1%.

Sharp edges introduce localised geometric distortion. This is caused by the electron beam bending as it approaches or leaves a charged area (Franken (1980)). This means that the localised distortion may be higher than the global values quoted by manufacturers and since it is a function of the object it cannot be corrected for.

Video cameras also suffer from shading. This is a variation in sensitivity across the field of view. Shading of up to 10% is common and this will lead to a significant error if a threshold is computed from bright and dark levels in a region of different sensitivity.

3.3.6 Image shearing.

If an image is split into two portions and sheared by an amount (p) then the resultant image is given by,

$$I_s(x) = 0.5 \left[I(x-p) + I(x+p) \right] \tag{3.3.2}$$

Provided that the object is symmetrical the two images will intersect at the same intensity threshold. The eye is a poor photometer and is unable to judge the intensity at the point of intersection. However the eye is very sensitive to rapid changes in intensity such as bright or dark fringes and so it is possible to set the two images very precisely so that they merge into one.

The image of an incoherently illuminated step edge has rotational symmetry about the point half way between the light I_p and dark I_m areas. Let the true edge lie at x=0.

$$I(-x) = I_p + I_m - I(x) \tag{3.3.3}$$

If two opposing edges are aligned then the sheared image is,

$$I_s = 0.5 \left[I(-x) + I(x) \right] \tag{3.3.4}$$

$$I_s(x) = 0.5 \left[I_p + I_m - I(x) + I(x) \right] = \frac{I_p + I_m}{2} \tag{3.3.5}$$

This means that when the edges are overlapping at the true edge then the image profile across the true edge is perfectly flat. This only holds for incoherent images of step edges. Equation 3.3.3 does not hold for partially coherent images.

Figure 3.3.3 shows the image profile across a partially coherently illuminated line with opposing edges set to around the "just touch" shear position. The profile never actually goes flat but the smoothest curve occurs when the two edges overlap at around the 52-54% intensity thresholds. This results in a theoretical linewidth measurement error of 0.15µm. The intensity at the overlap is twice the value of the intensity of each of the images at that point.
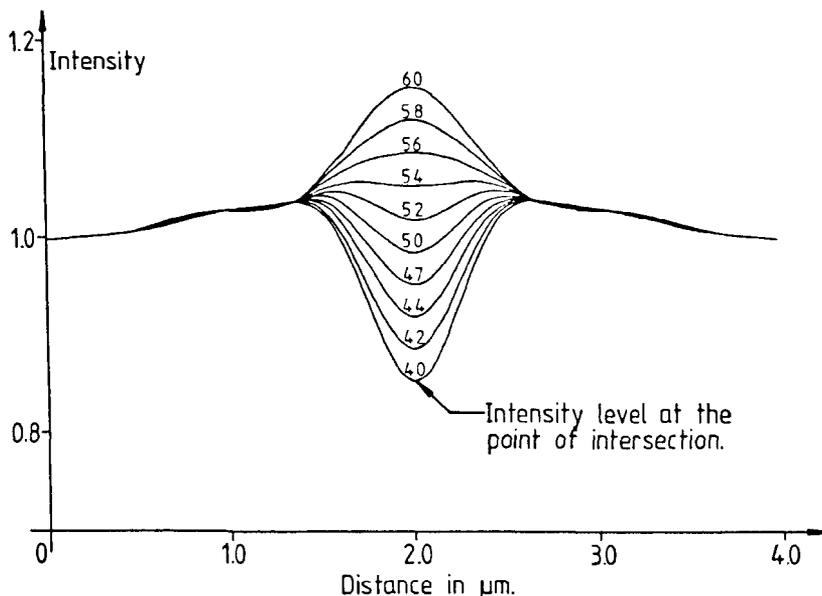
Figure 3.3.3. Image profiles of sheared partially coherent images of a line object set to different intensity levels. (0.9 NA objective and 0.6 NA condenser)

Instead of relying on the eye to perform the edge to edge setting, the sheared image may be relayed from a video camera. A display of the video waveform will then allow setting to any intensity threshold. The video signal will be given by equation 3.2.4 but the image I(x,y) is now the sheared image $I_s(x,y)$ in equation 3.3.2.

In order to perform a measurement to a particular threshold, the shearing is adjusted until the dip between the images falls to the correct intensity level. Since the measurement is performed by measuring the distance sheared and not from the video signal, geometric distortion will not affect it. Similarly if the line objects are first under-sheared and then over-sheared in order to set the bright and dark levels, then the pick up tube need not be isoplanatic as the measurement is being performed at only one site on the tube face.

Photometric non-linearity and resolution will affect the measurement. The effect of a non-linear response will be to cause the wrong edge threshold to be located. However by using a camera with a linear response this can be overcome.

The limited resolution of the video camera will blur the fringe produced at the overlapping edges and will result in the wrong threshold being located. Figure 3.3.4 shows the video signal which is produced by the image profiles of figure 3.3.3. The camera width parameter (σ) has been taken as 0.16µm which is typical for the magnifications used in most systems. It can be seen that the intensity level at the point of intersection is no longer exactly twice the intensity level of each edge.
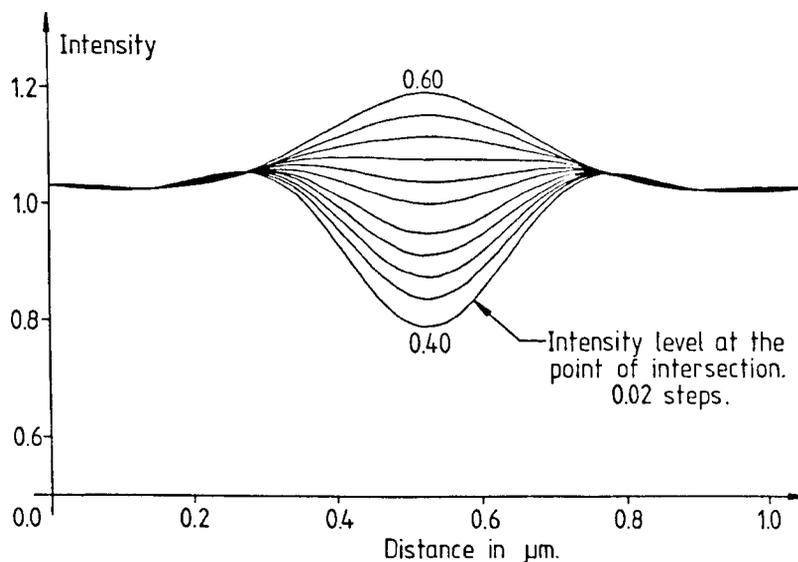
Figure 3.3.4. Image profiles from figure 3.3.3 after convolving with the point spread function of a video camera.

The smoothest profile still occurs for intersections around the 52-54% intensity level and so this edge setting criterion is unaffected by resolution. Since the profile is roughly constant under this condition then it will not be affected by photometric non-linearity either.

The video camera reduces the amplitude of the fringe produced by a shear error but it does not alter the threshold at which the shear is minimised. Thus the limited resolution does not introduce a systematic error but merely reduces the precision.

If however the technique is used to locate other thresholds (Downs and Turner (1983)) then the video resolution will introduce an error. Figure 3.3.5a shows how the camera resolution width parameter (σ) affects the location of the edge threshold. If the camera resolution is ignored then the wrong threshold is located. When attempting to set to the 25% threshold, the threshold which is actually located is given by curve (ii). The effect of setting to this incorrect threshold is to introduce a linewidth measurement error and this is given as a function of the camera resolution width parameter in figure 3.3.5b. Where the camera resolution is known, instead of setting to the 25% intensity threshold, the correct threshold can be computed and located (curve i figure 3.3.5a). In order to keep the linewidth measurement error to within 0.01µm using this technique, σ must be less than 0.05µm.
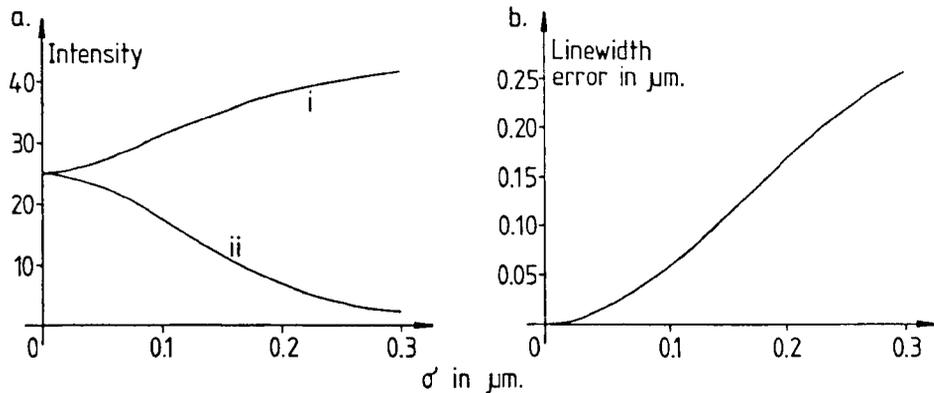
Figure 3.3.5.   a) For a given camera resolution width the shearing should be set to the threshold given by curve (i) in order to allow for the camera resolution. If the camera resolution is ignored and the 25% threshold used then the actual threshold will be given by curve (ii). This would result in the linewidth measurement error given by the curve in (b).

3.3.7 Coincidence setting shear (CSS).

Measurements by eye using the CSS technique enable the distances between the peaks of bright or dark fringes to be measured. The measurement method may be extended to images other than those with fringes by detecting thresholds in the image profile and using these for alignment. This is then equivalent to an image scanning system except that it enables a video camera to be used without incurring

errors from geometric distortion. This is because the measurement is made by performing and measuring the mechanical displacement of the images and the video signal is merely used to detect a null setting.

Resolution and photometric non-linearity will still effect the measurement in the same way as for a conventional image scanning system.

3.4 Practical investigation into linewidth measurement techniques.

The results of linewidth measurement surveys have appeared in the literature (see section 3.1). This work has concentrated mainly on the repeatability of the techniques and methods of determining a calibration curve for any one system using a calibration standard. There has been little work published which attempts to examine how well experimental data agrees with the models for linewidth measurement techniques.

3.4.1 Practical Results.

Several linewidth measurement techniques were tested on a photomask standard from the National Physical Laboratory (NPL)†. The standard was used to calibrate the instruments from pitch measurements and the different techniques were used to measure the width of ten clear lines on the pattern. These lines have nominal widths from 1μm to 10μm. The following measurement techniques were used.

a) Dark ground - Using the coincidence setting shear (CSS), part of the dark ground image was sheared relative to the rest. The setting from edge to edge was performed by eye, as the video camera was not sensitive enough to pick up the image.
    80X 0.85NA dark ground objective used in white incident light with a 10X eyepiece.

b) Image shearing by eye - Conventional image shearing through the eyepieces of the microscope, setting from edge to edge by eye.
    100X 0.9NA objective, 10X eyepieces, 0.6NA substage condenser and green filter, set up for Köhler illumination.

_____

† NPL Photomask Linewidth Standard No.24, 1983.

c) Image shearing by TV - Conventional image shearing using the video waveform for setting from edge to edge.
Optical conditions as in (b), but with 3X camera lens.

d) Image shearing to 25% by TV - Image shearing as in (c) but using the video waveform to set to the 25% point instead of the 50% point.
Optical conditions as in (c).

e) TV threshold detection 25% - Using coincidence setting shear (CSS) of the bright field image and an electronic system of threshold detection in the video signal it is possible to measure between the 25% thresholds in the video image.
Optical conditions as in (c).

f) TV threshold detection 50% - The same as (e) but measuring between the 50% intensity thresholds in the video image.
Optical conditions as in (c).

The results from these techniques are shown in table 3.4.1. Each measurement entered in the table is the mean of ten readings. The columns on the right hand side show the measured differences between different techniques.

On wide lines the dark ground measurement error is small according to the theory. If wide band illumination and circular apertures are used then it should be smaller than that predicted in section 2.4.3. The dark ground measurements have been taken as the reference in order to compare the offsets between the different techniques.

Using the models developed in section 2.3, the image shearing technique should give measured linewidth values which are $0.14\mu m$ too small. This is because the technique does not measure at the true edge intensity threshold. The use of the video signal to determine the edge setting should not change the image shearing measurement (section 3.3.6).

When the video signal is used with image shearing to locate other thresholds then an error is incurred due to the video camera resolution. In this particular case the measured linewidths should be $0.07\mu m$ larger than the true value when attempting to measure between the 25% thresholds.

Track number	a Dark ground	b Image shearing by eye.	c Image shearing by TV.	d Image shearing to 25% by TV.	e TV Threshold detection 25%.	f TV Threshold detection 50%	g c-b	h f-c	i d-a	j e-a	k a-c	l d-c	m e-f
1	1.16	0.98	0.99	1.20	1.45	1.11	0.01	0.12	0.04	0.29	0.17	0.21	0.34
2	2.18	2.01	2.01	2.21	2.49	2.05	0.00	0.04	0.03	0.31	0.17	0.20	0.44
3	3.21	3.09	3.09	3.27	3.54	3.10	0.00	0.01	0.06	0.33	0.12	0.18	0.44
4	4.25	4.11	4.11	4.31	4.55	4.12	0.00	0.01	0.06	0.30	0.14	0.20	0.43
5	5.36	5.22	5.22	5.45	5.65	5.22	0.00	0.00	0.09	0.29	0.14	0.23	0.43
6	6.31	6.20	6.21	6.41	6.64	6.19	0.01	-0.02	0.10	0.33	0.10	0.20	0.45
7	7.32	7.19	7.19	7.41	7.59	7.17	0.00	-0.02	0.09	0.27	0.13	0.22	0.42
8	8.32	8.20	8.21	8.41	8.60	8.19	0.01	-0.02	0.09	0.28	0.11	0.20	0.41
9	9.33	9.24	9.25	9.43	9.64	9.22	0.01	-0.03	0.10	0.31	0.08	0.18	0.42
10	10.37	10.28	10.28	10.49	10.68	10.24	0.00	-0.04	0.12	0.31	0.09	0.21	0.44
MEAN							0.00	0.00	0.08	0.30	0.13	0.20	0.42
THEORETICAL VALUES							0.00	0.00	0.07	0.30	0.14	0.21	0.44

Table 3.4.1. Comparison of different photomask linewidth measurement techniques.

In section 3.3.3 it was shown that the measurement between the 50% intensity thresholds was virtually unaffected by the video camera resolution but when the measurement was performed between the 25% intensity thresholds a significant error is incurred. For the conditions used here this results in bright lines being measured 0.3μm wider than the true width.

Table 3.4.1 shows that the measured offsets show good agreement with the values predicted by the theory. This suggests that the models are reasonably accurate.

Although the agreement between the measurements is encouraging, there are unexplained discrepancies between the theoretical and practical results. Two of the unresolved differences are,

i. For a 0.9 NA objective and 0.6 NA condenser (λ = 530 nm), the difference between measurements of lines between the 50% and 25% intensity thresholds should be 0.14μm. In practice even when the optical image is scanned with a narrow slit, the difference is often observed to lie between about 0.20 and 0.26 μm, depending on the photomask.

ii. The intensities of the bright lines at either side of a track in the dark ground image should be equal. In practice when the illumination is from one side only, these bright lines have been observed to have considerably different intensities (Kirk et al. (1984))† and the extent to which they differ varies from mask to mask and with different illumination polarisation angles.

The discrepancy between the measured and theoretical 25-50% differences requires further examination. When using video techniques, some of this error may be traced to the camera resolution but this does not explain the variation between photomasks. Nyyssonen (1977) has reported that good agreement between experimental and theoretical image profiles is possible. This suggests that the discrepancies may be partly due to aberrations and partly due to mask quality.

---

† It was reported to the author by Dr. F.H. Smith at Vickers Instruments, when one half of the annular aperture used for dark ground illumination was obscured, the two bright lines along the edges of the gap in chrome were of unequal intensity.

3.4.2 Investigation into the effect of the video camera resolution.

In section 3.3.6 it was demonstrated how the limited resolution of the video camera affected the linewidth measurement when used with image shearing. If the camera resolution is improved ($\sigma \rightarrow 0$) then the 25-50% difference should decrease and tend towards the asymptotic value of 0.14µm. The resolution is most easily controlled by changing the magnification of the image on the pick-up tube but the limiting value corresponding to $\sigma=0$ cannot be achieved as it requires infinite magnification. However significant improvements in σ are possible with only modest increases in magnification.

Figure 3.4.1 shows a comparison of practical measurement errors with the theoretical error curve as σ is varied (achieved by changing the magnification). The practical error values are relative to the linewidths quoted by NBS‡ and NPL for their linewidth standards. The practical results demonstrate the trend predicted by the theoretical error curve and this suggests that the video camera does indeed introduce a convolution error. However whereas the theory predicts that clear lines are being measured too wide, when the magnification is increased this error does not decrease as quickly as predicted and when extrapolated it suggests that the NBS lines are about 0.04µm narrower than claimed and the NPL lines are about 0.05µm narrower than claimed. This is not consistent with results obtained by scanning the image with a slit and suggests that there is an additional source of error.

Comparing 25% and 50% measurements of clear lines on various masks showed that the difference between them could be as low as 0.18µm when high magnification was used (no convolution error). This 25%-50% value of 0.18µm is still significantly higher than the theoretical value of 0.14µm and worse still it varies from mask to mask.

3.4.3 Optical conditions.

So far it has been assumed that the optics are free from aberrations and that the edge of the line object is ideal. In order to investigate this, a series of lenses was used with different numerical apertures.

‡ NBS AR Chromium Photomask Linewidth Standard, AR006-1, 1980.
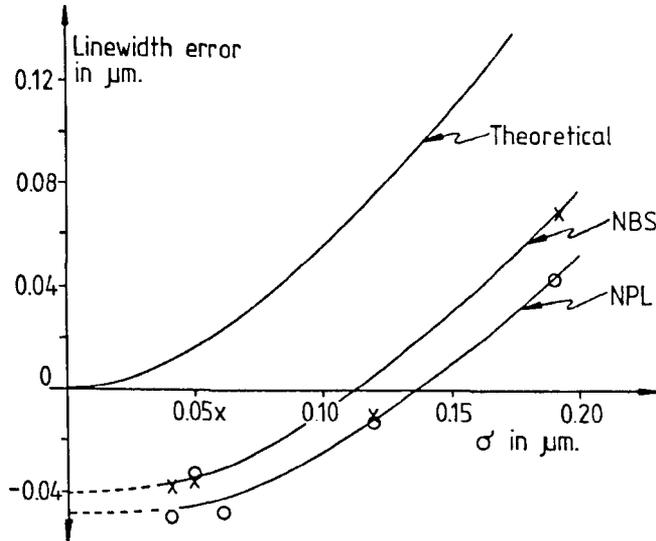
Figure 3.4.1.   The effect of camera   resolution on the   linewidth
                 measurement   error when the   video   profile of the
                 sheared images is used to set to the 25% intensity
                 threshold.

If   the   numerical   aperture   of   the   objective   is   reduced   without
decreasing   the   numerical   aperture   of   the   condenser,   then   the
illumination   becomes   less   coherent   and   the   true   edge   intensity
threshold   should   tend   towards   50%.   Thus   if   the   50%   intensity
threshold   is   used   as   the   measurement   criterion   (eg.   image   shearing),
then   the   linewidth   measurement   error   should   tend   to   zero.   If   however
the   condenser   aperture   is   reduced   with   the   objective   aperture   in   order
to   maintain   a   coherence   parameter   (s)   of   2/3   then   the   linewidth
measurement   error   between   the   50%   thresholds   will   increase   as   the
resolution   has   decreased.

The   first   experiment   investigates   the   effect   of   the   objective   aperture
on   the   linewidth   measurement   error   while   keeping   the   illumination
angle   constant.   The   second   experiment   investigates   the   effect   of   the
illumination   angle   while   keeping   the   coherence   parameter   constant.
The   results   are   shown   in   figure   3.4.2.

The   measurements   were   performed   on   the   NPL   and   NBS   linewidth
calibration   standards   and   the   errors   are   quoted   relative   to   the
calibration   values.   In   the   first   experiment   where   the   condenser
aperture   remains   constant,   the   measurement   error   does   not   fall   to   zero
as   the   objective   numerical   aperture   decreases.   Instead   it   falls   away
slowly   and   tends   towards   a   fixed   value.   The   sudden   rise   in   the   curve
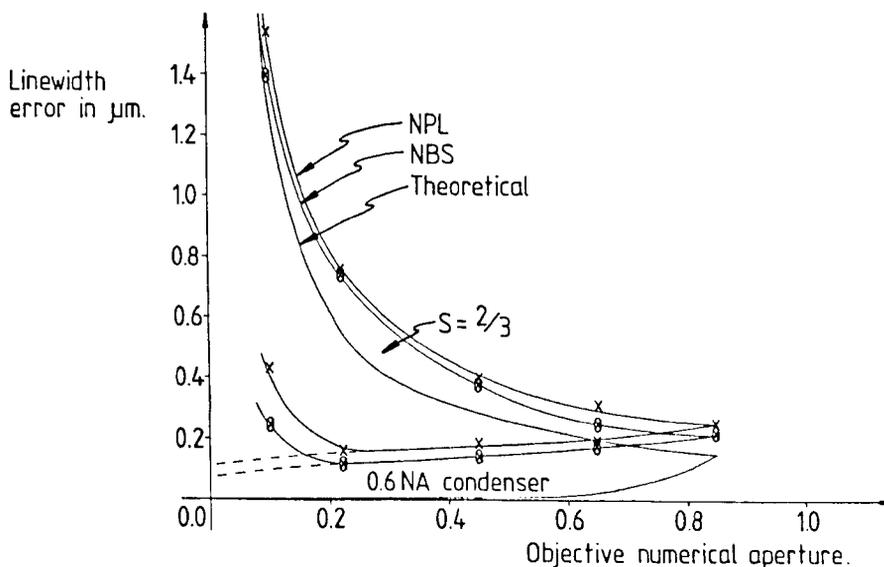
Figure 3.4.2. The effect of numerical aperture on linewidth
error measurements using image shearing to 50%.

for the measured results below 0.2 NA is due to interference between
adjacent line edges.

In the second experiment where the coherence parameter is kept
constant, the difference between the measured and theoretical errors
remains roughly constant but the ratio of this difference to the
theoretical error decreases with the numerical aperture. This
suggests that there is an effect due to the line edge but that this is
less significant at low illumination angles.

3.4.4 The dark ground image and polarisation.

The results of the previous section and work by other authors
(Richards and Wolf (1959) and Watrasiewicz (1965)) suggest that at
high illumination angles the conventional scalar diffraction theory
begins to break down and the image becomes polarisation dependent.
The effects of high illumination angles are most easily observed using
dark ground illumination as this removes the high intensity images of
the central orders from the final image.

In order to investigate this effect a photomask was illuminated from
outside the aperture of the objective (0.22 NA) using a polarised
633nm helium-neon laser. The illumination angle was $30°$ from the
optical axis and the beam lay in the y=0 plane. This illumination
angle lies within the aperture of a 0.6 NA condenser.

The laser was rotated in order to rotate the plane of polarisation and the dark ground image of a gap in chrome was recorded at different angles. It was found that the relative intensities of the two peaks varied as a function of the polarisation angle. Let the intensity of the trailing edge (the edge nearest the source) be I_2 and let the intensity of the leading edge (the edge furthest from the source) be I_1. Figure 3.4.3 shows how the ratio of I_2 to I_1 varies with polarisation angle for different photomasks. The relative intensities clearly vary from mask to mask, but the cyclic variations are in phase for all three masks.
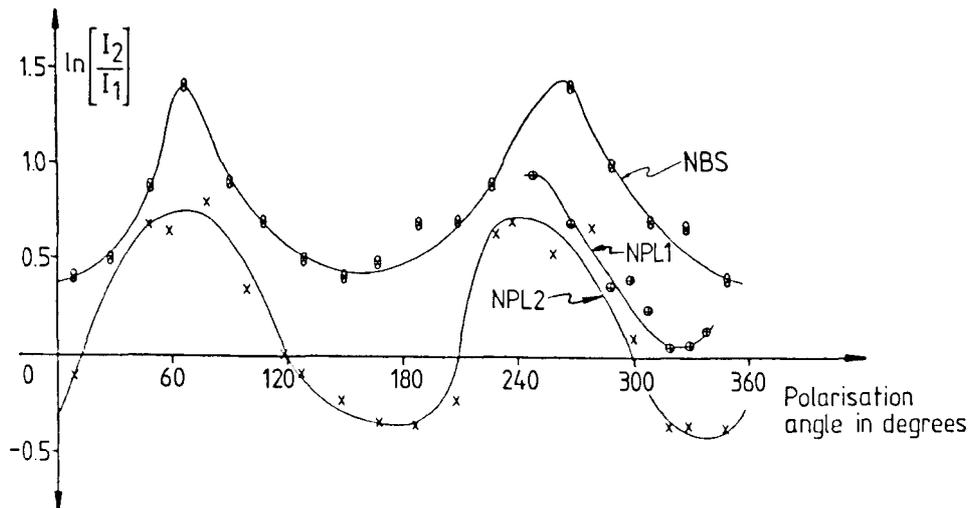
Figure 3.4.3. The effect of polarisation angle
on the dark ground image.

It was suspected that the effect may be due to asymmetry either on the mask edges or in the optics. This was tested by rotating the mask and comparing images and then moving the illumination source over to the other side and comparing images. Rotating the mask had no significant effect on the ratio of I_2 to I_1 which indicates that any asymmetry on the mask was negligible. Moving the source across to the other side completely reversed the image and this suggests that there is negligible asymmetry in the optical system.

In section 3.4.1 it was pointed out that the difference between the 25% and 50% linewidth measurements were often observed to be larger than predicted by theory. The three masks used in figure 3.4.3 were also used to compare 25% and 50% measurements. It was found that the edge difference in the bright field image was related to the ratio of the edge intensities in the dark field (table 3.4.2). Masks which

have sharp edge profiles in the bright field (ie. low 25%-50% measurement differences) also have large imbalances in the dark field image.

Mask	25%-50% in microns	Max(ln(I2/I1))
Ideal	0.14	0.00
NBS	0.21	1.40
NPL1	0.23	0.87
NPL2	0.25	0.72

Table 3.4.2   Comparison of the bright field 25%-50% measurement difference with the peak relative intensities of the dark ground image.

The results of the work presented in this section suggest that the imaging of chrome photomasks does not agree entirely with scalar diffraction theory. The deviations from the theory are particularly noticeable at high illumination angles.

Other workers in this field (Nyyssonen (1977)) have obtained very good agreement between theoretical and practical image profiles. This suggests that aberrations in the optics may be the cause of the discrepancies. On the other hand, variations between photomasks have been observed and this suggests that a photomask cannot be assumed to be a planar object with no polarisation dependence.

In order to obtain precise control over the critical dimensions of photomasks, it is necessary to monitor the edge quality as this will affect the linewidth measurement. The dark ground illumination technique described above is a very sensitive indicator of edge quality. Alternatively the edges may be monitored by using the difference between the 25% and 50% linewidth measurements.

CHAPTER FOUR

Image Processing for Measurement Automation.

In order to automate the measurement process it is necessary to capture the image profile. The amount of data collected will depend on what it is needed for. Feature locating routines will require coarse sampling of the field of view. Linewidth measurement algorithms will require relatively small, finely sampled one dimensional sections across the line object.

Different types of processing may be required for different stages of the measurement process. Pattern recognition algorithms may be required to locate features using a stored map or alternatively to build up a co-ordinate map of the image using edge and corner detecting algorithms. Automation of focusing and the measurement process itself will require very specific algorithms which will depend on the measurement technique being used.

4.1 Sampling the image profile.

The resolution limit of the optical system determines the closest two objects may come before they become indistinguishable. When locating features there is little point in sampling the image much beyond this limit. In an optical system with partially coherent illumination, the highest spatial frequency (F_m) which passes through the objective is produced by the point source at the outer edge of the condenser aperture (see section 2.5.1).

$$F_m = \frac{NA_o + NA_c}{\lambda} \qquad (4.1.1)$$

Thus the amplitude profile will consist of spatial frequencies of up to F_m. The intensity profile is the square of the amplitude profile and so the image resulting from the point source at the outer edge of the condenser aperture will contain spatial frequencies of up to $2F_m$. Since the images of all the illumination angles are linearly superposed, it follows that the image contains no spatial frequency components above $2F_m$.

In order to avoid aliasing it is necessary to sample the image at a frequency greater than twice this and for a 0.9 NA objective, 0.6 NA condenser and $\lambda = 0.53\mu m$, this requires the image to be sampled at

least every 0.18μm. This will ensure that any resolvable feature can be detected.

The usable field of view of an optical microscope using a high magnification objective will be at least 150μm in diameter and so to fully sample an image it would be necessary to capture over 5×10^5 pixels. This is roughly equivalent to a single frame using 625 line television format.

Sampling in this manner will ensure that the presence of any resolvable feature is detected but it is inadequate for measuring dimensions to precisions beyond the resolution limit. In chapter 3 it was demonstrated that isolated objects could be measured with very high precision provided that there exists a priori knowledge of the object and the measurement system. The limit to the precision of measurement of the image is given by the ability to locate a given intensity threshold. The distance between samples in the image must be less than half the value of the precision required and this would typically result in sampling the image at least once every 0.05μm. Coarser sampling is possible but this would require interpolation and for narrow lines where the edge profile is a function of linewidth this becomes a problem.

Some measurement techniques do not require such fine sampling of the image profile. The image shearing technique requires that the two images are adjusted until the intensity profile across the intersection is flat. If there is an error in the edge setting then there will be a dip or a peak at the intersection. When detected by the imaging transducer this dip or peak cannot be narrower than the width of the transducer point spread function. Provided that the image is sampled at intervals smaller than this, then precise edge setting is possible. In the case of the profiles in figure 3.3.4 this would require the image to be sampled about once every 0.2μm.

4.2 Threshold detection system for linewidth measurement.

The image shearing technique only works for images which do not have fringes at the line edges. The partially coherent images of low contrast objects tend to have edge fringes and this means that thresholds need to be located in the edge profile in order to perform a measurement.

The coincidence setting shear technique (Smith (1978)) enables a section of the image to be sheared relative to the rest of the field of view and features in the image such as fringes are used as verniers for the edge to edge alignment. In order to extend this technique to allow for measurements to be made at any threshold, Kirk and Moore (1983) proposed the generation of verniers corresponding to chosen intensity thresholds. These verniers are produced by detecting the appropriate intensity in the video signal and adding a very short bright pulse to the video signal at that point. The effect is to produce white lines along contours of equal intensity in the television picture.

The main advantage behind this technique is that it does not use the video signal for direct measurement but for a null setting. The measurement is performed by recording the displacement of the optical images between the alignments of the white lines. The calibration is thus determined by the optical displacement mechanism and not the video signal. Similarly geometric distortion in the video field will not affect the measurement.
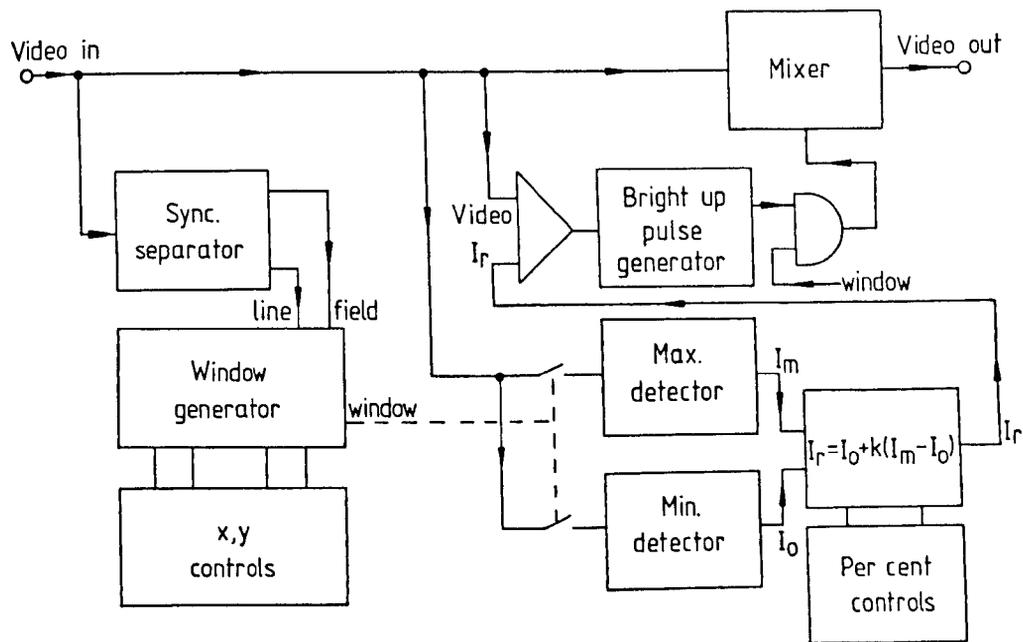
Figure 4.2.1. Block diagram of the edge threshold detector.

Figure 4.2.1 shows a block diagram of an instrument developed for this purpose and figure 4.2.2 shows how it operates on the video signal. A window is defined within the video frame and its size and position can be controlled by the operator. The borders to the window are defined by electronically generated bright lines and the video signal within

the window is reduced in amplitude in order to make the bright edge threshold lines stand out more clearly.
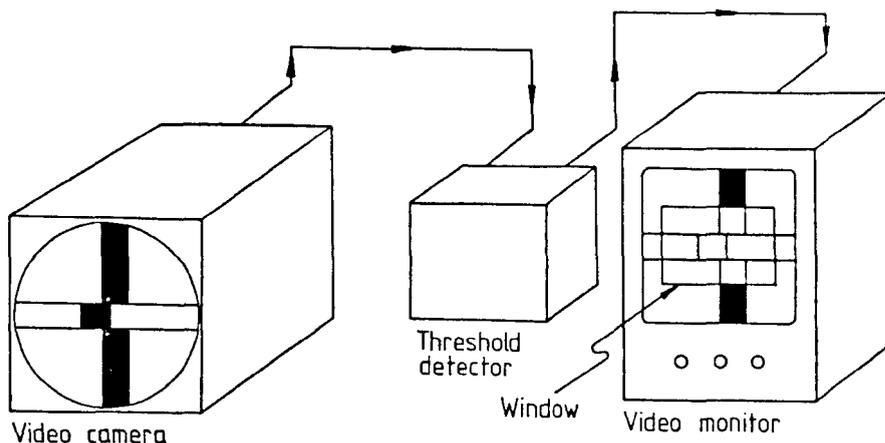
Figure 4.2.2. The video camera picks up the sheared image of a line object. The threshold detector locates the line edges of the object within a window and this image is displayed on the monitor.

Within the window area the maximum (I_m) and minimum (I_o) intensity levels are recorded and saved. The edge intensity threshold ($0<r<1$) is entered by the operator using the front panel controls and this defines the threshold in the video signal.

$$I_r = I_o + r.(I_m - I_o)$$ (4.2.1)

The location of the threshold is thus independent of the dark signal and the illumination intensity. A comparator compares the value I_r with the video signal and whenever this value is crossed a bright pulse is generated and added to the video display. These bright pulses form lines down the edges of the line object and can be used for alignment.

It was found that using this technique, measurement repeatability was improved (Blackburn (1983)) compared with just using the coincidence setting shear without the threshold detection. Although each operator's repeatability improved only slightly, the repeatability between operators improved significantly as the setting criterion was no longer subjective.

A method of automating the technique (Kirk (1983) and (1984a)) was proposed. By gating the appropriate rising and falling edges, it is
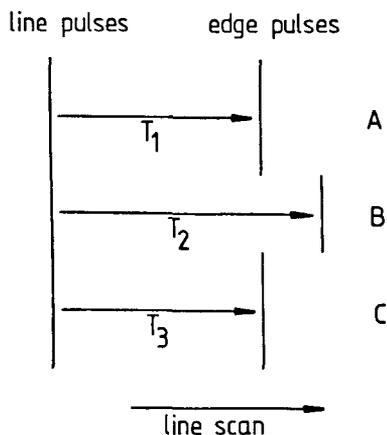
Figure 4.2.3. Detecting the alignment of two edge thresholds.

possible to isolate three edges for alignment (figure 4.2.3). The edges are aligned when the time delay between the line pulses and the bright up pulses in regions A,B and C are equal. The time intervals may be measured either with a high frequency clock and counter or using integrators. The effects of noise are minimised by timing over several (N) lines. The alignment error (e) is determined from the timed values for the groups of lines in the three regions.

$$e = \frac{k_c}{4N} \left[\sum_{n=1}^{N} T_{1,n} - \sum_{n=1}^{2N} T_{2,n} + \sum_{n=1}^{N} T_{3,n} \right] \qquad (4.2.2)$$

k_c = calibration constant.

The value of (e) can be used to compute the shearing error and then the shearing can be adjusted accordingly. The computed shear error may not be correct due to geometric distortion and so once the shearing is adjusted, (e) is recalculated and the process repeated until (e) has fallen to within the measurement tolerance. By gating different edges, this technique may be used to perform both edge to edge settings and hence determine the linewidth.

4.3 Spatial filter models for feature detection.

4.3.1 Introduction.

In the manufacture of integrated circuits it is necessary to inspect wafers at several stages of processing. The pattern printed on a wafer may vary in distance by as much as 50µm and the angular misalignment may be as much as 6°. This variation in positioning is greater than the field of view of a typical high magnification

objective and consequently the operator is required to search for a particular feature. This would normally be carried out using a low magnification objective with a correspondingly larger field of view. If the procedure is to be automated then the pattern recognition needs to be performed by a machine.

The co-ordinates of a feature are adequate for calculating the repositioning required in order to locate the target feature within the window of the inspection system. In order to assist with registration, global alignment marks are made on the silicon wafer and these are usually simple shaped artefacts such as T shapes. In this section a pattern recognition technique is described which is capable of locating the co-ordinates of such features. One method is to correlate a stored reference artefact with the image. However such a method has difficulty in detecting gross misalignment and it also requires that the target feature is of a reasonably similar size and shape as the stored reference. It would be preferrable to be able to locate the feature co-ordinates without needing a reference image.

4.3.2 The array filter.

The digitised image is stored in an N×M array which allows for filtering operations to be performed with n×m arrays (where n and m are much smaller than N and M). These filter arrays are convolved with the image array in order to modify the image in some way. Typical arrays include low pass filters which smooth images and reduce noise and high pass filters which enhance edges. The array filter may also be used to detect feature boundaries and corners and the design of these filters will be analysed here.
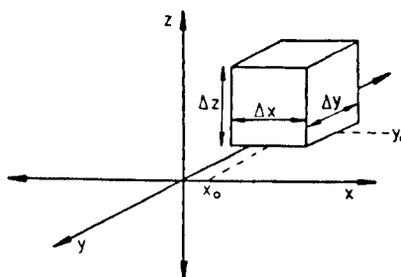
Figure 4.3.1. Single element in a filter array.

Each element in an array filter may be treated as a two dimensional RECT function in the image space (figure 4.3.1). Consider a pulse which is located at x_0 and y_0.

$$Z_o(x,y) = \begin{cases} \Delta Z_o & (x_o - \frac{\Delta x}{2} < x < x_o + \frac{\Delta x}{2}) \text{ AND } (y_o - \frac{\Delta y}{2} < y < y_o + \frac{\Delta y}{2}) \\ 0 & \text{elsewhere} \end{cases} \qquad (4.3.1)$$

Where ΔZ is the value of the filter element and may be complex. The pulse has width Δx and depth Δy. The filter may be represented in the frequency domain by its Fourier transform.

$$\tilde{z}(u,v) = \int\int_{-\infty}^{\infty} z(x,y) \cdot \exp(2\pi i(ux+vy)) \; dx \; dy \qquad (4.3.2)$$

Where $z(x,y)$ is the filter function in the image space and $\tilde{}$ and $\tilde{}[]$ denote the Fourier transform.

The function $z(x,y)$ is made up of $n \times m$ elements each defined by $Z_{j,k}(x,y)$ and by invoking the linearity property of the Fourier transform it is possible to consider each element separately.

$$\sum_{j=1}^{n}\sum_{k=1}^{m} \tilde{Z}_{j,k}(u,v) = \tilde{}[\sum_{j=1}^{n}\sum_{k=1}^{m} Z_{j,k}(x,y)] \qquad (4.3.3)$$

The Fourier transform of a single element is given by,

$$\tilde{Z}_o(u,v) = \int_{y_o-\frac{\Delta y}{2}}^{y_o+\frac{\Delta y}{2}} \int_{x_o-\frac{\Delta x}{2}}^{x_o+\frac{\Delta x}{2}} \Delta Z_o \cdot \exp(2\pi i(ux+vy)) \; dx \; dy \qquad (4.3.4)$$

$$\tilde{Z}_o(u,v) = \Delta Z_o \cdot \exp(2\pi i u x_o) \cdot \exp(2\pi i v y_o) \cdot \frac{\sin \pi u \Delta x}{\pi u} \cdot \frac{\sin \pi v \Delta y}{\pi v} \qquad (4.3.5)$$

Equation 4.3.5 describes the Fourier transform of any element in an array filter and it demonstrates several important results.

i) The two dimensions are completely independent of each other.

ii) The width of a transform at any angle through the origin is inversely proportional to its width in the image space and its height is directly proportional to the value of the filter element.

iii) The centre of the transform lies at $u=0, v=0$ regardless of the position of the pulse in the image space.

iv) The effect of moving the pulse off axis is to introduce a helical twist into the complex amplitude of the Fourier transform.

The function $\tilde{Z}_o(u,v)$ represents one pixel in the frequency domain. Thus using the linearity property of the Fourier transform described

by equation 4.3.3, the Fourier transform of the complete filter may be found by summing the Fourier transforms of all the filter elements. In general $\Delta x = \Delta y$ and for simplicity let $\Delta x = \Delta y = 1$ and thus x_0 and y_0 are integers. A general spatial filter in the image space is given by,

$$
G = \begin{bmatrix}
\Delta Z_{1,1} & \Delta Z_{1,2} & \cdot & \cdot & \Delta Z_{1,n} \\
\Delta Z_{2,1} & \Delta Z_{2,2} & \cdot & \cdot & \Delta Z_{2,n} \\
\cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot \\
\Delta Z_{m,1} & \Delta Z_{m,2} & \cdot & \cdot & \Delta Z_{m,n}
\end{bmatrix} \tag{4.3.6}
$$

and thus the general filter in the frequency domain is given by,

$$
\tilde{G}(u,v) = \sum_{j=1}^{n} \sum_{k=1}^{m} \int_{-\infty}^{\infty}\int \Delta Z_{j,k}(x,y) \cdot \exp(2\pi i(ux+vy)) \, dx \, dy \tag{4.3.7}
$$

Substituting equation 4.3.5 into 4.3.7 gives the general frequency domain expression for the filter.

$$
\tilde{G}(u,v) = \sum_{j=1}^{n} \sum_{k=1}^{m} \Delta Z_{j,k} \exp(2\pi u') \cdot \exp(2\pi v') \cdot \frac{\sin\pi u}{\pi u} \cdot \frac{\sin\pi v}{\pi v} \tag{4.3.8}
$$

Where,

$$
u' = u\left[j - \frac{n+1}{2}\right] \quad \text{and} \quad v' = v\left[k - \frac{m+1}{2}\right] \tag{4.3.9}
$$

Using equation 4.3.8 it is now possible to compute the Fourier transform of any array filter by building it up from its elements. This enables a filter to be designed from its required spatial frequency response.

This technique has been used to analyse the design of many commonly used filters (Kirk (1984b)) but the results presented here will concentrate on methods of detecting the co-ordinates of features.

4.4 The design of feature locating filters.

Most spatial filter applications which have appeared in the literature have used only real numbers in the arrays and have been designed without exploiting the phase properties of the Fourier transform. A common method of detecting edges has been to use a high pass filter with integer elements.

Equation 4.3.5 describes the Fourier transform of a single element anywhere in an array filter. When the element lies at $x_0=y_0=0$, the transform function is real. However when the element moves off axis the transform becomes complex and the direction vector of the amplitude describes a helix parallel to both axes. The frequency of this helix is proportional to the displacement of the filter element relative to the origin. It is important to realise that the modulus of the amplitude of the transform has not changed by moving the element off axis, only the phase has changed.

Conversely, the effect of shifting the Fourier transform is to introduce a helical complex amplitude across the filter array. Thus the position of the central peak of the Fourier transform is controlled by the phase profile across the filter. In order to construct a filter which is sensitive to edges only and not the steady state level, the gain at $u=v=0$ must be zero and so the phase shift must be chosen so as to cause the transform to pass through zero at $u=v=0$. Consider the general phased array filter below.

$$
\begin{bmatrix}
(\Delta Z_{1,1}, \theta+\delta\theta+\Delta\theta) & (\Delta Z_{1,2}, \theta+2\delta\theta+\Delta\theta) & . & . & (\Delta Z_{1,m}, \theta+m\delta\theta+\Delta\theta) \\
(\Delta Z_{2,1}, \theta+\delta\theta+2\Delta\theta) & (\Delta Z_{2,2}, \theta+2\delta\theta+2\Delta\theta) & . & . & (\Delta Z_{2,m}, \theta+m\delta\theta+2\Delta\theta) \\
. & . & . \, . & & . \\
. & . & . \, . & & . \\
(\Delta Z_{n,1}, \theta+\delta\theta+n\Delta\theta) & (\Delta Z_{n,2}, \theta+2\delta\theta+n\Delta\theta) & . & . & (\Delta Z_{n,m}, \theta+m\delta\theta+n\Delta\theta)
\end{bmatrix}
\quad (4.4.1)
$$

This filter allows for any amplitude at any point and for independent phase gradients across and down the filter.

The low pass filter has a narrow pass band and in its conventional form it attenuates both positive and negative frequencies above a certain value. Such a filter would be of the form,

$$
L = 0.04 \times
\begin{bmatrix}
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1
\end{bmatrix}
\quad (4.4.2)
$$

The Fourier transform of this filter is shown in figure 4.4.1. This filter forms a small "window" in the frequency domain and the reason why it is a low pass filter is because it is centred at $u=v=0$. If this "window" were moved away from the axis then it would pass a

different band of frequencies. It can be seen from figure 4.4.1 that the Fourier transform of this low pass filter has zeros in it and so by moving the response until a zero lies at u=v=0, the filter can be made into a band pass filter which does not respond to the steady state level.
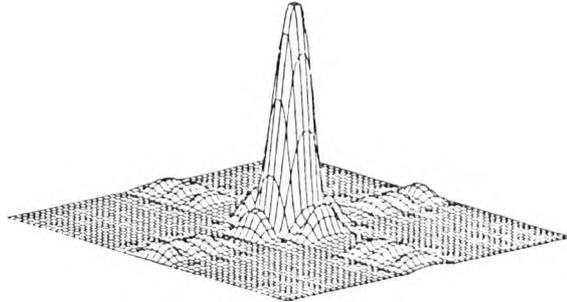
Figure 4.4.1. Fourier transform of the low pass filter L.

The low pass filter has $\Delta Z=1$ for all its elements but in order to shift its frequency response it is necessary to include a linear phase gradient. Thus each element in the filter array becomes,

$$\Delta Z_{j,k} = \exp(2\pi i k_1 j').\exp(2\pi i k_2 k').\Delta Z_{j,k} \qquad (4.4.3)$$

Where,

$$j' = j - \frac{m+1}{2} \qquad k' = k - \frac{n+1}{2} \qquad (4.4.4)$$

and k_1 and k_2 determine the frequency of the helix.
The Fourier transform of this filter is thus,

$$\tilde{z}(u,v) = \frac{\sin \pi n(k_1+u)}{\pi(k_1+u)} \cdot \frac{\sin \pi m(k_2+v)}{\pi(k_2+v)} \qquad (4.4.5)$$

The first zero in the frequency response for $k_1=k_2=0$, along the u and v axes will occur at $|u|=1/n$ and $|v|=1/m$ respectively. Thus in order to provide the frequency shift required for an edge detection filter it is necessary to choose k_1 and k_2 such that nk_1 and mk_2 are integers. For the 5×5 filter this gives $k_1=k_2=0.2$ for the first zero in the frequency domain, and in general it gives $k_1=1/n$ and $k_2=1/m$.

From equation 4.4.3 it can be seen that the terms nk_1 and mk_2 represent the number of phase revolutions from one side of the filter to the other. The condition $nk_1=mk_2=1$ results in one phase revolution across the filter. Higher integer values will also satisfy the condition for an edge detection filter but will move the ´window´

further from the axis. This may be undesirable as most of the information about an edge lies close to the origin in the frequency domain. If nk_1 and mk_2 are not integers then the filter gain at $u=v=0$ will not be zero and so the filter will be sensitive to the steady state level in the image. This will have the effect of producing an enhanced image rather than a high pass filtered one.

In conclusion there are three important points to remember when designing phase shifted arrays based on the low pass filter.

i)   The bandwidth is inversely proportional to the filter width in the image space along a given axis.

ii) The displacement from $u=v=0$ along a given axis, of the central peak of the frequency response of the filter, is directly proportional to the phase gradient (helix frequency) of the spatial filter in the image space along the corresponding axis.

iii) In order to implement an edge detection rather than an edge enhancement filter, it is necessary for the complex amplitudes to sum to zero along any axis of the filter. In the case of a filter where all the components have the same modulus, the complex amplitude vector must go through a whole number of revolutions along any axis.
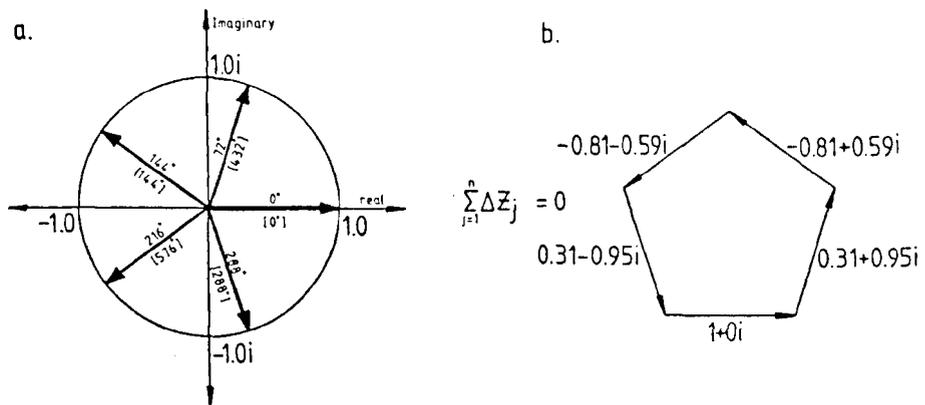
Figure 4.4.2. Design of a single turn phased edge detection
filter. a) Calculating the array elements.
b) All the elements sum to zero.

Figure 4.4.2a shows the design of a 5×5 spatial filter. The components are determined by segmenting a circle of unit radius into equal angles. Joining all these vectors shows that these components

sum to zero (figure 4.4.2b). If this phase shift is implemented in one direction only, then it produces the following filter.

$$
E = \begin{bmatrix}
.31-.95i & -.81-.59i & -.81+.59i & .31+.95i & 1+0i \\
.31-.95i & -.81-.59i & -.81+.59i & .31+.95i & 1+0i \\
.31-.95i & -.81-.59i & -.81+.59i & .31+.95i & 1+0i \\
.31-.95i & -.81-.59i & -.81+.59i & .31+.95i & 1+0i \\
.31-.95i & -.81-.59i & -.81+.59i & .31+.95i & 1+0i
\end{bmatrix}
\qquad (4.4.6)
$$

Figure 4.4.3 shows the Fourier transform of this filter and demonstrates how the phase gradient shifts the transform away from the centre. The advantage of this design over a conventional high pass filter is that the filter response is only high over the frequency range where it is required. It does not pass high frequencies away from the axes and so this will lead to an improved performance in the presence of noise.
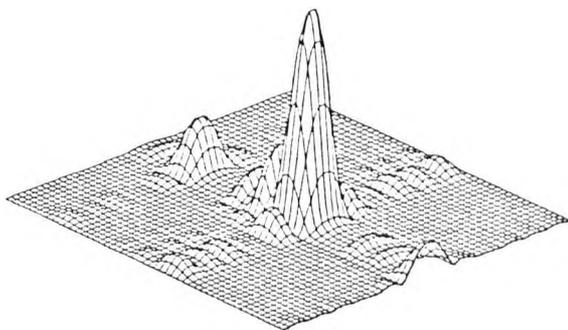
Figure 4.4.3. Fourier transform of the filter E.

The effects of these filters on image features can be demonstrated by considering a square object [R] after two applications of the low pass filter L (figure 4.4.4). The notation used here denotes a filter array by a capital letter and an image array by a capital letter in square brackets.

I = LL[R]
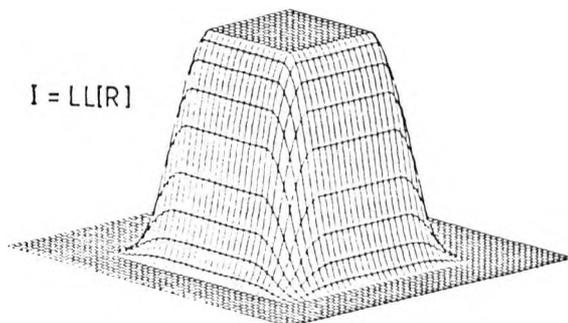
Figure 4.4.4. Low pass filtered image of a square.

[I]=ELL[R]
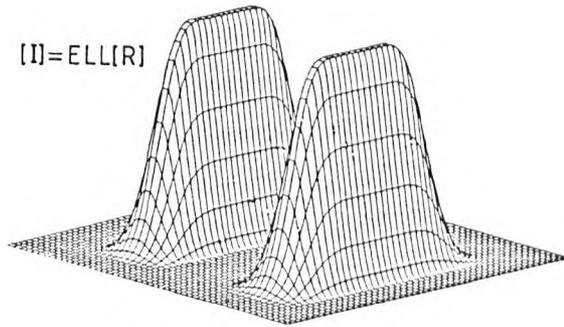
Figure 4.4.5.   Locating edges using the phased array filter E.

Figure 4.4.5 shows the effect of the filter E on the image in figure 4.4.4.   The filter response is of course complex and only the modulus is shown.   The effect of the filter is to locate edges which lie perpendicular to the phase gradient of the filter.   The phasing across the filter may be extended to two dimensions.

$$F = \begin{bmatrix} .31-.95i & -.81-.59i & -.81+.59i & .31+.95i & 1+0i \\ -.81-.59i & -.81+.59i & .31+.95i & 1+0i & .31-.95i \\ -.81+.59i & .31+.95i & 1+0i & .31-.95i & -.81-.59i \\ .31+.95i & 1+0i & .31-.95i & -.81-.59i & -.81+.59i \\ 1+0i & .31-.95i & -.81-.59i & -.81+.59i & .31+.95i \end{bmatrix} \quad (4.4.7)$$

The central peak of the Fourier transform of this filter lies away from both the u and v axes and thus it is only sensitive to features which are changing in two directions at once.   Figure 4.4.6 shows the effect of this filter on the smooth image of figure 4.4.4.   One single application of this filter successfully locates the corners of the image.

[I]=FLL[R]
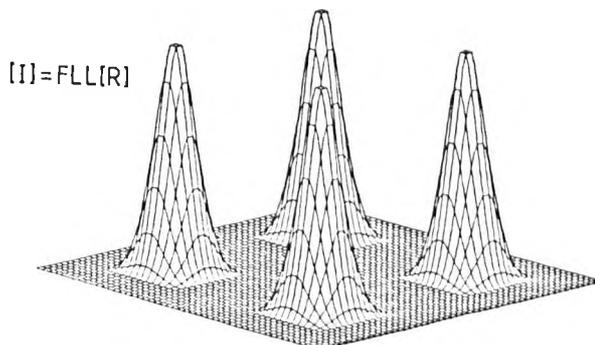
Figure 4.4.6.   Locating corners using a phased array filter.

As stated earlier, the displacement of the central peak of the transform away from u=v=0 is directly proportional to the phase shift from one side of the filter to the other (the phase gradient).   The

filters E and F have only one revolution of phase from side to side and by rearranging the components it is possible to make this into two revolutions. Three and four revolutions are not possible in a 5×5 filter as they would be the same as two and one revolutions respectively in the opposite direction. Similarly five revolutions are equivalent to no phase shift at all.

Although phased array filters allow for much greater design flexibility, they are not very efficient to implement in their basic form. The filters E and F would require to be implemented with floating point complex number multiplication routines and this would be slow on a microprocessor based system. Smaller arrays with no complex components and fixed multiplication would be desirable.

Smaller arrays can be implemented by calculating the appropriate phases for fewer elements.

Single turn
4×4 array.

$$\begin{bmatrix} 1 & i & -1 & -i \\ i & -1 & -i & 1 \\ -1 & -i & 1 & i \\ -i & 1 & i & -1 \end{bmatrix}$$

(4.4.8)

Double turn
4×4 array.

$$\begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix}$$

(4.4.9)

Single turn
3×3 array.

$$\begin{bmatrix} 1+0i & -.5+.87i & -.5-.87i \\ -.5+.87i & -.5-.87i & 1+0i \\ -.5-.87i & 1+0i & -.5+.87i \end{bmatrix}$$

(4.4.10)

A double turn 3×3 array would be the same as a single turn array in the opposite direction. The filters in equations 4.4.8 and 4.4.9 are of interest as they have whole number elements. The 4×4 double turn filter is particularly interesting as it contains only real, integer values. Although this simplifies computation, double turn filters are less desirable as they have only a small response between the origin and the first zero.

The filters described so far have their peak responses moved off the axes into one of the quadrants. The filter is only using the information from one quadrant despite the fact that the object will generate frequency components in all four quadrants. The filter response may be moved into any of the other quadrants by simply

changing the direction of the phase gradients. Thus the filter in equation 4.4.8 could also take the form,

$$\begin{bmatrix} 1 & -i & -1 & i \\ -i & -1 & i & 1 \\ -1 & i & 1 & -i \\ i & 1 & -i & -1 \end{bmatrix} \qquad (4.4.11)$$

In section 4.3 the linearity property of the Fourier transform was used to show that the sum of the frequency responses of a set of filters was equal to the frequency response of the sum of the set of filters. Now the two filters in equation 4.4.8 and 4.4.11 have their peaks in different quadrants and so the sum of these filters C will be a filter with a peak in two quadrants.

$$C = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \qquad (4.4.12)$$

The advantages of this filter are,

i)   No complex terms.

ii) Smaller than a 5×5 (16 instead of 25 elements).

iii) No multiplication required.

iv) Sparse data means that half the elements can be ignored.

This filter does not occupy all four quadrants of the frequency domain, in fact the peaks lie in only two. This is not making the most efficient use of the available image. Since the two filters have phase gradients in opposite directions along both axes, the peaks lie in diagonally opposed quadrants.

If an array is rotated through 90° then its Fourier transform will also rotate through the same angle. Thus by rotating the filter C through 90° the two peaks may be moved into the other two quadrants. The sum of C and its rotated image will thus be a filter occupying four quadrants.

$$D = C + C^{r} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \end{bmatrix} \qquad (4.4.13)$$

Figure 4.4.7 shows the Fourier transform of this four quadrant filter and figure 4.4.8 shows how it locates the corners in the smooth image of figure 4.4.4. Although filter D makes more efficient use of the frequency information than filter C, it also has twice as many non-zero elements and this will slow down the filtering process.
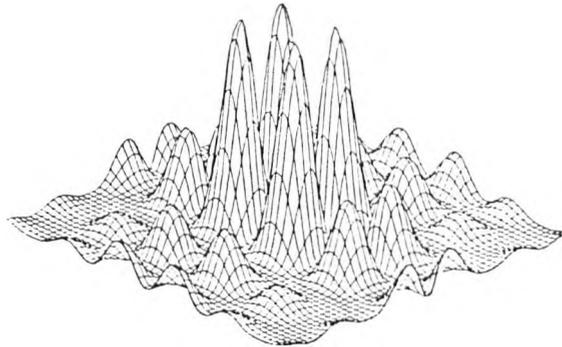
Figure 4.4.7.   Fourier transform of filter D.

$[I] = DLL[R]$
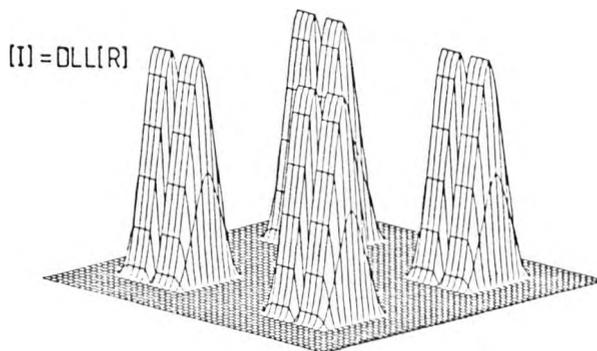
Figure 4.4.8.   Locating corners using filter D.

The performance of these filters in the presence of noise and angular misalignment of the image has been examined in detail (Kirk (1984b)). This work concluded that these corner detecting filters were able to locate corner co-ordinates in images with up to 7° misalignment.

4.5 Application of array filters to measurement automation.

In this section the design and application to measurement automation of one dimensional array filters will be considered. The two critical steps in performing a linewidth measurement are focusing the objective lens and locating the thresholds in the image profile. Array filters do not lend themselves to locating intensity thresholds directly from the image profile but they can be used to detect edge to edge settings in image shearing systems and this application is discussed in section 4.5.2.

4.5.1 Detecting critical focus using array filters.

In the work described so far, it has been assumed that the object is planar and lies in the focal plane of the objective lens. In order to obtain accurate linewidth measurements to within a certain tolerance, it is necessary to locate the focal plane to within an error bound.

Attempts to focus an optical microscope by eye are subjective. Individual operators can often locate the same plane of focus consistently, but different operators will focus to different planes. This variation between operators leads to variations in measurements and hence it is desirable to develop an objective focusing criterion. Ideally the method of detecting focus should lend itself to automation.

There are many patented focusing techniques and these have been reviewed (Kirk (1982)). Most image based focus detection algorithms attempt to implement similar focusing criteria to those used by operators when focusing by eye. Intuitively an image is thought to be in focus when it appears "sharp". This usually means that the edges in the image profile are at their steepest and that the curvature of fringes is at its greatest. Focusing criteria such as these are readily implemented using array filters and these are suitable for automating a focusing mechanism.

If an image $I(x)$ is sampled at intervals of Δx, then the image may be represented by the discrete profile $I(n)$. This image may be filtered using the $(2M+1)$ element array $f(m)$, to produce the filtered signal $I'(n)$.

$$I'(n) = \sum_{m=-M}^{M} f(m).I(m+n) \qquad (4.5.1)$$

If $f(m)$ takes the form of a difference filter $[-1,1]$ and Δx is small, then $I'(n)$ approximates to the sampled derivative of $I(x)$. Thus the peak in the array $I'(n)$ represents the steepest slope in the image profile.

The filter may be extended by adding zeros ($[-1,0,1]$, $[-1,0,0,1]$ etc.). The Fourier transform ($\tilde{G}(u)$) of this difference filter is given by,

$$\tilde{G}(u) = 2i.\sin(\pi\ u(M+1)).\frac{\sin\pi u}{\pi u} \qquad\qquad (4.5.2)$$

From equation 4.5.2 it can be seen that the effect of separating the -1 and 1 terms with zeros, is to reduce the bandwidth of the filter and also to increase the low frequency gain. Much of the edge information is carried in the low frequencies and so extending the filter length is desirable. There is of course a limit to this. If the filter is very long then then it will have a narrow bandwidth and so information about the higher frequencies will be lost. In section 4.1 it was demonstrated that there is an upper limit to the spatial frequency content of the image. This limit should be used to determine the bandwidth and hence the filter length.

f(m) may also take the form of a curvature detecting filter [-1,2,-1]. This will produce a peak response when the fringes in the image are sharpest. Again it may be padded with zeros in order to reduce the bandwidth.

In order to examine the behaviour of these two types of filter, four objects were considered.

a) Thin layer, 0.0 relative reflectance.
b) Thin layer, 1.0 relative reflectance, π relative phase.
c) Thin layer, 0.4 relative reflectance, $\pi/2$ relative phase.
d) Thick layer, 0.6µm thick silicon dioxide on silicon.

All the image profiles were generated for a 0.85 NA objective and 0.14 NA illumination aperture. A scanning aperture width of 0.2µm and illumination wavelength of 0.53µm were assumed. The field profile for the thick layer object was generated using a waveguide model which will be described in chapter 6. These four objects were chosen as they represent typical cases of photomask and wafer line objects. Once the far field diffraction pattern has been computed for each object, a through focus series of image profiles may be computed.

A through focus series of nine image profiles was generated for each object and sampled every 0.012µm. This profile was then filtered using the following arrays.

[-1,1] [-1,0,-1] [-1,0,2,0,-1] [-1,0,0,2,0,0,-1]

The largest amplitude response from each filter for each profile was

recorded and the results are shown in figure 4.5.1.
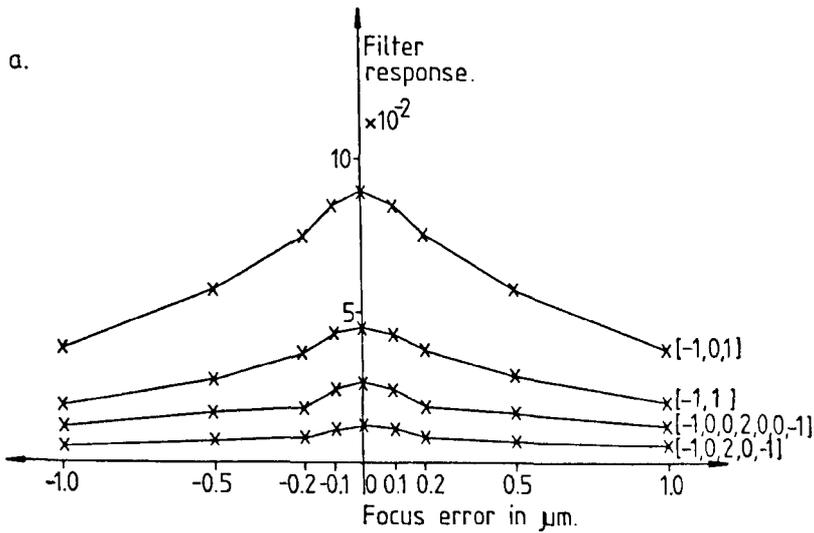
Figure 4.5.1a. Peak filter responses to a through focus series of images for an object with 0.0 relative reflectance.
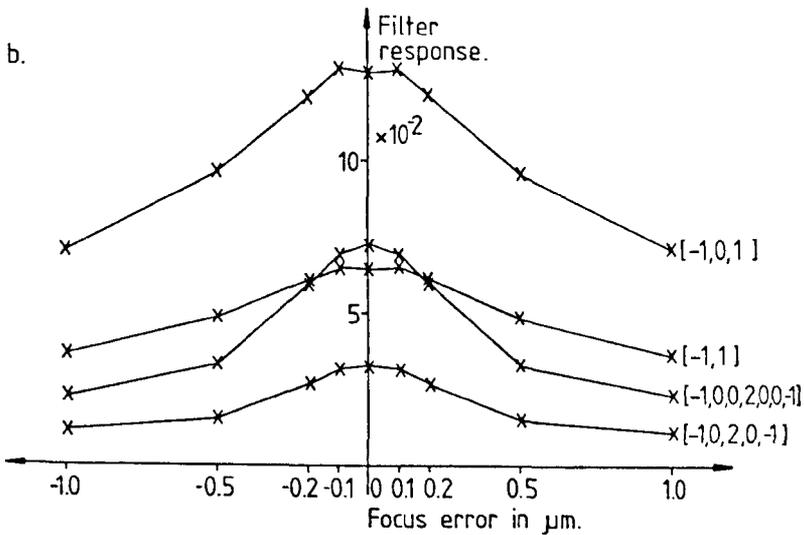
Figure 4.5.1b. Peak filter responses to a through focus series of images for an object with 1.0 relative reflectance and π phase shift.

As expected the results show that extending a filter improves its response. For both the slope and curvature detecting filters, adding zeros increases the signal response around focus.

Object (a) is the most ideal for locating the focal plane. All the filters produce the greatest response at the focal plane and also the response is symmetrical about the focal plane. Thus the steepest slope focus criterion is quite satisfactory for photomasks.
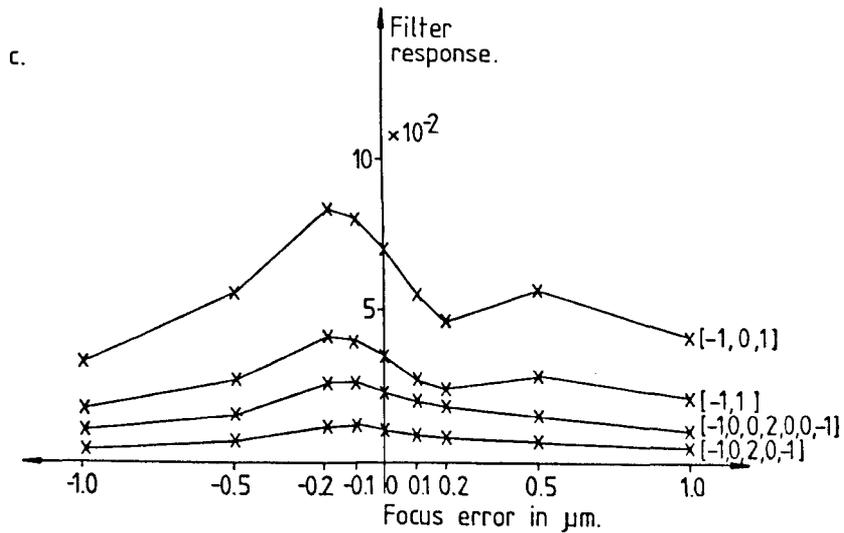
Figure 4.5.1c. Peak filter responses to a through focus series of
images for an object with 0.4 relative reflectance
and π/2 phase shift.
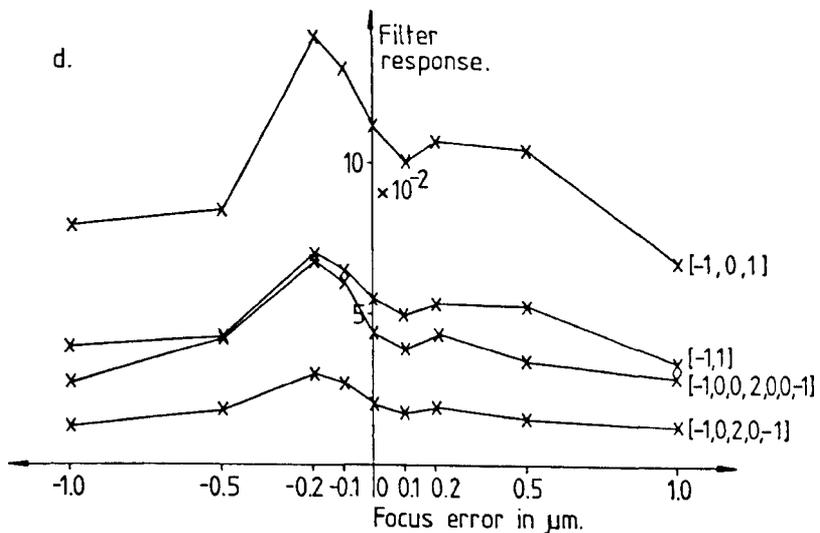
Figure 4.5.1d. Peak filter responses to a through focus series of
images for a 0.6 μm thick oxide layer on silicon.

Object (b) has no contrast and is only detectable because the light
reflected by the line is phase shifted relative to the light from the
substrate. The filter responses are not quite symmetrical about the
focal plane and it is the curvature filters which produce the sharpest
peaks.

Objects (c) and (d) demonstrate the problems of focusing on finite
contrast objects. Both the slope and curvature detecting filters fail
to locate the focal plane correctly. The steepest slope and sharpest

fringes do not necessarily occur when the object lies in the focal plane. Also the filter responses are not symmetrical about the focal plane and this would make interpolation difficult.

These focusing algorithms do at least make the focusing operation reproducible between operators but for many finite contrast objects, the "sharpest" image does not necessarily occur when the object is in the focal plane, which is contrary to what is often assumed.

4.5.2 Detecting the alignment of sheared images using array filters.

The technique of image shearing was discussed in section 3.3.6. It was demonstrated that when the opposing edges of an incoherently illuminated line object are aligned, the intensity profile becomes perfectly flat. Even when the illumination is partially coherent, the image profile becomes very smooth when the edges are aligned. Detecting this edge to edge alignment can be readily achieved using a digital filter. When the edges are not aligned, there will be a fringe and this will result in both slope and curvature components to the image profile. Thus the slope and curvature sensitive filters discussed earlier will produce a minimum response at the edge to edge shear setting.

A model of this technique was established in order to examine different filter responses. A 0.9 NA objective, 0.6 NA condenser and 0.53μm illumination wavelength were assumed. The camera point spread function width (σ) was taken as 0.16μm, with 14 video scan lines per μm. The image profile was sampled with an 8 bit analogue to digital converter. These parameters represent the optical microscope system which was set up to compare practical and theoretical results.

The image profile of a clear gap in opaque chrome was computed using the model in chapter 2 and the video signal of this image was computed using the models in chapter 3. The profile was sampled every 0.14μm. This represents 7 lines per μm and not 14 as stated above. The reason for the difference is that a 2:1 field interlace was used. The sampled signal was reduced to 8 bit resolution to simulate the precision of the analogue to digital converter.

The shearing was adjusted from the over shear condition through to the under shear condition in steps of 0.0082μm. At each position the response of different array filters was recorded. The range of the

sampled signal was restricted so that it only included the aligned edges and not the other edges of the sheared line image.

Figure 4.5.2 shows the responses of six array filters to the sheared image of a photomask line. The results show that there is good agreement between the experimental and theoretical filter responses. The curves (a),(b) and (c) show the response of curvature sensitive filters and demonstrate that lengthening the filter increases the sensitivity. Similarly the curves (d) and (e) show that lengthening a slope sensitive filter improves the response.

The filter in curve (f) is a combination of the curvature sensitive filter [-1,0,2,0,-1] and the slope sensitive filter [-1,0,1]. Although this filter is better than the slope sensitive filter on its own, it is no better than the curvature filter on its own.

In conclusion, the lengthened curvature sensitive filter [-1,0,0,2,0,0,-1] has the best response for detecting edge to edge alignment. It was found that increasing the length further made no significant improvement and if made too long, the sensitivity decreased. It is desirable to keep the filter as short as possible as the filter length is an important factor in determining the smallest linewidth which can be measured.
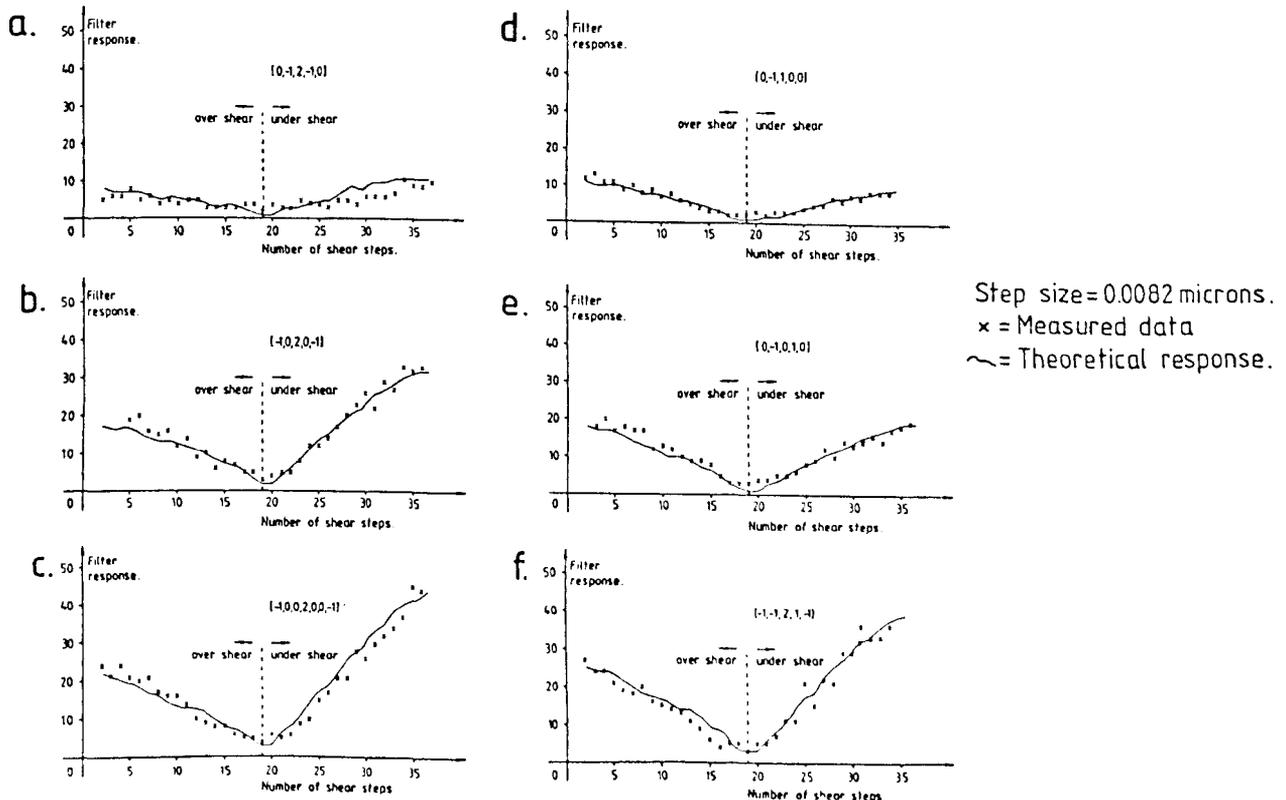
Figure 4.5.2. The through shear response of six array filters to the sheared image of a photomask line.

CHAPTER FIVE

Development of an Automated Measurement System.

In chapter 2, models were developed for the optical microscope and in chapter 3, different measurement techniques were discussed. Methods of sampling images in order to determine critical focus and edge to edge shearing alignment were analysed in chapter 4. These components are brought together in this chapter in order to design and analyse an automated critical dimension measurement system for measuring the length and width of gaps in magnetic recording heads.

5.1 Specification of the measurement problem.

During the manufacture of magnetic recording heads it is necessary to be able to make precise measurements of the length of the gap in the magnetic circuit. There are many types of magnetic head currently being manufactured, but this work is primarily concerned with measuring gaps of between 0.3μm and 3.0μm. These heads are used mostly for computer disc storage systems and video tape recorders.

Floppy disc heads have gap lengths of the order of 2μm and are often made in long blocks called loaves. These blocks are cut up into slices to form the individual heads. Most manufacturers choose to measure the length of the gap before the loaf is sliced. The allowed tolerance on floppy disc head gaps is usually about ±0.3μm, but future developments may bring this down to ±0.1μm. The audio recording heads for video recorders are made and measured in a similar manner, but generally have gap lengths of around 1μm.

Heads for winchester discs are made individually and require not only the gap length to be measured but also the gap width (figure 5.1.1). The gap lengths vary from about 0.5μm to 1.0μm while the gap widths vary from about 20μm to 50μm. The allowable tolerance on the gap length is typically about ±0.1μm while the tolerance on the width is considerably greater. The magnetic heads used for the video channel in video tape recorders have the smallest lengths of all. These heads are generally made in loaves with gap lengths of between 0.3μm and 0.6μm, and a manufacturing tolerance of the order of ±0.05μm[†].

Usually every head gap will need to be measured during manufacture and so a fast, reliable and low cost measurement system is required. The

---

[†] The details of this market background were provided by Mr. V. Burke of Vickers Instruments.
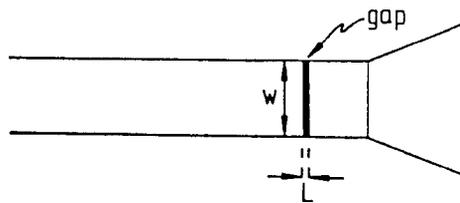
Figure 5.1.1. Top view of a winchester head,
with gap length L and width W.

high precisions demanded of these measurements and the high
throughputs involved, make it desirable to automate the measurement
process. Even at the end of the manufacturing process each head still
has a relatively low value and so measuring the gap can add
considerably to the production cost. Time consuming cleaning
processes need to be avoided and so the measurement system must be
able to cope with dirty heads. In order to successfully replace
existing manual measurement systems, an automated system must fulfill
the following specification.

i)   Automatically focus and measure in under 4 seconds.

ii) Measure gaps as short as 0.5µm to a precision of ±0.04µm (3σ).

iii) Measure gap widths (winchester heads) of 20µm to 50µm to a
     precision of ±0.1µm (3σ).

iv) Be able to measure in the presence of scratches, dirt, chips etc.

In this particular application it is not important for the
measurements to be accurate, but they do need to be precise.
Manufacturers are concerned with maintaining the same gap dimensions
within a batch and these are compared against a known acceptable head.

This approach is not ideal as it requires a reference to act as a
comparative standard. The reference must be very similar in size and
shape to the heads being measured or else it will not be possible to
use it as a calibration value. Since the measurements are not
absolute, they cannot be compared with those made on other systems and
it is not possible to reproduce the measurement calibration without
the reference head.

The measurement system needs to be robust as it will have to operate
in adverse conditions. Semiconductor linewidth measurements are
performed in clean rooms with stable temperature and humidity levels.
The measurement system and the operator will be closely monitored and

frequent calibration checks will be performed against national calibration standards. The manufacture of magnetic heads is quite different. Much of the production takes place in workshops where the environment is dirty and there is no control over temperature. The operators are generally unskilled and the importance of calibration checks is unlikely to be appreciated. The heads themselves are usually dirty and scratched and in these conditions the optics will also pick up dirt very quickly.

The measurement system must be reliable, require very little operator involvement and need only infrequent recalibration. Techniques which require careful adjustments of components need to be avoided.

In section 3.3.6, the image shearing technique was shown to be insensitive to many system parameters and this results in very good calibration stability and avoids a lot of expensive electronic hardware. Many recording head manufacturers are currently using systems based on image shearing but since it requires the operator to perform the focusing and edge to edge alignment, there is an element of subjectivity. Methods of automating the detection of critical focus were discussed in section 4.5.1 and array filter techniques for automating the image shearing were discussed in section 4.5.2.

5.2 System models.

The analysis presented in the previous chapters was used to model the image shearing measurement technique when applied to measuring gap lengths in magnetic recording heads. A head gap consists of a thin layer of glass sandwiched between two bulk areas of metal. Several magnetic heads were examined and no reflected light could be detected from the glass region. However if the refractive index of the glass is assumed to be about 1.5, then the theoretical reflectivity should be about 4% (normal incidence).

Initially it was assumed that there was no light reflected from the glass region and the optical microscope image profile was computed according to the method described in section 2.3.2. Later, the reflectivity of the glass was taken into account and this is discussed in section 5.4. The models of image shearing and a video camera were also included in the simulation and the final image profile was filtered using a variety of array filters[†]. The simulation of the

---

[†] The software for this simulation was written in Fortran 77 on the University of Leeds Computer Services' Amdahl 470 and is not subject to copyright.

measurement system was used to investigate the relationship between the optical measurements and the true gap length, together with how this is affected by parameters such as gap length, focus error, illumination wavelength and camera resolution.

A 0.85 NA objective was used with the illumination source filling the back aperture. The system magnification resulted in an image scale on the pick up tube of 0.03μm between the raster scan lines and a camera resolution width (σ) of 0.06μm. The mean illumination wavelength was taken as 0.53μm.

The responses of different edge setting filters were investigated. The image of a 2.0μm long gap was sheared over a range of 0.5 to 1.5μm and the video image profile was sampled and filtered over a range of ±1.0μm. The filter responses were recorded in two ways. For each shear position, the peak response of the array filter was recorded ($\hat{f}(p)$). Since $\hat{f}(p)$ represents the filter response of only one position in the image profile, it is very sensitive to noise. An improvement on this method is to record the sum of the absolute values of the filter response ($f_s(p)$) across the whole of the 2.0μm centre section of the profile. This reduces the effect of noise as more points are contributing to the filter response.

The model was used to compute $\hat{f}(p)$ and $f_s(p)$ for a variety of slope and curvature sensitive filters. The responses of two of these filters are shown in figure 5.2.1. As expected, the filter responses go through a minimum close to the position where the two edges overlap (at 1.0μm). This observed minimum is taken as the edge to edge alignment setting when performing the measurement. The response will not quite be zero as the images are partially coherent and for the same reason, even on long gaps, the response will not be symmetrical about the edge to edge setting. The summation of the curvature sensitive filter responses provides the sharpest definition of the edge to edge alignment and so this filter profile was used in the practical system.

When the gap length becomes comparable with the resolution limit of the system, the image profiles will distort and this will affect the measured value (see section 2.4.3). A range of gap lengths from 0.5μm to 3.0μm were examined (figure 5.2.2a) and it was found that the measured gap length was consistently 0.12μm longer than the true
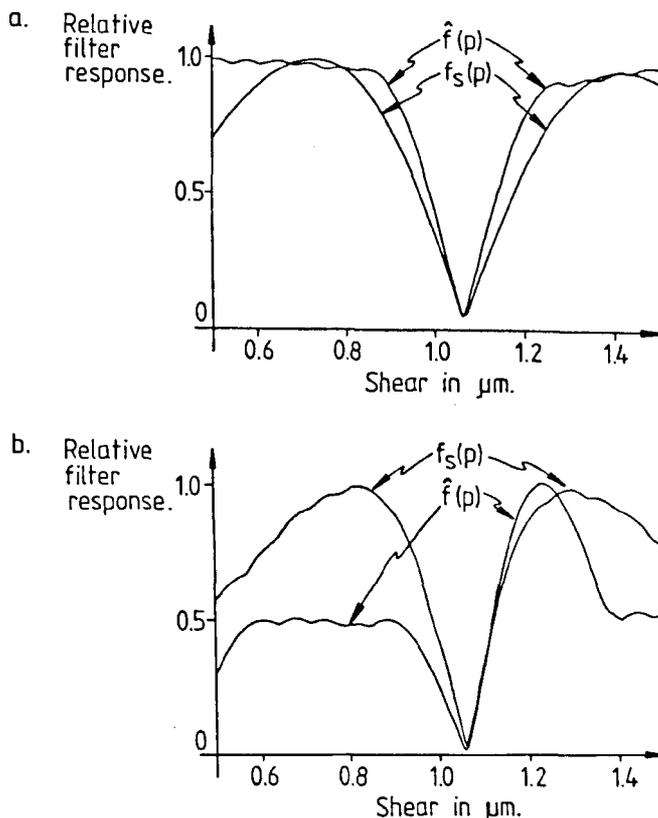
Figure 5.2.1. The response of, a) a slope sensitive filter [-1,0,1] and b) a curvature sensitive filter [-1,0,2,0,-1] to the video image profile as a function of shear position (p). The fs(p) values are scaled down by a factor of 10.

value. The optical gap length measurements are thus in error by a fixed offset which can be subtracted from the readings.

Focus detection algorithms using array filters were discussed in section 4.5.1 and it was shown that for a high contrast object, it was quite possible to locate the focal plane to within ±0.2μm using the steepest slope method. A theoretical investigation into the effects of defocus (figure 5.2.2b) showed that focus errors of up to ±0.5μm produced measurement errors of only ±0.01μm and so this method of focus detection is quite adequate for this application. Since the object field is real and symmetrical, the effects of defocus are the same in either direction. Figure 5.2.3 shows a through focus series of image profiles for a 2.0μm gap and figure 5.2.4 shows how the filter responds to these images when they are sheared.

The through focus image profiles pivot about a threshold close to the 50% intensity level and so focus errors do not introduce very large errors into the measurement. One effect of defocus is to reduce the
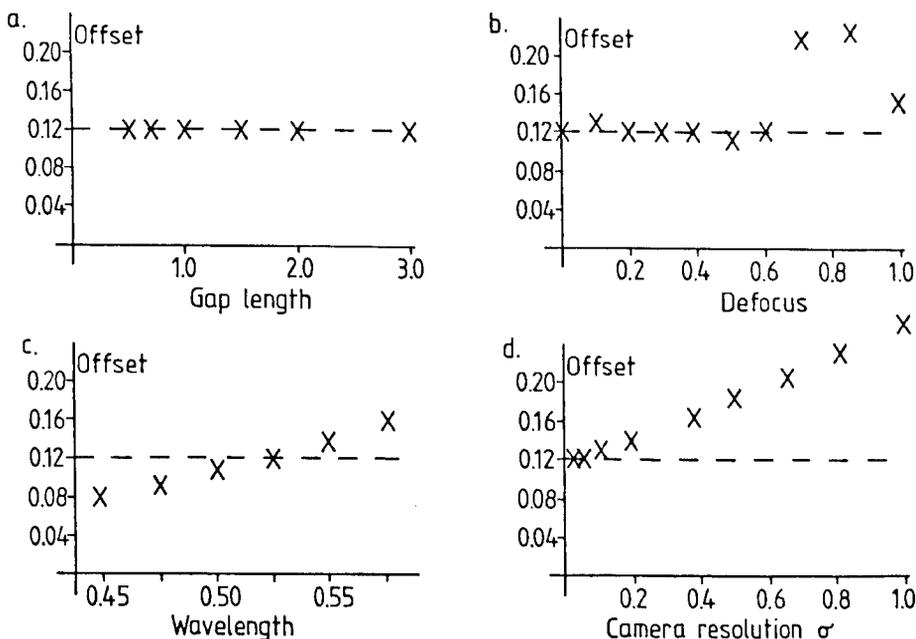
Figure 5.2.2. The theoretical offset between the measured and actual gap length as a function of gap length (a), defocus (b), wavelength (c), and camera resolution width (d).
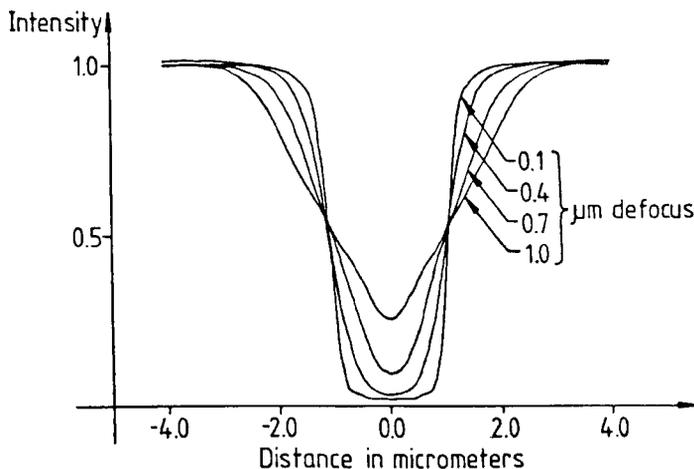
Figure 5.2.3. Through focus series of image profiles for a 2.0um long gap.

slope of the gap edge in the image profile. This reduces the sharpness of the fringe when the gap edges are aligned and so reduces the sensitivity with which the filter detects the edge to edge alignment condition.

The image profile is a function of illumination wavelength and this will affect the linewidth measurement (figure 5.2.2c). Short wavelengths allow more diffraction orders to be imaged and so produce images with sharper edges. Thus reducing the wavelength has the
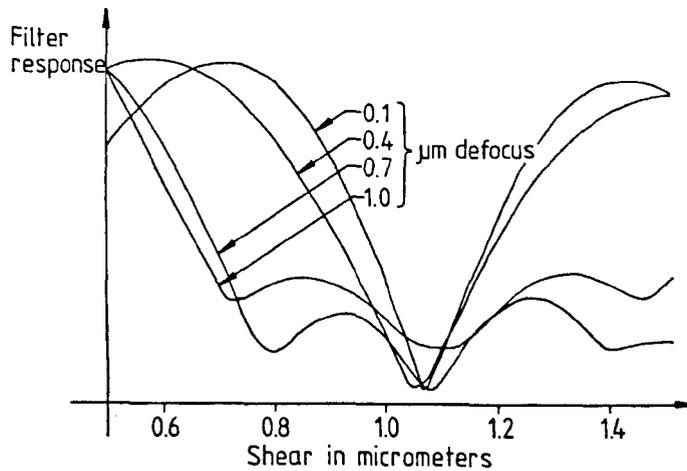
Figure 5.2.4. Filter response as a function of shear position for the through focus set of image profiles in figure 5.2.3.

effect of reducing the measurement offset.

The theoretical edge profile produced by a broadband green filter is very similar to the profile produced by monochromatic green light (λ = 0.53μm). However Nyyssonen (1977) suggests that the image edge profile is more sensitive to illumination bandwidth because of chromatic aberrations which have been ignored here. In practice it was found that the measurement technique used was insensitive to illumination bandwidth. Measurements performed using a green filter were found to agree with those using white light, to within ±0.01μm.

The effective camera resolution (σ) is determined by the magnification on the tube face. If the resolution is poor (large σ) then the camera will blur the fringe where the two sheared images meet and this will reduce the sharpness of the filter response. Figure 5.2.5 shows the filter responses for a sheared, focused image of a 2.0μm long gap for different camera resolution widths and figure 5.2.2d shows how this affects the gap length measurement offset. If the value of σ is kept below about 0.08μm then according to the model this introduces a measurement error of less than 0.01μm on gap lengths as short as 0.5μm. The effect the camera has on the measurement offset depends on the gap length. As the gap becomes shorter it is no longer small compared with the resolution width (σ) of the camera and so the outer edges of the sheared line image will distort the image profile and introduce a measurement error.

Filter response

σ in μm

0.02
0.20
0.50
1.00

Shear in micrometers
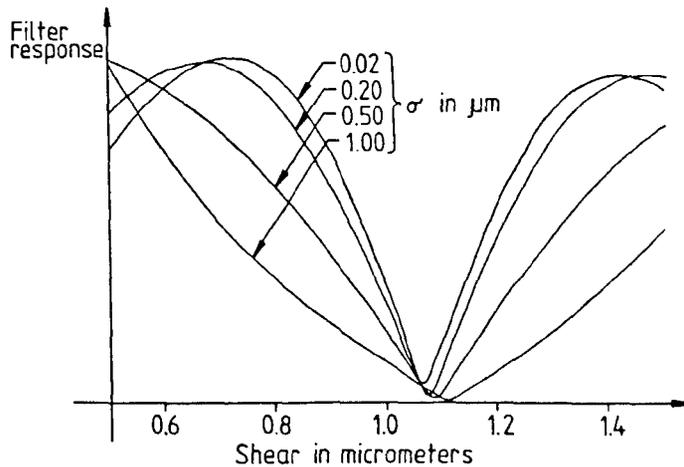
Figure 5.2.5. Theoretical filter response as a function of shear position using different camera resolution widths.

In conclusion, the model predicts that the focusing and measurement techniques are adequate for achieving precise measurements on gaps as narrow as 0.5μm. A broad band green filter may be used but the magnification must be high enough to achieve an effective camera resolution width (σ) of less than 0.08μm.

5.3 System architecture and design.

The theoretical results presented in the last section enable a technical specification for a gap length measurement system to be determined, based upon an image shearing optical microscope. The technique of filtering sheared images in order to measure gap lengths has been developed for the purpose of achieving a reliable and low cost system. The technique does not require a video camera with high linearity or low geometric distortion, nor does it require a computer with a fast processor or large amounts of memory. The instrument is in fact based around a conventional image shearing microscope, uses a Z80 microprocessor and requires less than 30 Kbytes of memory.

5.3.1 Hardware description.

The major components of the system are shown in the block diagram in figure 5.3.1. The operation of the instrument is controlled by a Z80 microprocessor based, S100 bus microcomputer. Input and output to the rest of the instrument is achieved using 8 bit parallel input/output ports and communications with the operator are via a terminal with an RS232 link to the microcomputer.
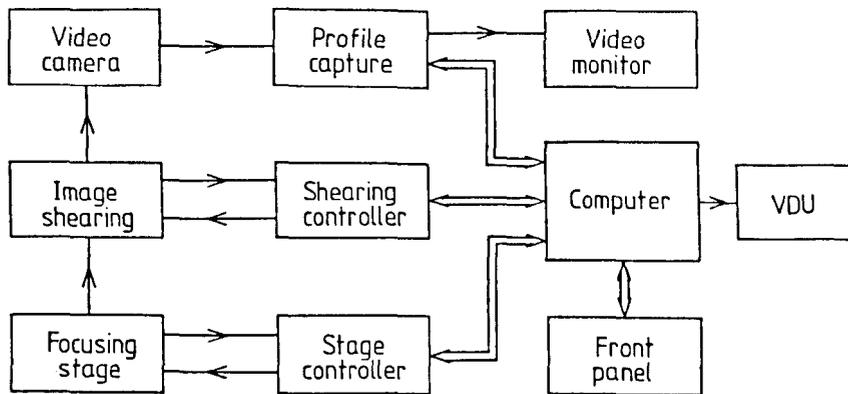
Figure 5.3.1. System block diagram.

The microcomputer controls the position of the object along the z axis by sending an 8 bit value to the stage controller via an output port. The stage controller then moves the stage in order to track this demanded value. The stage can thus be moved to 256 different positions over a range of about 30µm. A separate control loop is used for the stage in order to free the microprocessor from this task.

The object is imaged by a conventional optical microscope with incident illumination. The image is split using an image shearing module and the two displaced images are detected by an ultricon video camera. The image shearing module is controlled in a similar manner to the stage. Again a separate control loop is used to track the position demanded by the microcomputer in order to free the processor from this task. The shearing is adjusted by sending 12 bits via two 8 bit ports to the shearing controller, thus allowing 4096 different shear settings.

The video camera generates a conventional 625 line, 25 frames per second, 2:1 interlace video signal and this is sampled once every line and digitised to 8 bit precision. The 8 bit value is loaded into the microcomputer via an 8 bit parallel port.

a) Shearing controller.

The mechanical arrangement of the shearing module is shown in figure 5.3.2. A conventional manually controlled shearing module was adapted for the purpose by replacing the lead screw mechanism with a linear actuator. A current is applied to the linear actuator causing it to push the shaft outwards and so push on the lever. This in turn releases the force against the flexure spring and causes the mirror to move.
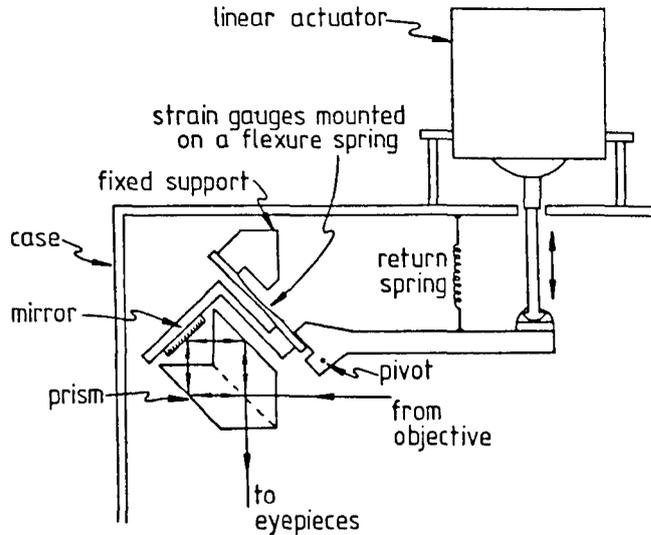
Figure 5.3.2. The mechanical arrangement of the shearing mechanism.

The return spring ensures that there is always a force against the flexure spring when there is no current to the actuator and thus the mirror is pushed to one side. In order to bring the mirror back to the zero shear position or to shear in the opposite direction, it is necessary to drive a current through the actuator.

Strain gauges mounted on a flexure spring provide a signal proportional to the mirror displacement and in the existing manual system, this signal was used for the measurement. The response of this displacement measurement transducer is extremely linear. The linearity was measured using a stage micrometer imaged with a 25X, 0.45 NA objective operating in dark ground. The images were sheared using the coincidence setting shear (CSS) mode and the strain gauge amplifier output voltage was recorded at 16 different positions. The shearing range was ±0.4mm measured at the linear actuator end of the lever. A best straight line was fitted to the data using the least squares curve fitting technique and the errors were recorded relative to this. All the values lay within ±0.2μm of their correct positions and this represents an overall non-linearity of less than 0.05%. This would result in a measurement error of less than 0.01μm on a feature as large as 10μm. Most of the points were in fact randomly distributed. This would be due to the difficulty of aligning the image edges accurately and also the positions of the markings on the stage micrometer would not be entirely accurate.

The "linear" actuator is far from linear and also suffers from severe hysteresis. Figure 5.3.3 shows the open loop response of the entire mechanism from the actuator to the output of the strain gauge amplifier.
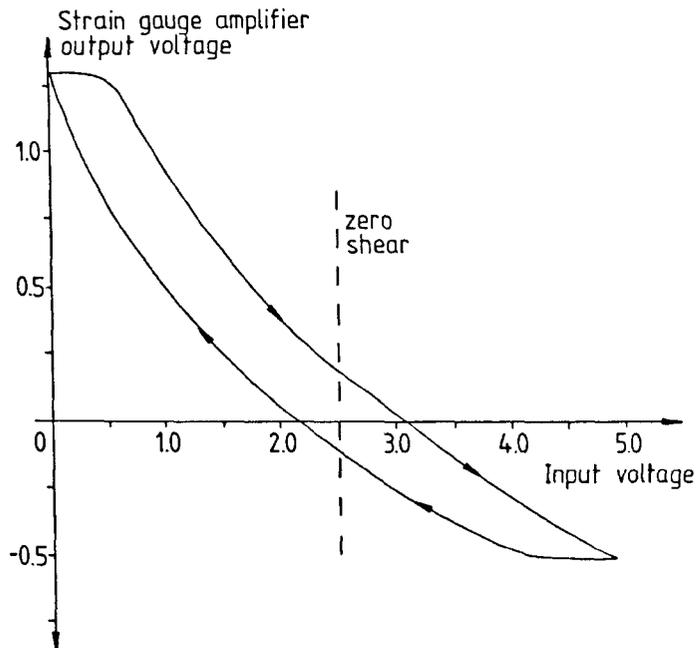
Figure 5.3.3. Open loop response of the shearing mechanism.

In order to obtain a linear system response, the shearing mechanism was enclosed within an integrating control loop (figure 5.3.4). The input voltage was provided by a 12 bit digital to analogue convertor and the overall closed loop non-linearity (including the digital to analogue convertor) was again found to be better than 0.05%.
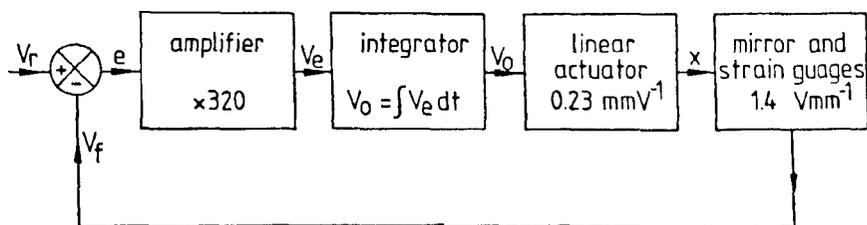
Figure 5.3.4. Image shearing module control loop.

Time domain plots of the system response, together with theoretical calculations showed that the forward transfer function may be described by,

$$x(s) = \frac{73.6}{s} \cdot e(s) \qquad (5.3.1)$$

$x(s)$ = actuator position in mm and $e(s)$ = error signal in volts.
The feedback transfer function is given by,

$$V_f(s) = 1.4 \times x(s) \qquad (5.3.2)$$

$V_f(s)$ = output voltage from strain gauge amplifier.

This gives the overall closed loop response of,

$$x(s) = \frac{73.6}{s+103} \cdot V_r(s) \qquad (5.3.3)$$

$V_r(s)$ = input signal.

The settling time can be defined as the time taken for the position error to fall to less than a specified value. In a situation such as this where the demanded value is a digital signal, it is convenient to define the system as having settled when the position error is less than one bit. If a 12 bit control signal is used, then from equation 5.3.3 it may be shown that the settling time (t) to within one bit for a step change of x, is given by,

$$t = \frac{1}{103} \cdot \ln(5120.x) \qquad (5.3.4)$$

t in seconds and x in mm.

The largest change in x is 0.8mm and this gives a worst case settling time of less than 80ms.

The shearing mirror controller was designed to operate to 12 bit precision and so is very sensitive to noise. For this reason, the digital to analogue convertor was completely optically isolated from the microcomputer and separate stable power supplies were used.

The shearing controller thus tracks the demanded position rapidly and accurately and so there is no need to measure the position of the shearing mirror. This has the advantage that it avoids the need for a 12 bit analogue to digital convertor which would be expensive.

The linear actuator was later replaced with a torque motor which has a much more linear open loop response. There was no significant change in overall closed loop performance.

b) Stage movement.

The stage movement was based on a design described by Kirk (1982). For this application, the movement was achieved by a torque motor instead of the linear actuator. The mechanism is shown in figure 5.3.5.
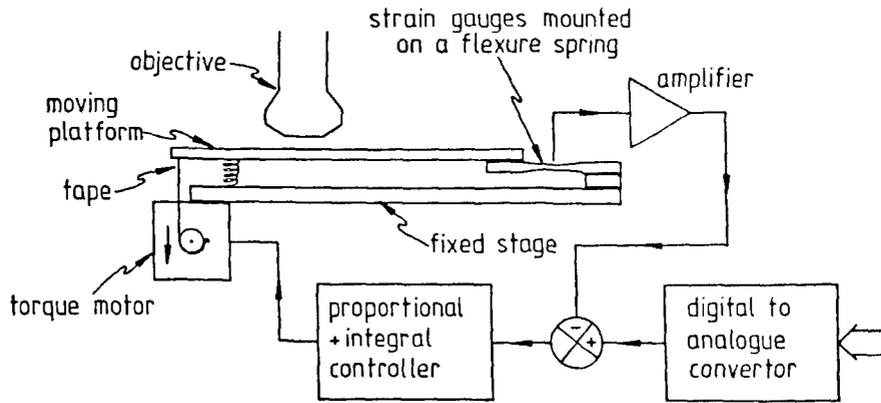
Figure 5.3.5. Schematic diagram of the focusing stage.

The stage movement is very similar in operation and design to the shearing mechanism. Worst case settling time to within one least significant bit is less than 250m*s*. The stage platform is much heavier than the shearing mirror and this accounts for its longer settling time.

The stage movement tilts the specimen but the angle is sufficiently small that it is not a problem. If the optical axis lies 50mm away from the flexure pivot and the video field of view is 50µm, then a ±15µm focus change (the full range of the movement is 30µm) will result in the extremes of the field of view being less than 0.008µm away from the focal plane. The autofocus algorithm can only focus to within about ±0.2µm and the measurement technique only requires the object to lie within ±0.5µm of the focal plane. Thus the slope of the object due to the stage pivoting can be neglected.

Sloping object problems can occur when the heads are not loaded correctly into the chip trays. For the purpose of gap length measurement, they need to be flat to within 3°. This would keep the object to within ±0.5µm of the focal plane over a 20µm range.

Although the stage movement has a very linear response, this is not necessary for this application. It is only important that the stage does not suffer from short term drift and that it can return to a demanded position precisely. Using a precise focus indicator (PFI) (Smith (1980)) and a 0.85 NA objective, it is possible to measure movement along the optical axis to within ±0.1µm. Using this device, there was no detectable error when the platform was moved by as much as 100µm and then returned to its original position.

<u>c) Profile capture.</u>

The video image profile is captured according to the approach shown in figure 5.3.6. This method has appeared in different forms in many different commercially available video sampling systems. At the beginning of each scan line a timer is started. The length of the time interval (T_1) is programmed into the timer by an 8 bit value from the microcomputer. At the end of the time interval T_1, a sample and hold is turned on for a short period ΔT. At the end of this period, the sample and hold goes into its hold mode and a bright line (cursor) is added to the video display to indicate the sampling position.
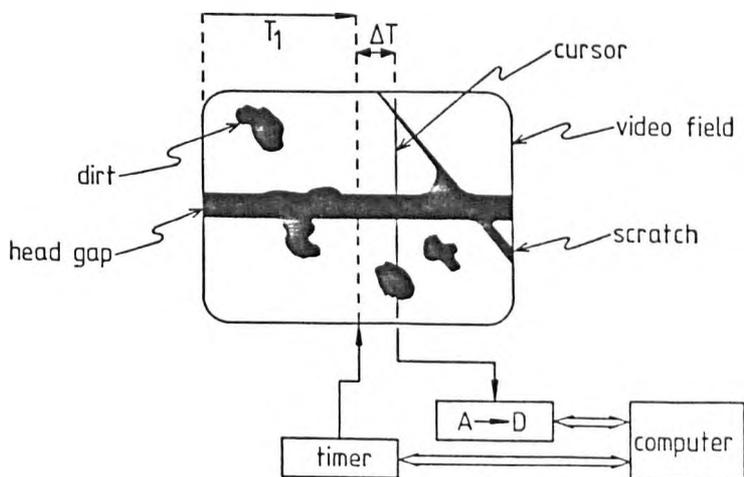
Figure 5.3.6.   Video field of a gap in a magnetic recording head with dirt   and   scratches.   The   cursor   position determines the location of the   sampling   window.

When the interval ΔT ends, an 8 bit analogue to digital conversion is initiated. Even very simple devices can perform this operation in under 20μS and so this leaves at least another 40μS before the next analogue to digital conversion starts (assuming a 64μS line scan period). The end of conversion signal from the analogue to digital convertor, together with the field pulse, are fed to one of the microcomputer input ports in order to provide synchronisation when loading the digitised profile into memory.

Using this approach it is possible to move the cursor to 256 different positions (or more if a higher precision timer is used) and at each position to sample $312\frac{1}{2}$ lines. Even when the flyback lines and the distorted edge regions have been allowed for, a 256×256 pixel image can still be loaded into the microcomputer in under 6 seconds. The

advantage of this technique is that it does not require rapid analogue to digital conversion and the sampling is slow enough to allow simple handshaking with a microprocessor.

During the sampling interval ΔT, the sample and hold tracks the video signal and at the end of ΔT when it enters the hold mode it should hold the value of the video signal at that instant. If the video signal is noisy, then the sampled signal will be noisy. Here it is possible to make use of the fact that the line object is one dimensional and so the video signal can be averaged over a small time interval before it is held. This is achieved by a simple low pass RC filter in front of the sample and hold.

If the illumination intensity is too low, then the video signal will have a poor signal to noise ratio. However, if the lamp brightness is turned up too high, the camera will saturate and this will introduce errors. The system is set up so that the analogue to digital convertor input range covers the voltage range from the black level to just below saturation. A pair of peak detecting comparators monitor the sampled video profile. One comparator detects if the peak level falls below $\frac{3}{4}$ of the saturation voltage, the other detects if the peak level exceeds a value just less than the maximum input of the analogue to digital convertor. If the peak value of the video profile goes outside this range this is indicated by lamps on the front panel. The operator uses these as a guide to adjust the illumination intensity before performing measurements.

5.3.2 Software description.

The software for the system[†] can be broken down into six distinct modules which are called up by the user. These modules can be combined to create useful routines which are shown in figure 5.3.7.

When the system is turned on, it automatically enters the calibration routine. The system can also be recalibrated at any time by selecting this routine. The front panel allows for different combinations of routines for focusing, site finding and measuring. Alternatively the system can be put into a manual mode which is useful for measuring features other than dark lines.

[†] Developed with the UNIX operating system on the Microsystems Unit PDP11/44 at Leeds University and not subject to copyright.
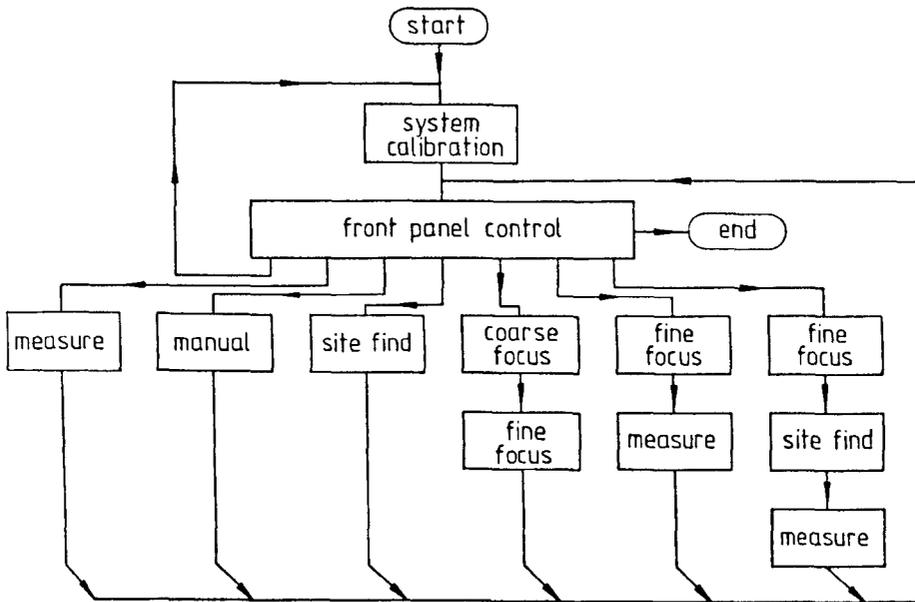
Figure 5.3.7. Software structure for gap length measurement.

a) System calibration.

This routine starts by setting the focusing and shearing to mid-position. The object is then coarsely focused manually and the shearing mirror is set to zero shear. The routine then calibrates the scale of the video camera (effective width of one scan line in μm) and the shearing movement (number of μm per step from the digital to analogue convertor).

It is necessary to calibrate the video camera in order to be able to make approximate gap length measurements. This rough mesurement is then used by the measurement routine to calculate how far to adjust the shearing.

The image shearing calibration is achieved by performing a manual shear and recording the number of shear steps. The operator then enters the gap length. If the objective or any other lenses are changed, it will be necessary to recalibrate using this routine.

b) Coarse focus.

This routine uses an extremely crude but quick method of focus detection. The stage is scanned through 16 positions across its full range and the video profile is captured at each position. The profiles are examined to find the maximum (I_m) and minimum (I_o) values in each. The profile with the largest difference $(I_m - I_o)$ is taken

as being the one closest to the focal plane. The stage is then returned to this position.

For objects with no phase change across the surface, the largest difference $(I_m - I_o)$ will occur when the object lies in the focal plane and so this algorithm should be capable of detecting critical focus. However, the profile amplitude changes very slowly around focus and so if the signal contains even a slight amount of noise, it will be difficult to locate the focal plane precisely. The advantage of this routine is that away from the focal plane, it has a response which continuously decreases as the object moves out of focus. This makes it ideal for correcting large focus errors. Despite its simplicity, this routine was found to focus repeatably to within $\pm 2\mu m$ of the focal plane.

c) Fine focus.

This routine steps the object through ± 12 steps about its present position. At each step, the video profile is captured and filtered using a [-1,1] digital filter. The profile which produces the greatest response to this filter is taken as being the one closest to the focal plane and the stage is returned to this position.

This algorithm detects the steepest slope in the image profile. For an object with no phase change across its surface, the steepest slope will occur at critical focus. The edge slope falls off rapidly as the object moves out of focus and so this algorithm is very sensitive to small focus errors. The object does not have to move far from focus before the filter response has fallen to a very low value and thus the algorithm is less useful for detecting large focus errors.

d) Site find.

When making automatic measurements it is important to find a clean site as dirt chips and scratches will affect the measurement. Figure 5.3.6 shows how these artefacts can appear in the image.

There are several assumptions behind the clean site finding routine. It assumes that all artefacts will be darker than the background metal material. Also it requires the head gap to be parallel to the horizontal line scan and assumes that no long scratches lie parallel to the gap. Since dirt can only reduce the intensity at any point in

the image, it follows that dirt which lies on the edge of a gap will make it appear wider and thus the narrowest point along the gap must also be the cleanest. This is not unreasonable for the purpose of measurement as parts of the gap which are wide because they are chipped, are normally regarded as inactive in the magnetic circuit and so it is not necessary to measure them.

Once the head gap has been found, the cleanest site is thus found on the basis that it is the narrowest point. However it is also necessary to find the gap in the first place, rather than trying to measure a scratch or a piece of dirt. The gap is located by virtue of the fact that it is known to be parallel to the horizontal line scan. The video image profile is sampled at 16 different cursor positions. The head gap will produce a dip in each profile at the same position. Artefacts such as dirt and chips will appear in some profiles but not others and scratches will appear in different positions in different profiles. Thus if all the profiles are added together, then the resultant profile will have a large amplitude dip at the position of the head gap, but only small dips elsewhere (figure 5.3.8). Thus the gap can be located approximately by scanning the array to find the position of this deepest dip (M). The maximum (I_m) and minimum (I_o) values are located in the array and the threshold I_e is taken as the mean of these two. The gap edges are defined as the two points S_1 and S_2 which cross the intensity threshold I_e closest to the point M. The gap centre is now taken as the mean of S_1 and S_2.

This locates the head gap and then each profile is examined in turn at the position where the gap is known to lie (p) and a rough gap length measurement is made from the video profile. The profile producing the shortest gap length is assumed to be the cleanest site. The rough gap length measurement is made by finding the highest (I_m) and lowest (I_o) values in each image profile. The line edges are then defined as the first two points on either side of the gap location which cross the intensity threshold $0.5(I_m-I_o)$ (figure 5.3.8).
The position of these gap edges in each profile is used for the rough gap length measurement and the profile with the shortest value is taken as lying across the cleanest site.
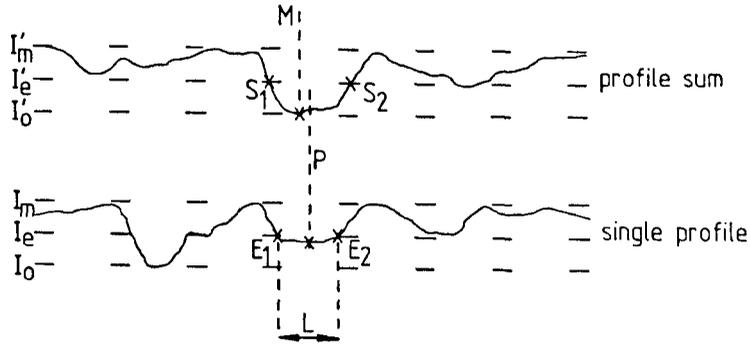
Figure 5.3.8. Locating the gap edges in a video profile.

e) Measurement.

The location of the cleanest site is passed from the site find routine to the measurement routine and then the measurement is performed at this position.

The video profile is captured and the gap length is measured by locating the points on either side of (p) which cross the threshold (I_e) (where I_e is the mean of I_0 and I_m). This gives a rough gap length measurement. This is performed several times and the mean gap length is calculated. This gives a reasonably precise value for the gap length.

Using the video scale value from the calibration routine, the image is sheared by an amount just less than this gap length. The shearing falls short of the rough edge to edge setting by 32 shear steps. The shearing is then scanned through 8 jumps, each of 8 shear steps in order to cover the range of ±32 shear steps about the edge setting. At each point, the image profile is captured and filtered and the filter response saved. The shearing is then returned to the position where the filter response was the least. The shearing is then scanned through ±20 single shear steps about this position and the profile captured and filtered at each step. The shearing is then returned to the position where the filter response was least.

This is taken as one of the edge settings (p_1). The sheared images are then switched over to the position $-p_1$ so that the opposing edges should now be in alignment. It is very unlikely that the system will have been accurately set to zero shear and so the edges will not necessarily line up exactly. The images are aligned by shearing them through ±30 shear steps about this point and the image profile is captured and filtered at each step. The shearing is returned to the position where the filter response was at a minimum and this is taken

as the second edge setting (p_2).

The difference p_1-p_2 should now be a measure of the gap length. Using the shearing calibration value derived by the calibration routine, the measurement is calculated and displayed.

When the image profiles are filtered, it is necessary to window them in order to ensure that it is only the signal corresponding to the aligned edges which is detected. If a gap is sheared over a range from $\frac{1}{2}$ its length to $1\frac{1}{2}$ times its length then the image may be filtered over a range of $\pm\frac{1}{2}$ the gap length about the centre, without introducing errors due to the outer edges (figure 5.3.9). The shearing window is thus defined by the line edges of the unsheared image.
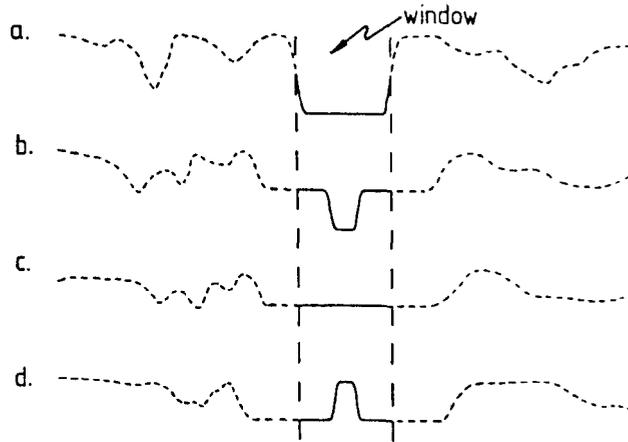
Figure 5.3.9.   A window is located in the unsheared image profile (a) between the two edges, and the profile filtering is restricted to this region. Curves b,c and d show the under, exact and over shear conditions respectively.

f) Manual.

Sometimes it is desirable to be able to perform manual measurements of images. When this routine is entered, the shearing is controlled directly by two front panel switches. When one switch is pressed, the output to the shearing controller is incremented and when the other is pressed, it is decremented.

A third switch allows measurements to be made. Whenever this switch is pressed, the current shear position is subtracted from the shear position when the switch was last pressed. This difference is multiplied by the calibration value and displayed. The current shear

position is then recorded as the reference for when the switch is next pressed.

5.4 Experimental performance verification.

Experimental tests were performed on the system using four floppy disc heads with nominal gap lengths of 1.4 to 1.7μm and four winchester disc heads with nominal gap lengths of 0.6 to 1.2μm. The winchester head gap widths ranged from 20 to 50μm. The aim of the experiments was to determine the absolute system performance in terms of accuracy and reproducibility. The validity of some of the assumptions made in the model could be determined and the theoretical performance predictions verified. The performance was also compared with the existing manual system.

Using the system in its manual mode, four operators each performed 25 measurements at the same site on a nominally 1.8μm long gap. The 3σ reproducibilities varied from 0.03μm for an experienced operator to 0.05μm for an inexperienced operator. Although the operators achieved good reproducibilities individually, when their measurements were compared the agreement was not as good. The four operators achieved 3σ reproducibilities of 0.03,0.04,0.04 and 0.05μm individually, but collectively, the 3σ reproducibility was 0.06μm. Thus each operator judges the focus and edge setting conditions differently. One advantage of an automated system is that it eliminates this subjectivity.

In order to assess the precison of the automated measurement technique, the system was operated without any autofocus. The microscope was focused manually using the precise focus indicator (PFI) and the focus setting was checked periodically during the experiment. 25 automatic and 25 manual gap length measurements were performed on each of the heads. The 3σ reproducibilities ranged from 0.025 to 0.035μm for the automatic measurements and from 0.035 to 0.045μm for the manual measurements. The manual measurements were all performed by the same experienced operator. The automated measurements demonstrate a significant improvement in reproducibility over the manual measurements.

The automated measurements were repeated at the same sites but allowing the system to focus automatically. The 3σ reproducibilities

now varied from 0.027 to 0.048μm (with a mean of 0.036μm). This is an improvement over the 0.06μm 3σ reproducibility achieved between operators performing manual measurements.

The results compare favourably with those published by LeMaster (1984). 3σ reproducibilities of 0.044 to 0.063μm were achieved using an image scanning technique. Prior to this a manual measurement system had achieved 3σ reproducibilities of 0.15 to 0.18μm.

Measurements were performed using both white light and a green filter, and they were found to agree to within 0.01μm. This suggests that the assumption that the illumination is monochromatic is reasonable. It does not appear to be necessary to restrict the illumination bandwidth with a green filter. However in extreme circumstances where the lamp is running at an unusually high or low voltage, the colour temperature of the filament will change significantly and this could result in a variation in the measurements. For this reason a green filter would be desirable.

The effect of magnification was investigated by making measurements with the magnification changer in both the 1.0X and 1.6X positions. There was no significant change in reproducibility.

Using the precise focus indicator and a Z axis measuring instrument, a series of automated measurements were made around the focal plane. It was found that the measurements varied by only ±0.01μm even with focus errors as large as 0.5μm in either direction. Outside this focal tolerance, the measurements changed rapidly as predicted in figure 5.2.2b. The model predicted that focus errors of up to ±0.5μm would introduce negligible errors into the measurement and so the focusing algorithm was designed to locate critical focus to within ±0.2μm. The practical measurements agree with these theoretical predictions.

Figure 5.4.1 shows the spread of measurement data from four winchester heads. It can be seen that the automated measurements lie closer together than the manual ones and that when autofocus is added, there is a slight increase in the spread. When the site find routine is used, the spread becomes large and the data can be seen to be clustered around 2 or 3 values. The reason for this spread is that the gap length measurements made when selecting a site are not very precise and if any noise is present different sites may be selected.

The gap length varies across the head and so the data is grouped into 2 or 3 clusters representing 2 or 3 clean and narrow measurement sites. This spread is thus a measure of the variation in gap length rather than the reproducibility of the system. This is reflected by the mean gap lengths in figure 5.4.1. The manual, automatic and autofocus measurements all result in similar mean values. However when the system is given the freedom to find its own site, it seeks out the shortest gap length and so the mean value can be quite different from the other measurements.

Figure 5.4.2 shows how the measurement reproducibilities compare on different heads. In general, the automated measurement is an improvement over the manual, but the autofocus increases the spread slightly. When the site find routine is included, the spread increases significantly on some heads as they have ragged edges causing multiple measurement sites to be selected.

The relationship between the optical gap length measurements and the true values was investigated by examining two of the heads in a scanning electron microscope (SEM). The SEM measurements range from 1.67 to 1.73µm along one head gap and from 1.60 to 1.67µm along the other. The corresponding optical measurements were 1.73 to 1.82µm and 1.60 to 1.73µm respectively. Most of the optical measurements were between 0.04 and 0.08µm larger than the SEM measurements.

According to the analysis in section 5.2, the optical measurements should be 0.12µm greater than the true gap lengths which is significantly greater than the offsets observed. One possible source of error could be that it was assumed that no light was reflected from the glass region. If the glass has a refractive index of around 1.5, then it will reflect about 4% of the normally incident illumination. If this is included in the model, then the theoretical optical edge threshold becomes 43% and the optical measurement should be 0.06µm larger than the true measurement. This compares well with the 0.04 to 0.08µm range of offsets obtained from comparing the SEM and optical measurements.

The technique can also be adopted to measure gap widths on winchester heads. These are much greater (20 to 50µm) than the gap lengths and require less precision (0.1µm, 3σ) when being measured. Since the width is large and the required precision is not particularly
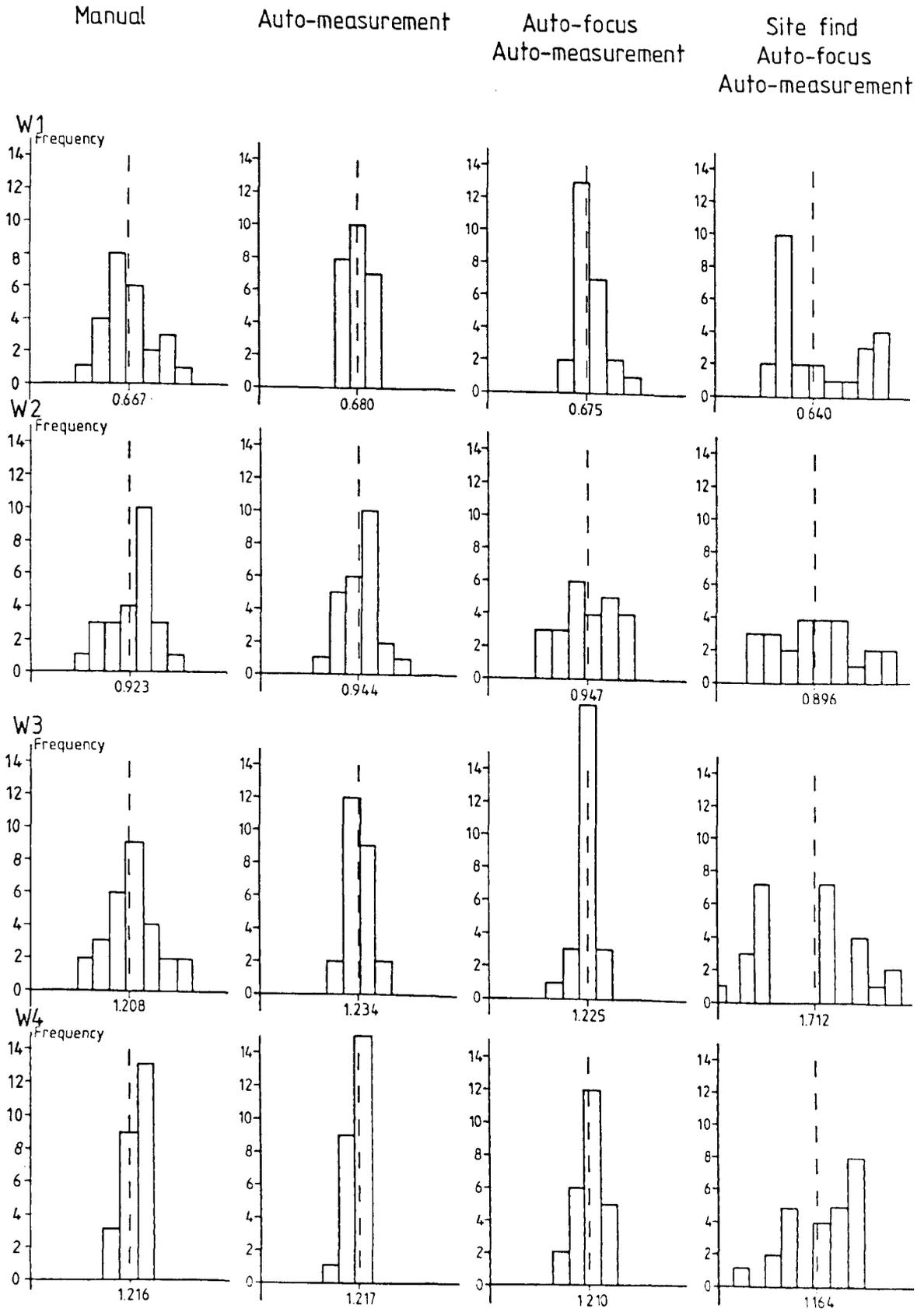
Figure 5.4.1.   Histograms   of the   measurements made on the gap lengths of four winchester heads using different measurement programs.
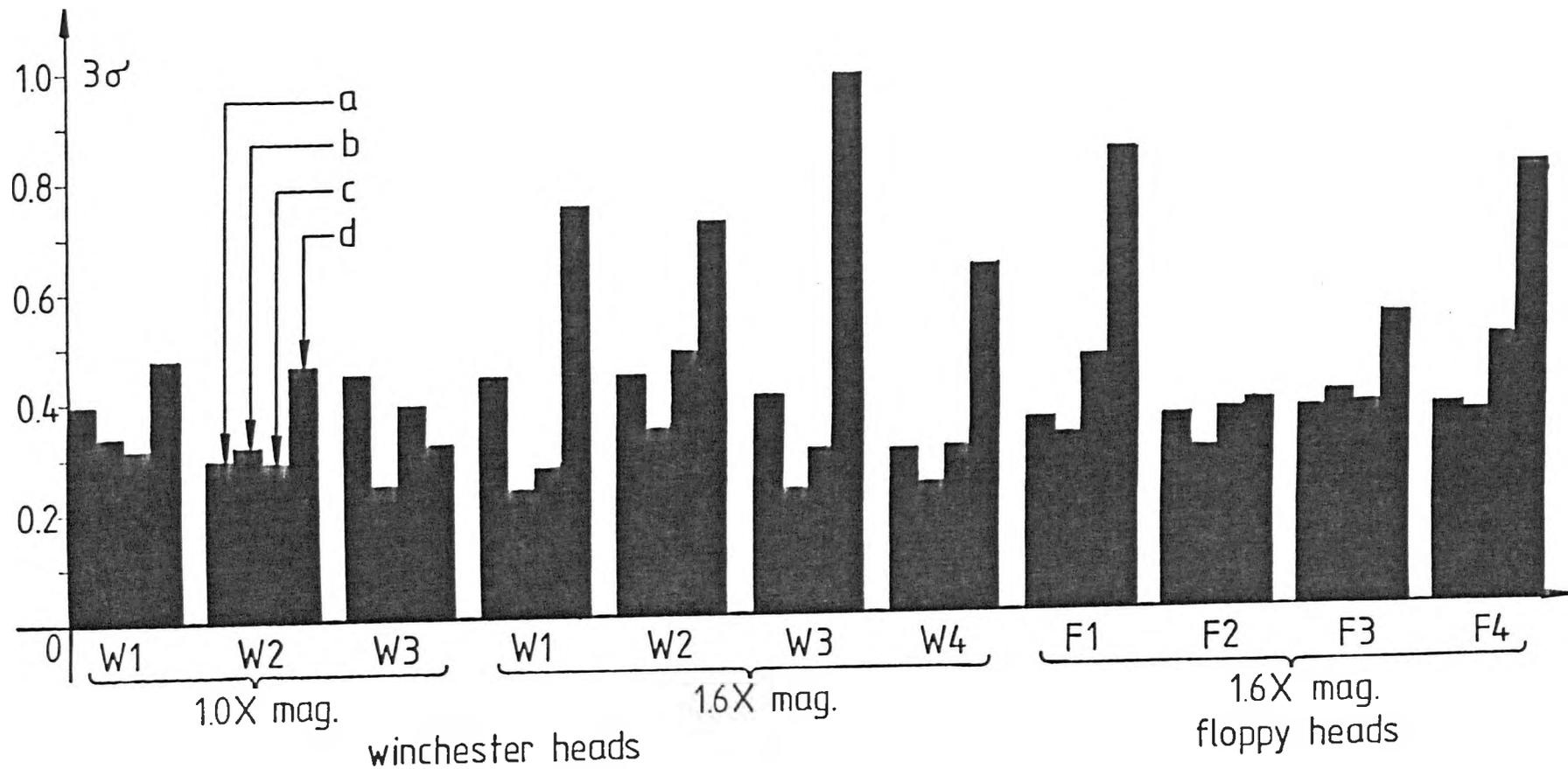
Figure 5.4.2. Measurement 3σ reproducibilities for different heads. a) manual, b) auto-measurement, c) auto-focus and auto-measurement and d) automatic site find, auto-focus and auto-measurement.

exacting, a low magnification can be used, resulting in much more light, less vibration and a greater focus tolerance (as the objective NA can be reduced). This makes the measurements considerably easier. On a batch of four winchester heads, the 3σ reproducibilities varied from 0.04 to 0.07µm.

In conclusion, the system models have been used to develop an automated critical dimension measurement system for measuring the length and width of the gaps in magnetic recording heads. A complete focusing and measurement cycle takes about 4 seconds with an overall 3σ reproducibility of under 0.05µm (gap lengths of 0.5 to 3.0µm). The offset between the optical measurement and the true gap length compares favourably with the 0.06µm predicted by the models. The system is also capable of locating clean measurement sites.

CHAPTER SIX

Modelling the Optical Images of Thick Layer Objects.

6.1 The thick layer imaging problem.

The work presented so far in this thesis has been based on the assumption that an object can be described by an amplitude and phase distribution which does not change significantly over a small range of illumination angles. It has been assumed that the object can be described by step change material discontinuities which result in step changes in the field profile. This approximation is valid for thin ($<0.2\mu m$) layers where the object is not thick enough to support propagation away from a feature such as a line edge.

It has been shown (Nyyssonen (1982c)) that the image profiles of thicker layers do not agree with the predictions of scalar imaging theory. When light is incident upon a thick object, it is diffracted and coupled into propagation modes which are supported by the waveguide structure. Most features of interest for semiconductor linewidth measurement fall into this category of thick layers. The polysilicon layers for transistor gates generally range from 0.3 to 0.7μm in thickness and photoresist layers range from about 0.3 to 1.0μm but can be up to 1.5μm thick.

For thin layers the edge geometry can be reasonably ignored as any transition takes place within a small fraction of a wavelength. However as the layer becomes thicker, the edge begins to cover a significant distance. A line which is only 0.6μm thick with edge walls as steep as 60° will have edges which are 0.3μm wide. This is comparable with the illumination wavelength.

A model is required which can accommodate thick layer objects, take into account different edge slopes and also cope with sublayers, asymmetric structures and variations within materials such as the granular structure of polysilicon.

Burckhardt (1966) presented a model for the diffraction of plane waves incident upon sinusoidally stratified dielectric gratings. This analysis was aimed at the problem of hologram reconstruction from plates of photographic emulsion of appreciable thickness. Unfortunately, the analysis was limited to objects which had

dielectric constant profiles which were constant in z but varied sinusoidally in x. Burckhardt established Maxwell's differential wave equation for three regions; the air layer, the patterned layer and the substrate. Within the patterned layer, it was shown that the H-mode satisfied an expression of the form of a Mathieu equation (Whittaker and Watson (1940)) and that the computation of integration constants from the boundary conditions could be reduced to an eigenvalue problem.

Kaspar (1973) extended Burckhardt's solution to include complex dielectric constants and non-sinusoidal stratifications. The analysis still assumed that the object profile was symmetrical in x and constant in z. Kaspar expanded the dielectric constant profile into a Fourier series and showed that when this was inserted into the differential wave equation, it generated a series of differential wave equations known as Hill's equation (Whittaker and Watson (1940)). The method of solving for the integration constants was essentially the same as that proposed by Burckhardt.

Nyyssonen (1982d) applied Kaspar's solution to the thick layer imaging problem of semiconductor line objects. The model was able to predict the image profiles of vertical edge wall dielectric patterns on infinite substrates. The solution was extended to allow for a sub-layer and later (Nyyssonen (1984)) it was extended to include materials with complex dielectric constants. Nyyssonen's approach was to use Kaspar's method to solve for the electric field profile across the top of the patterned layer and then to compute the image of this using conventional scalar diffraction theory.

In this chapter, Burckhardt's method will be extended to more general structures which are not constant in z (Kirk and Nyyssonen (1985)). The technique proposed involves splitting the line object up into separate layers which are constant in z but have independent widths, thicknesses and refractive index profiles. The waveguide modes are solved for each layer and then equated with adjacent layers. This enables the boundary conditions at the top and bottom of the patterned layer to be linked by a series of matrix operations. It is then possible to solve for the field at the top of the patterned layer and this can be used to compute the image profile.

6.2 Analysis of the field distribution in a thick layer object.

The analysis presented here is based on Burckhardt's method which has appeared in the literature (Burckhardt (1966)). Although the solution starts in the same manner, there are several differences which make the present work more general and so the analysis will be presented here in full.

6.2.1 Solution of the field within the patterned layer.

It is assumed that the specimen is illuminated by a coherent, monochromatic source in a direction perpendicular to the object. It will also be assumed that the illumination is polarised such that the E-vector is perpendicular to the plane of incidence and lies parallel to the lines of the object (figure 6.2.1). Thus $E_x = E_z = 0$.
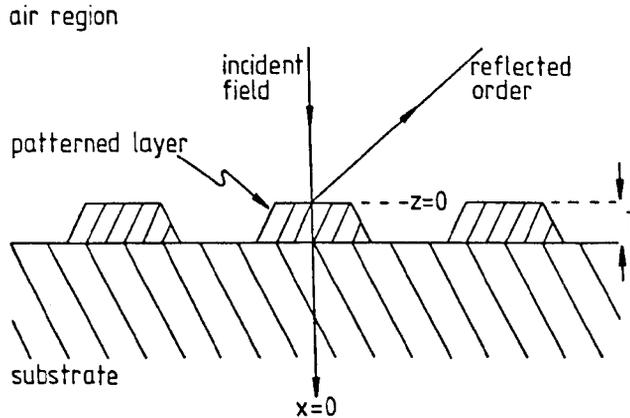
Figure 6.2.1. Orientation of the illumination relative to the object.

Burckhardt's solution starts from Maxwell's equation for a wave in an inhomogeneous space. In this case, it may be used to obtain the following equation for the field within the patterned layer.

$$\frac{\partial^2 E_y}{\partial x^2} + \frac{\partial^2 E_y}{\partial y^2} + \frac{\partial^2 E_y}{\partial z^2} + \text{grad}\left[\frac{E}{\varepsilon}.\text{grad}\varepsilon \right] = \mu_o \varepsilon \frac{\partial^2 E_y}{\partial t^2} \qquad (6.2.1)$$

Where the permittivity of the structure is defined by the function ε.

Now E_y lies parallel to the lines and hence is perpendicular to grad ε. Thus the scalar product term in equation (6.2.1) is zero for the H-mode. The object consists of lines parallel to the y-axis and so this partial derivative term reduces to zero.

$$\frac{\partial^2 E_y}{\partial x^2} + \frac{\partial^2 E_y}{\partial z^2} - \mu_o \varepsilon \frac{\partial^2 E_y}{\partial t^2} = 0 \qquad (6.2.2)$$

$E_y(x,z,t)$ is a sinusoidal function of time and may be written as,

$$E_y(x,z,t) = E_y(x,z).e^{i\omega t} \qquad (6.2.3)$$

Thus,

$$\frac{\partial^2 E_y}{\partial t^2} = -\omega^2.E_y \qquad (6.2.4)$$

Now,

$$\omega^2 = (2\pi f)^2 = [\frac{2\pi c}{\lambda_o}]^2 = k_o^2 c^2 = \frac{k_o^2}{\mu_o \epsilon_o} \qquad (6.2.5)$$

and

$$\epsilon = \epsilon_r.\epsilon_o \qquad (6.2.6)$$

Thus equation (6.2.2) may be rewritten as,

$$\frac{\partial^2 E_y}{\partial x^2} + \frac{\partial^2 E_y}{\partial z^2} + k_o^2 \epsilon_r E_y = 0 \qquad (6.2.7)$$

It is assumed that there exists a small interval in z such that ϵ_r may be taken as a function of x only and is independent of small changes in z. This means that if the dielectric region is patterned, then the pattern is not a function of z. This would represent structures with vertical edge walls in one material. For structures where the material changes as a function of z or the pattern within the dielectric varies with depth, then it will be necessary to assume that both the material and the pattern remain constant over a small interval of z.

Having made this assumption it is now possible to uncouple the partial differential equation in equation (6.2.7) into two equations. Let,

$$E_y(x,z) = X(x).Z(z) \qquad (6.2.8)$$

Since ϵ_r is not a function of z then,

$$\frac{\partial^2 E_y(z)}{\partial z^2} = \alpha^2 E_y(z) \qquad (6.2.9)$$

or

$$\frac{\partial^2 Z(z)}{\partial z^2} = \alpha^2 Z(z) \qquad (6.2.10)$$

α^2 is the separation constant and this will be determined later.

Substituting equation (6.2.8) into equation (6.2.7) gives,

$$X(x) \frac{\partial^2 Z(z)}{\partial z^2} + Z(z) \frac{\partial^2 X(x)}{\partial x^2} + \varepsilon_r k_o^2 X(x) Z(z) = 0 \qquad (6.2.11)$$

Substituting equation (6.2.10) into equation (6.2.11), and rearranging gives,

$$\frac{\partial^2 X(x)}{\partial x^2} + k_o^2 \varepsilon_r X(x) + \alpha^2 X(x) = 0 \qquad (6.2.12)$$

This equation holds for all functional forms of ε_r but the method according to Burckhardt relates to a sinusoidal function. Kaspar extended the method to cover functions which may be expanded into Fourier series.

Let ε_r be defined by,

$$\varepsilon_r(x) = \sum_{q=0}^{Q} E_q \cdot \cos(2\pi q b x) \qquad (6.2.13)$$

Where b is the fundamental spatial frequency of the pattern and the E_q terms are the coefficients of the Fourier expansion of $\varepsilon_r(x)$. Substituting equation (6.2.13) into equation (6.2.12) produces a transcendental differential equation known as Hill's equation.

$$\frac{\partial^2 X(x)}{\partial x^2} + k_o^2 \sum_{q=0}^{Q} E_q \cdot \cos(2\pi q b x) \cdot X(x) + \alpha^2 X(x) = 0 \qquad (6.2.14)$$

Floquet's theory (Whittaker and Watson (1949)) produces a solution to Hill's equation of the form[†],

$$X(x) = \sum_{j} B_j \ e^{2\pi i j b x} \qquad (6.2.15)$$

The differential equation reduces to Mathieu's equation when E_0 and E_1 are real and all the other E_q are zero. This equation was tackled by Burckhardt, and Kaspar's solutions relate to the more general case of Hill's equation.

Substituting equation (6.2.15) into equation (6.2.14) gives,

$$\sum_{j}(2\pi j b)^2 B_j \cdot e^{2\pi i j b x} - k_o^2 \sum_{q=0}^{Q} E_q \cdot \cos(2\pi q b x) \cdot \sum_{j} B_j \cdot e^{2\pi i j b x} = \alpha^2 \sum_{j} B_j \cdot e^{2\pi i j b x}$$

$$\text{[†]} \sum_{j} \text{ denotes the summation } \sum_{j=-\infty}^{\infty} \qquad (6.2.16)$$

Using the cosine identity this becomes,

$$\sum_j (2\pi jb)^2 B_j \cdot e^{2\pi ijbx} - \frac{k_o^2}{2} \sum_{q=0}^{Q} E_q [e^{2\pi iqbx} + e^{-2\pi iqbx}] \sum_j B_j \cdot e^{2\pi ijbx} = \alpha^2 \sum_j B_j \cdot e^{2\pi ijbx}$$

(6.2.17)

and rearranging gives,

$$\sum_j \left[(2\pi jbx)^2 B_j - \frac{k_o^2}{2} \sum_{q=0}^{Q} E_q (B_{j-q} + B_{j+q}) \right] \cdot e^{2\pi ijbx} = \alpha^2 \sum_j B_j \cdot e^{2\pi ijbx}$$

(6.2.18)

Now this equation must hold for each order and so it may be written as,

$$(2\pi jb)^2 B_j - \frac{k_o^2}{2} \sum_{q=0}^{Q} E_q (B_{j-q} + B_{j+q}) = \alpha^2 B_j \qquad (6.2.19)$$

This equation should hold for all j.

It is possible to truncate this infinite series without introducing a large error. The value of j represents the order of each mode and the number of modes which can be imaged is determined by the object period (b). In the case of normally incident coherent illumination, orders which represent diffraction of more than $90°$ cannot be imaged and so in this analysis they will be ignored. These high order modes will be evanescent and although they cannot be imaged directly, they will influence the field pattern within the object and so will affect the lower orders. From equation (2.3.25) the limit for j may be determined and this produces the condition,

$$J = \frac{n_0}{\lambda b} \qquad -J < j < J \qquad (6.2.20)$$

The set of equations in equation (6.2.19) may now be written in matrix form and their solution reduces to an eigenvalue problem.

$$D \cdot \overline{B}_m = \alpha^2 \overline{B}_m \qquad -J < j < J \qquad (6.2.21)$$

Where,

$$
\bar{B}_m = \begin{bmatrix} B_{-J,m} \\ \cdot \\ \cdot \\ B_{-2,m} \\ B_{-1,m} \\ B_{0,m} \\ B_{1,m} \\ B_{2,m} \\ \cdot \\ \cdot \\ B_{J,m} \end{bmatrix} \qquad (6.2.22)
$$

Where D is a matrix with the elements,

$$
d_{j,m} = \begin{cases} (2\pi j b)^2 - k_o^2 E_o & j = m \\ \dfrac{-k_o^2}{2} E_{|j-m|} & 0 < |j-m| \leqslant Q \\ 0 & |j-m| > Q \end{cases} \qquad (6.2.23)
$$

j denotes the rows of the matrix and m the columns. So far the matrix is symmetrical about the leading diagonal but later, problems generating an asymmetric matrix will be considered.

This produces a $(2J+1) \times (2J+1)$ matrix which will have $(2J+1)$ eigenvalues and $(2J+1)$ eigenvectors. The eigenvalues are the α^2 terms in equation (6.2.21) and the eigenvectors are the $\bar{B}_m$ vectors in equation (6.2.22). Each value of α^2 will have two corresponding values for α (α and $-\alpha$) and so this produces $2(2J+1)$ modes which have to satisfy the differential equation in equation (6.2.14).

At this point we diverge from Burckhardt and Kaspar's analysis to consider more general object structures. So far it has been assumed that the object may be described over a finite interval along the z axis, by equation (6.2.13). This represents an object which is symmetical about x=0, as equation (6.2.13) describes an even function. In order to model more complex asymmetric structures it is convenient to be able to solve the field equations for functions which are not necessarily symmetrical about x=0. For these cases the E_q terms will not have the Hermitian form assumed in equation (6.2.13).

In order to cover more general structures the series in equation (6.2.13) may be rewritten as,

$$\varepsilon_r(x) = \sum_{q=-Q}^{Q} E_q \cdot \exp(2\pi iqbx) \qquad (6.2.24)$$

Using this new definition produces a new form for the expression in equation (6.2.19).

$$(2\pi jb)^2 B_j - k_o^2 \sum_{q=-Q}^{Q} E_q \cdot B_{j+q} = \alpha^2 B_j \qquad (6.2.25)$$

This in turn results in a new definition for the elements of the matrix D.

$$d_{j,m} = \begin{cases} (2\pi jb)^2 - k_o^2 E_0 & j=m \\ -k_o^2 \, E_{j-m} & 0<|j-m|<Q \\ 0 & |j-m|>Q \end{cases} \qquad (6.2.26)$$

D is no longer a symmetric matrix and in general the E_q terms will be complex. The eigenvectors will in general be complex, but the eigenvalues will still be real unless the materials used in the structure have a complex refractive index. The eigenvalues relate to the propagation constants of the modes and for dielectrics with no loss, they will be real and negative regardless of any complex phase term introduced into the Fourier series by the geometry.

The α_m values and B_m vectors may be found for all the modes in equations (6.2.10) and (6.2.15). The general field solution can be found by substituting equations (6.2.10) and (6.2.15) back into equation (6.2.8) and this gives,

$$E_y(x,z) = \sum_{m=-J}^{J} \left(A_m e^{\alpha_m z} + A'_m e^{-\alpha_m z} \right) \sum_{j=-J}^{J} B_{j,m} e^{2\pi ijbx} \qquad (6.2.27)$$

The field is given by the sum of the modes and since each mode separately satisfies Maxwell's differential field equations, the total field will comprise of the sum of all the modes with each one multiplied by a particular constant. The values of these constants (A_m and A'_m) will be determined by the boundary conditions imposed by the geometry of the structure.

There are $(2J+1)$ modes and hence $2(2J+1)$ unknowns for A_m and A'_m. It is not necessary to solve for all these values throughout the structure but it is necessary to solve the set of A_m and A'_m at the top surface of the object. Once the solution has been found for the top surface then the optical image may be found from conventional diffraction theory.

6.2.2 Imposing the boundary conditions on the solution.

So far it has been assumed that the object structure can be taken as remaining constant over a finite interval in z, and nothing has been said about defining the physical structure. Figure 6.2.2a shows a typical physical structure for a line on a semiconductor wafer. This structure does not remain constant over intervals in z and so it is necessary to approximate the real structure with a layered model. Figure 6.2.2b shows how this structure can be divided up into layers which may be assumed to be constant over small z intervals. The structure need not be symmetrical and it can consist of any number of materials. The structure shown has been divided up into 11 layers. 5 layers of photoresist, 5 layers of polysilicon and one layer of oxide. In general, the more complex the structure, the more layers are required and the rules for setting up the layers will be discussed later.
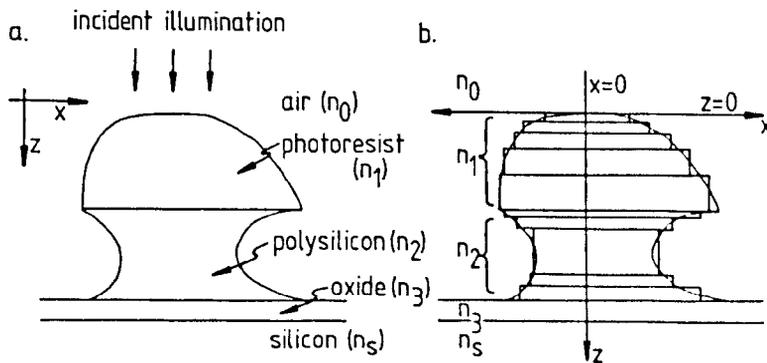
Figure 6.2.2. Cross section of a typical thick line object (a) and the corresponding multi-layer representation (b).

Each layer may now be treated as being constant over a finite interval in z. The extent of this interval will depend on the slope of the object at any given z position.

It is convenient but not essential, to define the first (top) layer as either having one edge at x=0 or lying symmetrically about x=0 (as in figure 6.2.2b). Each layer will have a refractive index (n_n), width (W_n) and an offset (Δ_n) relative to x=0. Assuming that the object is surrounded by one material with a refractive index of n_0, then the Fourier series expansion of the profile for the dielectric constant in any one layer is given by,

$$\epsilon_r(x) = \sum_q E_q \cdot e^{2\pi iqxb} \qquad (6.2.28)$$

Where,

$$E_q = \begin{cases} W_n bn_n{}^2 + (1-bW_n)n_0{}^2 & q=0 \\[2mm] \dfrac{(n_n{}^2-n_0{}^2)}{q\pi} \exp\left(2\pi i\Delta_n qb\right).\sin(q\pi W_n b) & q\neq 0 \end{cases} \qquad (6.2.29)$$

A series expansion has been used for each layer and so the object is assumed to be periodic with period 1/b. Figure 6.2.3a shows the dielectric constant profile of each layer. For the structure in figure 6.2.2b each layer will have this form but in general, any function may be represented. Figure 6.2.3b shows a dielectric constant profile for a line with another small line close to it on one side. The second line has a different refractive index from the first line. Structures such as this, enable the effects neighbouring features have on the image profile to be examined. This latter structure can be represented by the Fourier series,

$$E_q = \begin{cases} W_n bn_1{}^2 + d_n bn_2{}^2 + \left(1-b(W_n+d_n)\right)n_0{}^2 & q=0 \\[2mm] \dfrac{(n_1{}^2-n_0{}^2)}{q\pi}\exp\left(2\pi i\Delta_n qb\right).\sin(q\pi W_n b) + \dfrac{(n_2{}^2-n_0{}^2)}{q\pi}\exp\left(2\pi i\delta_n qb\right).\sin(q\pi d_n b) & q\neq 0 \end{cases}$$

$$(6.2.30)$$

This provides the E_q terms for equation (6.2.28).

Having described how the object physical structure can be modelled, we now return to the field equations in order to find a solution given the boundary conditions imposed by the air layer and the substrate. It is required to solve the A_m and A'_m terms for the top layer (equation (6.2.27)). The air space is denoted by the subscript 0, each layer by 1 to N and the substrate by s. The superscript + denotes an incident wave and − denotes a reflected wave.

Following Burckhardt's method and adding the incident and reflected waves gives the equation,

$$E_{y0} = E_{00}^{+}\exp(-ik_0 z)+ \sum_{j=-J}^{J} E_{j0}^{-}\exp(2\pi ijbx).\exp\left(ik_0(1-\lambda^2 j^2 b^2)^{1/2}z\right)$$

$$(6.2.31)$$

The equation for the H-field may be obtained from Maxwell's relation,
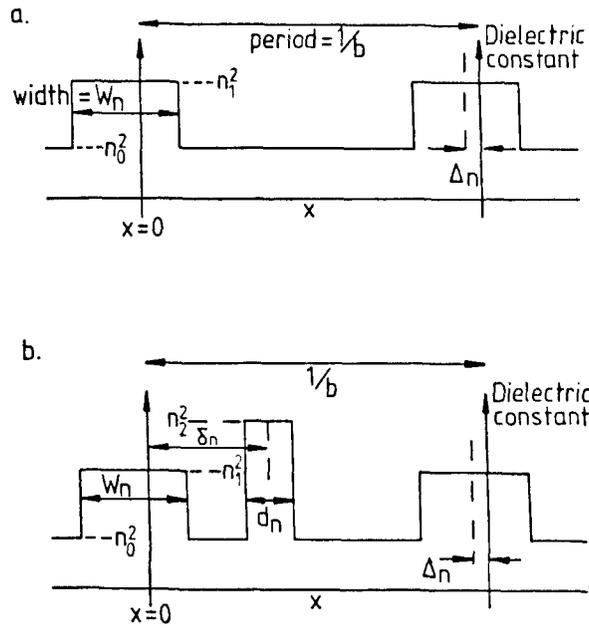
Figure 6.2.3. Dielectric constant profiles of a) a single line repeated with period 1/b and b) a line with a small line of higher refractive index close beside it.

$$\text{curl } E = ik_o H \qquad (6.2.32)$$

$$H_x = \frac{-i}{\mu_o \omega} \cdot \frac{\partial E_y}{\partial z} \qquad (6.2.33)$$

$$H_{x0} = \frac{1}{\mu_o \omega}\left[-K_0 E_{00}^{+}\exp(-ik_o z) + \sum_{j=-J}^{J} E_{j0}^{-} K_j \exp(2\pi ijbx)\cdot\exp(iK_j z)\right]$$

$$(6.2.34)$$

Where,

$$K_j = k_o\left[1 - \lambda^2 j^2 b^2\right]^{1/2} \qquad (6.2.35)$$

Now in each of the layers the field is given in equation (6.2.27) as,

$$E_{yn}(x,z) = \sum_{m=-J}^{J}\left(A_{m,n}\exp(\alpha_{m,n}z) + A'_{m,n}\exp(-\alpha_{m,n}z)\right)\sum_{j=-J}^{J} B_{j,m,n}\exp(2\pi ijbx)$$

$$(6.2.36)$$

Where the subscript n indicates the nth layer.

Using Maxwell's relation in equation (6.2.33) the H-field may be written as,

$$H_{x,n}(x,z) = \frac{-i}{\mu_o \omega} \sum_{m=-J}^{J} \left\{ \left[A_{m,n} \alpha_{m,n} \exp(\alpha_{m,n}z) - A'_{m,n} \alpha_{m,n} \exp(-\alpha_{m,n}z) \right] \right.$$

$$\left. \times \sum_{j=-J}^{J} B_{j,m,n} \exp(2\pi i j b x) \right\} \qquad (6.2.37)$$

The substrate consists only of the transmitted wave and so this may be written as,

$$E_{y,s} = \sum_{j=-J}^{J} E_{j,s} \exp(2\pi i j b x) \cdot \exp(-iK_{j,s}z) \qquad (6.2.38)$$

and again using Maxwell's relation between the E and H fields gives,

$$H_{x,s} = \frac{1}{\mu_o \omega} \left[\sum_{j=-J}^{J} -K_{j,s} E_{j,s} \exp(2\pi i j b x) \cdot \exp(-iK_{j,s}z) \right] \qquad (6.2.39)$$

where,
$$K_{j,s} = k_o \left[n_s^2 - \lambda^2 j^2 b^2 \right]^{1/2} \qquad (6.2.40)$$

At each interface the E_y and H_x fields must be continuous and thus,

$$H_{x,n}(x) = H_{x,(n+1)}(x) \quad \text{and} \quad E_{y,n}(x) = E_{y,(n+1)}(x) \qquad (6.2.41)$$

Consider the first interface (z=0) which consists of air on the top and the upper layer of the line object on the bottom. According to equation (6.2.41) it is necessary to equate equations (6.2.31) and (6.2.36) and also equations (6.2.34) and (6.2.37). Setting z=0 in equations (6.2.31) and (6.2.36) gives,

$$E_{y0} = E_{00}^+ + \sum_{j=-J}^{J} E_{j0}^- \exp(2\pi i j b x) \qquad (6.2.42)$$

and

$$E_{y1} = \sum_{m=-J}^{J} \left[A_{m,1} + A'_{m,1} \right] \sum_{j=-J}^{J} B_{j,m,1} \exp(2\pi i j b x) \qquad (6.2.43)$$

Now since the fields must equate for all diffraction orders, these equations may be combined to give,

$$E_{00}^+ + E_{00}^- = \sum_{m=-J}^{J} \left[A_{m,1} + A'_{m,1} \right] B_{0,m} \qquad (6.2.44)$$

This represents the on axis incident field together with the reflected

on axis field. The off-axis reflected field will thus be given by,

$$E_{j,0}^{-} = \sum_{m=-J}^{J} (A_{m,1} + A'_{m,1}) B_{j,m} \qquad j \neq 0 \qquad (6.2.45)$$

Nyyssonen (1982d) considered only the on-axis incident field and by setting $E_{00}^{+} = 1$, expressed the reflected field as,

$$E_{j,0}^{-} = \sum_{m=-J}^{J} (A_{m,1} + A'_{m,1}) B_{j,m} - \delta_{j} \qquad (6.2.46)$$

Where δ_{j} is a delta function such that,

$$\delta_{j} = \begin{cases} 1 & j=0 \\ 0 & j\neq0 \end{cases} \qquad (6.2.47)$$

Equation (6.2.46) describes the reflected field at the top of the object. It is this field which is imaged by the optical microscope.

Similarly equations (6.2.34) and (6.2.37) can be equated and this leads to the expression,

$$K_{j} E_{j0}^{-} = -i \sum_{m=-J}^{J} (A_{m,1} \alpha_{m,1} - A'_{m,1} \alpha_{m,1}) B_{j,m} + \delta_{j} K_{j} \qquad (6.2.48)$$

Multiplying equation (6.2.46) by K_{j} and adding to equation (6.2.48) enables these expressions to be written as,

$$K_{j} \sum_{m=-J}^{J} (A_{m,1} + A'_{m,1}) B_{j,m} + i \sum_{m=-J}^{J} (A_{m,1} \alpha_{m,1} - A'_{m,1} \alpha_{m,1}) B_{j,m} = 2\delta_{j} K_{j}$$

$$(6.2.49)$$

$$\sum_{m=-J}^{J} (K_{j} + i\alpha_{m,1}) B_{j,m} A_{m,1} + \sum_{m=-J}^{J} (K_{j} - i\alpha_{m,1}) B_{j,m} A'_{m,1} = 2\delta_{j} K_{j}$$

$$(6.2.50)$$

This may be written in matrix form.

$$[M] \bar{A}_{1} + [M^{-}] \bar{A}'_{1} = R \qquad (6.2.51)$$

Where,

$$R = \begin{cases} 2k_{o} & j=0 \\ 0 & j\neq0 \end{cases} \qquad (6.2.52)$$

and the elements of M and M' are given by,

$$m_{j,m} = \left(K_j + i\alpha_{m,1}\right)B_{j,m} \qquad (6.2.53)$$

$$m'_{j,m} = \left(K_j - i\alpha_{m,1}\right)B_{j,m} \qquad (6.2.54)$$

$\bar{A}_1$ and $\bar{A}'_1$ are vectors consisting of the elements $A_{m,1}$ and $A'_{m,1}$, where $-J \langle m \langle J$. This produces $(2J+1)$ equations for $2(2J+1)$ unknowns.

So far only the interface between the top layer and the air region has been considered. Burckhardt obtained another $(2J+1)$ equations by considering the interface between the bottom layer and the substrate. However this will only hold for objects which are constant with z from the top to the bottom. The objects of interest here are not constant with z but can be modelled by objects which are constant over small intervals in z.

We will continue with Burckhardt's analysis for the bottom interface and then divert to include the intermediate layers. According to equations (6.2.36) and (6.2.37), the field within the bottom layer at the interface within the substrate is given by,

$$E_{y,N}(x,T) = \sum_{m=-J}^{J} \left(A_{m,N}\exp(\alpha_{m,N}T) + A'_{m,N}\exp(-\alpha_{m,N}T)\right) \sum_{j=J}^{J} B_{j,m,N}\exp(2\pi i jbx)$$

$$(6.2.55)$$

and

$$H_{x,N}(x,T) = \frac{-i}{\mu_o \omega} \sum_{m=-J}^{J} \left[A_{m,N}\alpha_{m,N}\exp(\alpha_{m,N}T) - A'_{m,N}\alpha_{m,N}\exp(-\alpha_{m,N}T)\right]$$

$$\times \sum_{j=-J}^{J} B_{j,m,N} \cdot \exp(2\pi i jbx) \qquad (6.2.56)$$

Where T is the overall thickness of the patterned layer.

According to equations (6.2.38) and (6.2.39), the field within the substrate at the interface is given by,

$$E_{y,s}(x,T) = \sum_{j=-J}^{J} E_{j,s}\exp(-iK_{j,s}T) \cdot \exp(2\pi i jbx) \qquad (6.2.57)$$

and

$$H_{x,s}(x,T) = \frac{1}{\mu_o \omega} \sum_{j=-J}^{J} -K_{j,s} E_{j,s}\exp(-iK_{j,s}T) \cdot \exp(2\pi i jbx) \qquad (6.2.58)$$

Equating the E_y and H_x fields for each order at this interface gives,

$$\sum_{m=-J}^{J} \left[A_{m,N} \exp(\alpha_{m,N} T) + A'_{m,N} \exp(-\alpha_{m,N} T) \right] \sum_{j=-J}^{J} B_{j,m,N} = \sum_{j=-J}^{J} E_{j,s} \exp(-iK_{j,s} T)$$

(6.2.59)

and

$$\sum_{m=-J}^{J} \left[A_{m,N} \alpha_{m,N} \exp(\alpha_{m,N} T) - A'_{m,N} \alpha_{m,N} \exp(-\alpha_{m,N} T) \right] \sum_{j=-J}^{J} B_{j,m,N}$$

$$= i \sum_{j=-J}^{J} -K_{j,s} E_{j,s} \exp(-iK_{j,s} T)$$

(6.2.60)

Now the fields must equate for each mode separately and so these equations may be rewritten as,

$$\sum_{m=-J}^{J} \left[A_{m,N} \exp(\alpha_{m,N} T) + A'_{m,N} \exp(-\alpha_{m,N} T) \right] B_{j,m,N} = E_{j,s} \exp(-iK_{j,s} T) \quad \text{for all } j$$

(6.2.61)

and

$$\sum_{m=-J}^{J} \left[A_{m,N} \alpha_{m,N} \exp(\alpha_{m,N} T) - A'_{m,N} \alpha_{m,N} \exp(-\alpha_{m,N} T) \right] B_{j,m,N}$$

$$= -iK_{j,s} E_{j,s} \exp(-iK_{j,s} T) \quad \text{for all } j$$

(6.2.62)

Rearranging and eliminating $E_{j,s}$ gives,

$$\sum_{m=-J}^{J} \left[(iK_{j,s} + \alpha_{m,N}) \exp(\alpha_{m,N} T) B_{j,m,N} A_{m,N} + (iK_{j,s} - \alpha_{m,N}) \exp(-\alpha_{m,N} T) B_{j,m,N} A'_{m,N} \right] = 0$$

(6.2.63)

This may be written in matrix form,

$$[N] \, \bar{A}_N + [N^-] \, \bar{A}'_N = \bar{0}$$

(6.2.64)

Where the elements of the N and N' matrices are given by,

$$n(j,m) = (K_{j,s} - i\alpha_{m,N}) \exp(\alpha_{m,N} T) \, B_{j,m,N}$$

(6.2.65)

and

$$n'(j,m) = (K_{j,s} + i\alpha_{m,N}) \exp(-\alpha_{m,N} T) \, B_{j,m,N}$$

(6.2.66)

Burckhardt's analysis related to a single patterned layer (ie. constant with depth) and so $A_{m,N} = A'_{m,1}$ and $A_{m,N} = A'_{m,1}$. For this case, equation (6.2.64) would generate (2J+1) equations in 2(2J+1) unknowns which when combined with equation (6.2.51) would provide enough equations to solve for all the A_m and A'_m of the patterned layer. However we are interested in multilayer structures and so equations (6.2.51) and (6.2.64) are insufficient to solve for the $A_{m,1}$ and $A'_{m,1}$ terms. It is at this point that we diverge from Burckhardt's analysis in order to develop a method for relating the E_y and H_x field equations of adjacent layers. If this can be done then it will be possible to relate the boundary equations in equation (6.2.64) to the interface immediately above the substrate. This in turn may be related to the next interface up and so on, until the top layer is reached and then the $A_{m,1}$ and $A'_{m,1}$ terms may be solved.

In order to generate a general relationship, consider the interface between the nth and (n+1)th layers. From equations (6.2.36) and (6.2.37) expressions for the E_y and H_x fields in the nth layer can be written.

$$E_{y,n}(x,z) = \sum_{m=-J}^{J} \left[A_{m,n} \exp(\alpha_{m,n}z) + A'_{m,n}\exp(-\alpha_{m,n}z) \right] \sum_{j=-J}^{J} B_{j,m,n}\exp(2\pi ijbx)$$

$$(6.2.67)$$

and

$$H_{x,n}(x,z) = \frac{-i}{\mu_o\omega} \sum_{m=-J}^{J} \left[A_{m,n}\alpha_{m,n}\exp(\alpha_{m,n}z) - A'_{m,n}\alpha_{m,n}\exp(-\alpha_{m,n}z) \right]$$

$$\times \sum_{j=-J}^{J} B_{j,m,n}\exp(2\pi ijbx) \qquad (6.2.68)$$

Similarly in the (n+1)th layer.

$$E_{y,(n+1)}(x,z) = \sum_{m=-J}^{J} \left[A_{m,(n+1)}\exp(\alpha_{m,(n+1)}z) + A'_{m,(n+1)}\exp(-\alpha_{m,(n+1)}z) \right]$$

$$\times \sum_{j=-J}^{J} B_{j,m,(n+1)}\exp(2\pi ijbx) \qquad (6.2.69)$$

and

$$H_{x,(n+1)}(x,z) = \frac{-i}{\mu_o\omega} \sum_{m=-J}^{J} \left[A_{m,(n+1)}\alpha_{m,(n+1)}\exp(\alpha_{m,(n+1)}z) \right.$$

$$- A'_{m,(n+1)} \alpha_{m,(n+1)} \exp(-\alpha_{m,(n+1)} z)] \sum_{j=-J}^{J} B_{j,m,(n+1)} \exp(2\pi i j b x) \quad (6.2.70)$$

The fields must be equivalent for each individual mode and order and so the following equalities must hold.

$$\sum_{m=-J}^{J} [A_{m,n} \exp(\alpha_{m,n} z) + A'_{m,n} \exp(-\alpha_{m,n} z)] B_{j,m,n} = \sum_{m=-J}^{J} [A_{m,(n+1)}$$

$$\times \exp(\alpha_{m,(n+1)} z) + A'_{m,(n+1)} \exp(-\alpha_{m,(n+1)} z)] B_{j,m,(n+1)} \quad (6.2.71)$$

and

$$\sum_{m=-J}^{J} [A_{m,n} \alpha_{m,n} \exp(\alpha_{m,n} z) - A'_{m,n} \alpha_{m,n} \exp(-\alpha_{m,n} z)] B_{j,m,n} = \sum_{m=-J}^{J} [A_{m,(n+1)} \alpha_{m,(n+1)}$$

$$\times \exp(\alpha_{m,(n+1)} z) - A'_{m,(n+1)} \alpha_{m,(n+1)} \exp(-\alpha_{m,(n+1)} z)] B_{j,m,(n+1)} \quad (6.2.72)$$

These two sets of equations may be written in matrix form.

$$[P_{n,(n+1)}] \bar{A}_n + [P'_{n,(n+1)}] \bar{A}'_n = [P_{(n+1),n}] \bar{A}_{(n+1)}$$

$$+ [P'_{(n+1),n}] \bar{A}'_{(n+1)} \quad (6.2.73)$$

and

$$[Q_{n,(n+1)}] \bar{A}_n + [Q'_{n,(n+1)}] \bar{A}'_n = [Q_{(n+1),n}] \bar{A}_{(n+1)}$$

$$+ [Q'_{(n+1),n}] \bar{A}'_{(n+1)} \quad (6.2.74)$$

Where $P_{n,(n+1)}, P'_{n,(n+1)}, Q_{n,(n+1)}$ and $Q'_{n,(n+1)}$ represent the nth layer at the (n+1)th interface and $P_{(n+1),n}, P'_{(n+1),n}, Q_{(n+1),n}$ and $Q'_{(n+1),n}$ represent the (n+1)th layer at the (n+1)th interface. The elements of these matrices are given by,

$$P_{n,(n+1)}(j,m) = B_{j,m,n} \exp(\alpha_{m,n} z) \quad (6.2.75)$$

$$P'_{n,(n+1)}(j,m) = B_{j,m,n} \exp(-\alpha_{m,n} z) \quad (6.2.76)$$

$$P_{(n+1),n}(j,m) = B_{j,m,(n+1)} \exp(\alpha_{m,(n+1)} z) \quad (6.2.77)$$

$$P'_{(n+1),n}(j,m) = B_{j,m,(n+1)} \exp(-\alpha_{m,(n+1)} z) \quad (6.2.78)$$

$$q_{n,(n+1)}(j,m) = B_{j,m,n} \alpha_{m,n} \exp(\alpha_{m,n} z) \quad (6.2.79)$$

$$q'_{n,(n+1)}(j,m) = -B_{j,m,n} \alpha_{m,n} \exp(-\alpha_{m,n} z) \quad (6.2.80)$$

$$q_{(n+1),n}(j,m) = B_{j,m,(n+1)}\alpha_{m,n}\exp(\alpha_{m,(n+1)}z) \qquad (6.2.81)$$

$$q'_{(n+1),n}(j,m) = -B_{j,m,(n+1)}\alpha_{m,n}\exp(-\alpha_{m,(n+1)}z) \qquad (6.2.82)$$

The matrix equations (6.2.73) and (6.2.74) may be combined into a single equation.

$$\begin{bmatrix} P_{n,(n+1)} & P'_{n,(n+1)} \\ Q_{n,(n+1)} & Q'_{n,(n+1)} \end{bmatrix} \cdot \begin{bmatrix} \bar{A}_n \\ \bar{A}'_n \end{bmatrix} = \begin{bmatrix} P_{(n+1),n} & P'_{(n+1),n} \\ Q_{(n+1),n} & Q'_{(n+1),n} \end{bmatrix} \cdot \begin{bmatrix} \bar{A}_{(n+1)} \\ \bar{A}'_{(n+1)} \end{bmatrix}$$

$$(6.2.83)$$

This equation relates the field coefficients of one layer to the next layer. At this point it is convenient to introduce a more compact notation.

$$[PQ_{n,(n+1)}] \cdot \begin{bmatrix} \bar{A}_n \\ \bar{A}'_n \end{bmatrix} = [PQ_{(n+1),n}] \cdot \begin{bmatrix} \bar{A}_{(n+1)} \\ \bar{A}'_{(n+1)} \end{bmatrix} \qquad (6.2.84)$$

Hence,

$$\begin{bmatrix} \bar{A}_{(n+1)} \\ \bar{A}'_{(n+1)} \end{bmatrix} = [PQ_{(n+1),n}]^{-1} \cdot [PQ_{n,(n+1)}] \cdot \begin{bmatrix} \bar{A}_n \\ \bar{A}'_n \end{bmatrix} \qquad (6.2.85)$$

This recursive relationship may be used to relate the substrate interface equations in equation (6.2.64) with the top layer interface equations in equation (6.2.51).

$$\begin{bmatrix} \bar{A}_N \\ \bar{A}'_N \end{bmatrix} = [PQ_{N,(N-1)}]^{-1} \cdot [PQ_{(N-1),N}] \cdots [PQ_{(n+1),n}]^{-1} \cdot [PQ_{n,(n+1)}] \cdots$$

$$[PQ_{3,2}]^{-1} \cdot [PQ_{2,3}] \cdot [PQ_{2,1}]^{-1} \cdot [PQ_{1,2}] \cdot \begin{bmatrix} \bar{A}_1 \\ \bar{A}'_1 \end{bmatrix} \qquad (6.2.86)$$

For an N layer object, equation (6.2.86) will contain 2(N-1) matrices. These occur in pairs as shown and if two adjacent layers are identical then the matrix pair will reduce to the identity matrix. Each matrix will be a 2(2J+1)×2(2J+1) non-singular square matrix. This chain of matrices may be used to form a single matrix and this reduces equation (6.2.86) to,

$$\begin{bmatrix} \bar{A}_N \\ \bar{A}'_N \end{bmatrix} = [PQ] \cdot \begin{bmatrix} \bar{A}_1 \\ \bar{A}'_1 \end{bmatrix} \qquad (6.2.87)$$

This may be substituted into equation (6.2.64) to give,

$$[N \ N'] \cdot [PQ] \cdot \begin{bmatrix} \bar{A}_1 \\ \bar{A}'_1 \end{bmatrix} = \bar{0} \qquad (6.2.88)$$

Since matrix multiplication is distributive, this may be written as a new matrix equation.

$$[NP \ NP'] \cdot \begin{bmatrix} \bar{A}_1 \\ \bar{A}'_1 \end{bmatrix} = \bar{0} \qquad (6.2.89)$$

Where,

$$[NP \ NP'] = [N \ N'] \cdot [PQ] \qquad (6.2.90)$$

Finally equations (6.2.89) and (6.2.51) can be combined into a single matrix equation.

$$\begin{bmatrix} M & M' \\ NP & NP' \end{bmatrix} \cdot \begin{bmatrix} \bar{A}_1 \\ \bar{A}'_1 \end{bmatrix} = \begin{bmatrix} R \\ \bar{0} \end{bmatrix} \qquad (6.2.91)$$

This now brings us back to the same form as Burckhardt's analysis. Equation (6.2.91) represents a set of $2(2J+1)$ linear simultaneous equations in $2(2J+1)$ unknowns. These equations may now be solved to find the $(2J+1)$ A_m terms and $(2J+1)$ A'_m terms for the top layer at the air interface.

Following Nyyssonen's formulation for single layer structures, the Fourier series of the reflected E_y field may be written as,

$$g(j) = \sum_{m=-J}^{J} [A_{m,1} + A'_{m,1}] B_{j,m,1} - \delta_j \qquad (6.2.92)$$

Where, $g(j)$ is the Fourier coefficient of the jth order.

Thus the reflected E_y field at $z=0$ is given by,

$$E_{yR}(x,0) = \sum_{j=-J}^{J} g(j) \cdot \exp(2\pi i j b x) \qquad (6.2.93)$$

This field can now be regarded as the field due to a planar pseudo-object and the optical image may be computed using classical scalar diffraction theory. The $g(j)$ terms may be used in equation (2.3.29) in order to compute a partially coherent image.

6.2.3 Limitations of the multi-layer model.

There are several assumptions on which this model is based which impose limitations on its application and accuracy. It was assumed that the illumination is of a single wavelength, is normally incident on the object and is polarised such that the E-field component lies parallel to the lines. In practice none of these will be exactly true and making these assumptions restricts the range of problems to which the model may be applied.

a) Monochromatic illumination.

The diffraction effects in an optical microscope are wavelength dependent, but change sufficiently slowly that the illumination from a green filter with a bandwidth of 0.1µm may be assumed to be monochromatic. However, the field profile across the top surface of a thick layer object is very sensitive to wavelength and this sensitivity increases with thickness and refractive index. The model is restricted to features which do not have long optical path lengths.

In order to be able to obtain a good match between practical and theoretical profiles it is necessary to use very narrow bandwidth illumination as the image profile can change quite significantly for even a small change in wavelength. It is possible to compute the image profile for broad band illumination by summing a series of profiles for different wavelengths.

b) Normally incident illumination.

The illumination has been assumed to be normally incident but this can only be true for completely coherent illumination. In practice the illumination aperture will be finite and so the illumination will be incident at different angles. The image profile is a function of illumination angle and the sensitivity increases with thickness and refractive index. The validity of this assumption thus decreases as the features become thicker.

The field profile is only computed for the case of normal incidence and is then assumed to be constant for the different illumination angles in order that a partially coherent image can be computed.

c) H-mode polarisation.

The analysis has been restricted to the H-mode incident wave, but in practice the illumination will comprise of both the E and H modes. For normal incidence, both modes produce the same image profile (Burckhardt (1966)) but when there is a finite illumination angle this is only an approximation.

d) Repetition period of the object.

The analysis presented here relies on being able to express the object in terms of a Fourier series and so it is assumed that the object is periodic. In practice very few structures will even approximate to a periodic grating and so it is necessary to use a large period in order to model isolated structures.

The image profile contains fringes near the line edges which are due to both diffraction and waveguide effects. Diffraction fringes are small and fall off rapidly away from the line edge. However, the guided wave fringes produced by thick layer objects can propagate a considerable distance away from a line edge and so the period needs to be large in order to avoid interference from adjacent lines. The greater the period, the more orders will be imaged and so the matrices produced by the algorithm will become very large. This greatly slows down the computation and reduces the accuracy.

e) The finite layer approximation.

It is assumed that the object can be split into a finite number of layers where each layer is constant in z. The problem is not ill-posed and so as the number of layers is increased and adjacent layers become more and more similar, the product of the corresponding matrices tends to the identity matrix.

If too few layers are used then the model will be a poor approximation to the actual object, but if too many are used, then the computation will take a long time and will be less accurate.

f) One dimensional object structures.

It has been implicitly assumed that the line object is truly one dimensional in equation 6.2.1. If the object is not one dimensional then the scalar product of gradɛ and gradE is not zero and so

Maxwell's differential wave equation does not reduce to the form of Hill's equation.

The model cannot be applied to short lines or where the lines have ragged edges. It is not possible to model ragged edges by integrating over a range of linewidths as the form of Maxwell's differential wave equation will not be valid.

g) Truncating the field modes.

In order to limit the solution to finite sized matrices, it was argued that only those modes leaving the surface of the object need be considered. This is not true as there will be higher order modes which propagate within the object. Although they will not contribute directly to the image, they will perturb the field and this will affect the lower orders which are imaged.

According to equation 6.2.25, all the modes are coupled by the Fourier series of the dielectric constant profile. Since this series is infinite, it follows that all the modes are coupled to all the others and so truncating the higher orders (effectively setting them to zero) will introduce an error into the computed image. Each mode is most tightly coupled to its immediate neighbours and so it may be necessary to discard the highest computed orders as they will be most in error.

6.3 Computational methods for the calculation of image profiles.

6.3.1 Defining edge shapes.

The algorithm requires that the object is split up into layers, as in figure 6.2.2. For each layer, the width, offset, refractive index and depth must be defined. Nyyssonen (1982d) suggested a useful notation for edge shapes which was later used (Kirk and Nyyssonen (1985)) to compare edge effects. The linewidth can be defined as a polynomial function of depth.

$$W(z) = \sum_{i=0}^{R} X_i (z - Z_i)^i \qquad (6.3.1)$$

R defines the order of the edge and the parameters X_i and Z_i define the coefficients and offsets of the polynomial. Figure 6.3.1 shows how many commonly encountered edge shapes can be defined using this approach.
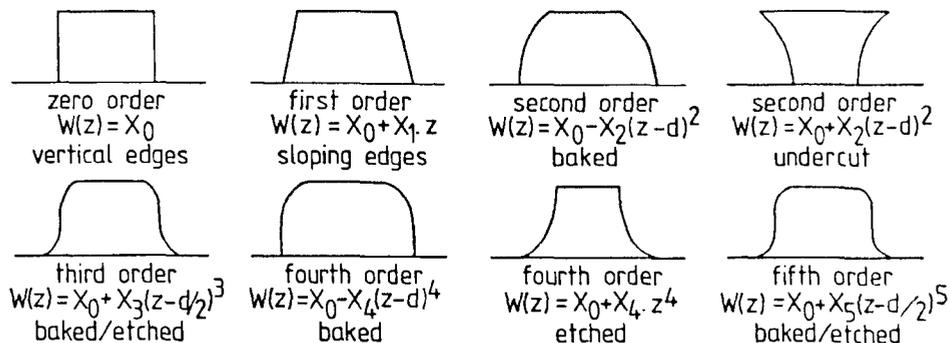
Figure 6.3.1. Polynomial models for typical line shapes.

These shapes are frequently encountered in different stages of wafer processing and they have been assigned generic names to indicate the typical areas where they are found. This method of defining an edge is convenient as complex edge shapes can be defined by only a few parameters and it enables quantitative comparisons between different edges to be made.

The line should be sectioned into layers which are sufficiently thin that dividing them up further will make no significant difference to the image profile. The layers need not all be of the same thickness and can be made a function of the edge shape in order to optimise the number of layers used. This can be achieved by a simple algorithm which starts a new layer every time the linewidth has changed by a certain amount (as a function of z).

6.3.2 Program architecture.

Once the object has been sectioned into layers, it is necessary to compute the Fourier series of the dielectric constant profile as described by equations 6.2.28 and 6.2.29. The diffracted orders all lie between $-J$ and J and so the series may be bounded by these limits.

The Fourier series for each layer can then be used to construct the D matrix according to equation 6.2.26. Each D matrix will be a $(2J+1)\times(2J+1)$ matrix and the α_m and $B_{j,m}$ terms for equations 6.2.21 and 6.2.22 can be found from the eigenvalues and eigenvectors of this D matrix. The $\bar{B}_m$ vectors are the eigenvectors of this matrix and the α_m terms are the square roots of the eigenvalues. This process must be repeated for each layer.

The boundary conditions can now be brought into the solution. The M and M$^-$ matrices corresponding to the top layer boundary conditions in

equation 6.2.51 can be constructed using equations 6.2.53 and 6.2.54. Likewise the $\left[PQ_{n,(n+1)}\right]$ and $\left[PQ_{(n+1),n}\right]^{-1}$ matrices may be constructed for all the interfaces between the layers using equations 6.2.75 to 6.2.82. The $\left[PQ\right]$ matrix can then be constructed from the daisy chain of matrices in equation 6.2.86.

The N and N˘ matrices described by equation 6.2.64 can be constructed according to equations 6.2.65 and 6.2.66. The $\left[N\ N˘\right]$ and $\left[PQ\right]$ matrices can now be combined according to equation 6.2.90 and the $\left[NP\ NP˘\right]$ matrix computed. This in turn can be combined with the $\left[M\ M˘\right]$ matrix as in equation 6.2.91 and used to solve for the $\overline{A}_1$ and $\overline{A}'_1$ vectors. Finally the Fourier coefficients of the field of the pseudo-object may be computed from these vectors together with the eigenvectors of the first layer, according to equation 6.2.92.

This algorithm was implemented in the form shown in the flow chart in figure 6.3.2[†].

6.4 Testing the model.

The model is based on certain assumptions which have already been discussed and it is also reliant upon computational accuracy. This restricts the range of problems to which it can be applied and so it is necessary to test the program in order to explore where it begins to break down. The most crucial test for the theory is whether or not it correctly predicts the image intensity profiles which are seen in practice. It is difficult to make direct tests of the theory against practical image profiles as there are so many parameters which affect the profile shape and many of them can be difficult to determine. Instead it is useful to start with some theoretical tests which enable a quantitative assessment of the accuracy of the model and its software implementation to be obtained.

6.4.1 Comparison with the thin layer model.

The thin layer model was described in chapter 2 and good agreement with experimental results has been obtained. If the object consists of a thin layer, then the thick and thin layer models should produce the same image profile. As the patterned layer becomes thinner, the

† This software was written in Fortran 77 on the National Bureau of Standards˘ (NBS), Sperry Univac and is not subject to copyright.

```
┌──────────┐      ┌──────────────┐
│  Input   │ ───▶ │ Set up the   │
│data file │      │sectioned line│
└──────────┘      └──────────────┘
                         │
                         ▼
                  ┌──────────────┐
                  │  Do  n=1 to N│ ──────────────┐
                  └──────────────┘               │
                         │                        │
                         ▼                        │
                  ┌──────────────┐                │
                  │Compute theFourier│            │
                  │series for layer n│            │
                  └──────────────┘                │
                         │                        │
                         ▼                        │
                  ┌──────────────┐                │
                  │Set up the D  │                │
                  │matrix for layer n│            │
                  └──────────────┘                │
   ┌──────────┐          │                        │
   │Store the │          ▼                        │
   │αm terms  │ ───▶ ┌──────────────┐             │
   └──────────┘      │Find the eigenvalues│       │
                     │of the D matrix│            │
   ┌──────────┐      └──────────────┘             │
   │Store the │          │                        │
   │Bm vectors│ ───▶ ┌──────────────┐             │
   └──────────┘      │Find the eigenvectors│ ─────┘
                     │of the D matrix│
                     └──────────────┘
                         │
                         ▼
                  ┌──────────────┐
                  │Set up the M  │
                  │and M' matrices│
                  └──────────────┘
                         │
                         ▼
                  ┌──────────────┐
             else │  If N>1 then │
            ◀─────└──────────────┘
            │            │
            │            ▼
            │     ┌──────────────┐
            │     │T = Identity matrix│
            │     └──────────────┘
            │            │
            │            ▼
            │     ┌──────────────┐
            │     │ Do n =1 to N-1│ ───────┐
            │     └──────────────┘         │
            │            │                  │
            │            ▼                  │
            │     ┌──────────────┐          │
            │     │Set up [PQn,(n+1)]│      │
            │     └──────────────┘          │
            │            │                  │
            │            ▼                  │
            │     ┌──────────────┐          │
            │     │Set up [PQ(n+1),n]│      │
            │     └──────────────┘          │
            │            │                  │
            │            ▼                  │
            │     ┌──────────────┐          │
            │     │Find [PQ(n+1),n]⁻¹│      │
            │     └──────────────┘          │
            │            │                  │
            │            ▼                  │
            │     ┌──────────────┐          │
            │     │T=[PQ(n+1),n]⁻¹[PQn,(n+1)].T│ ─┘
            │     └──────────────┘
            │            │
            ▼            ▼
          ┌──────────────┐
          │Set up the N  │
          │and N' matrices│
          └──────────────┘
                 │
                 ▼
          ┌──────────────┐
          │Compute the   │
          │NP and NP' matrices│
          └──────────────┘
                 │
                 ▼
          ┌──────────────┐
          │Set up the [M M']│
          │matrix    [NP NP']│
          └──────────────┘
                 │
                 ▼
          ┌──────────────┐
          │Solve for     │
          │A1 and A1     │
          └──────────────┘
                 │
   ┌──────────┐  ▼
   │Store the │ ┌──────────────┐
   │Fourier series│◀│Compute the Fourier│
   └──────────┘ │series of the field│
                └──────────────┘
                      │
   ┌──────────┐       ▼
   │Store the │ ┌──────────────┐
   │image profile│◀│Compute the  │
   └──────────┘ │image profile │
                └──────────────┘
```
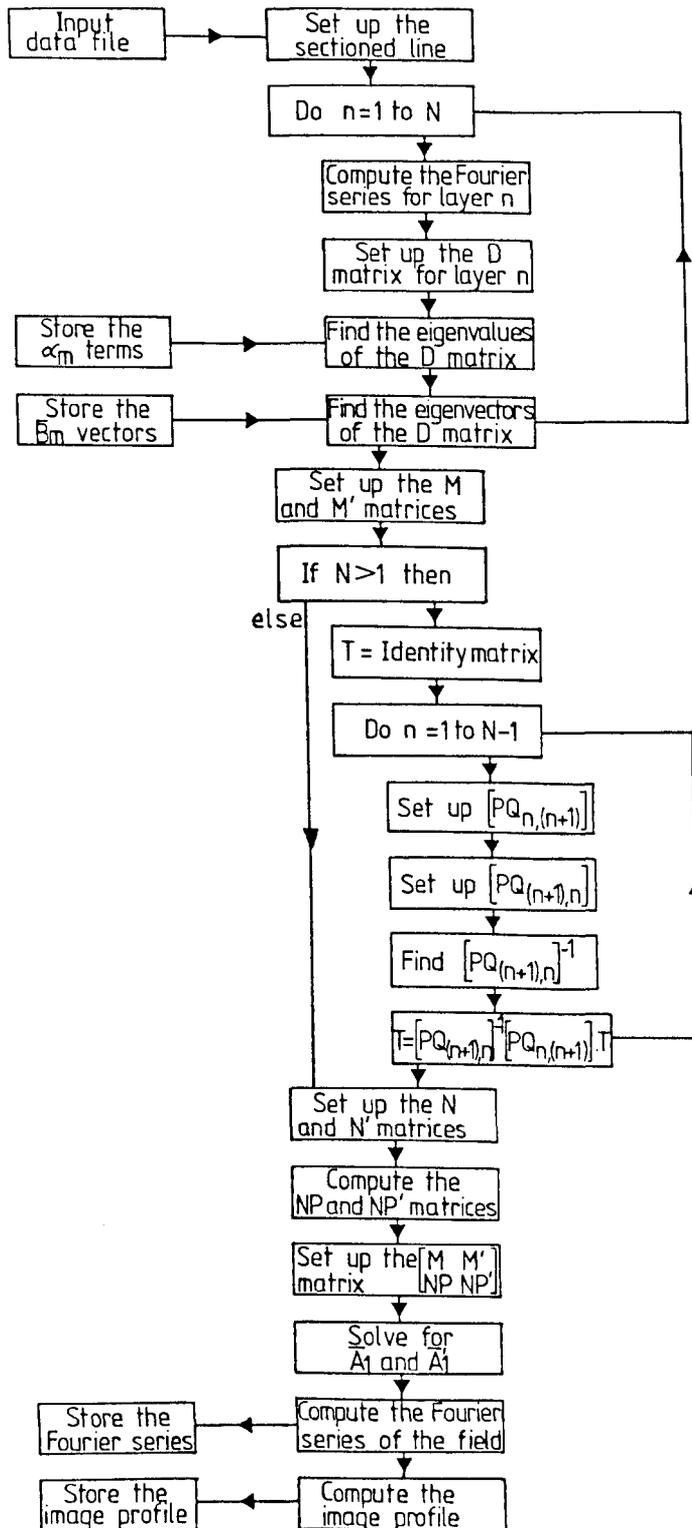
Figure 6.3.2. Flow chart of the program used to implement the
algorithm for computing thick layer image profiles.

agreement should improve. Complete agreement will not be obtained as
no layer is infinitely thin and so there will always be a small error
in the thin layer model.

Image profiles of a 6.0µm wide oxide line on silicon with a 12µm pitch
were generated for an illumination wavelength of 0.53µm. Two
thicknesses of 0.02µm and 0.05µm were considered and profiles were
computed using both the thin and thick layer models. The normalised
amplitude Fourier coefficients of the on-axis illumination field
profiles were compared and the two models were found to agree to four
decimal places. The agreement on the thinner layer was better than
the thicker layer.

The test was repeated for a 0.07µm thick metal layer with a refractive
index of 2.9+4.2i. This is a typical refractive index value for
chrome. The agreement was not as good as for the thin layer
dielectrics. This is because both the refractive index and thickness
have been increased. The amplitude Fourier coefficients agreed to
three decimal places in the lower orders.

The similarity between the results of the thin and thick layer models
is good enough to suggest that they are in agreement. The test cases
examined do not show fundamental inconsistencies between the models.
Second order disagreements are to be expected as both models are based
on different assumptions.

6.4.2 Comparison with the Fresnel equations.

The Fresnel equations enable the relative reflectivities of thick
layers to be computed and these should agree with the reflectivity
values predicted by the thick layer model. There are two methods for
examining the relative reflectances predicted by the thick layer
model. The most obvious approach is to define a patterned layer with
large areas of the coating material and large areas of the exposed
substrate. The relative intensities of these two regions in the image
should be equal to the relative reflectivity predicted by the Fresnel
equations.

The relative reflectivity for normal incidence is a function of
refractive index and the ratio of thickness to wavelength. Using a
6.0µm wide line on a 12µm pitch with 0.53µm wavelength illumination a
series of profiles was computed for different thicknesses of the
patterned layer. Table 6.4.1 shows how the relative reflectivities
measured from these curves compare with the values from the Fresnel
equations.

| Thickness in | Relative reflectivity | |
microns	Fresnel equations	Thick layer model
0.04	0.76	0.76
0.05	0.64	0.64
0.06	0.51	0.51
0.10	0.30	0.30
0.20	0.95	0.95
0.55	0.99	1.00
0.60	0.57	0.58
0.65	0.34	0.35
0.95	0.73	0.74
1.00	0.27	0.29
1.05	0.78	0.80
1.45	0.99	1.00
1.50	0.66	0.64
1.55	0.29	0.28

Table 6.4.1.   A comparison of the relative reflectivities of
a patterned layer of oxide on silicon using the
Fresnel equations and the thick layer model.

The results show that the thick layer model agrees with the Fresnel
equations for lines up to 1.5µm thick which is the region of interest
for this work. The difference between the two values increases with
thickness. This may not be real but could arise due to the difficulty
in measuring the reflectivity when large edge fringes are present.

The second method of comparing the thick layer model with the Fresnel
equations avoids the problem of oscillations in the image profile. If
the patterned layer is defined as having a width equal to the period,
then it becomes a continuous layer with no edges. The diffracted
field should then consist of a central order and no others. The
Fourier coefficients of the pseudo-object represent the complex
amplitude coefficients of the diffracted field. Thus the relative
reflected intensity can be found from the central diffraction orders
from the field from the layer ($F_m(0)$) and from the field from the
substrate ($F_s(0)$).

$$\text{Relative reflectance} = \frac{|F_m(0)|^2}{|F_s(0)|^2} \qquad (6.4.1)$$

This method was used to find the relative reflectance for different layer thicknesses and the results are shown in table 6.4.2.

Thickness in microns	Relative reflectance		Non-zero order intensity noise
	Fresnel equations	Thick layer model	
0.05	0.637	0.637	$< 10^{-16}$
0.60	0.566	0.566	$< 10^{-14}$
1.00	0.272	0.272	$< 10^{-12}$

Table 6.4.2. Comparison of reflectivities of oxide layers on silicon relative to the substrate for unpatterned layers.

Ideally the non-zero orders should not be present but in practice there will be some computational noise and it can be seen that this increases with layer thickness. The reflectivity values show extremely good agreement with the Fresnel equations, suggesting that a large part of the error incurred in the data in table 6.4.1 was due to the imprecision in determining the contrast from the image intensity profiles. These results show that this method of modelling sub-layers is very accurate. More importantly, it shows that the model agrees with the Fresnel equations.

6.4.3 Stability tests for multi-layer structures.

In section 6.2 it was stated that where two adjacent layers were identical, the matrix product linking them reduced to the identity matrix. This implies that no matter how many layers are used to define a vertical edge wall object, the image profile should always be the same. In practice, there will be numerical noise introduced due to the increased amount of computation.

In order to test the effect of different numbers of layers, the image profile of a 0.6μm thick oxide line was computed using one layer and ten layers. When the normalised Fourier coefficients of the complex amplitude fields were examined, they were found to agree to seven

decimal places. This suggests that the matrix operations introduce very little noise.

Having established that the series is stable for vertical edge wall structures where all the layers are the same, the test can now be extended to cover non-vertical edge wall structures. The sloping edge is approximated by a series of layers and the more layers used, the greater the accuracy. Provided that the algorithm is stable, the image profiles should converge to a limiting profile as the number of layers is increased. Eventually a point will be reached where increasing the number of layers further makes no significant difference to the image profile.

In order to test for this stability, a 6.0μm wide, 1.0μm thick oxide line with 45° edge slopes was modelled. Figure 6.4.1 shows the effect of increasing the number of layers used to define the object. The first curve represents the image which arises from one layer. The one layer definition approximates the sloping edges to vertical walls which is obviously unreasonable. The last curve represents the image of an object defined by 20 layers and this is the best approximation to the sloping edge walls. The curves show that as the number of layers increases, the images converge to one profile and so the algorithm is stable. Also the results show that there is little difference between the 10 layer and 20 layer models and so adding additional layers will not make any significant improvements to the accuracy.

As a rule of thumb, it appears it is adequate to define enough layers such that the step width at any one edge is less than one fifth of the wavelength of light. This enables most objects of up to about 2.0μm thick to be modelled with 20 layers or less. The program is restricted to 20 layers as its accuracy becomes doubtful on very thick layers.

6.4.4 Asymmetric object structures.

Extending the model to cover structures with sloping edges has enabled asymmetric objects to be considered. It was assumed in section 6.2 that the material in each layer did not have to lie symmetrically about the origin and this would allow asymmetric structures to be modelled.
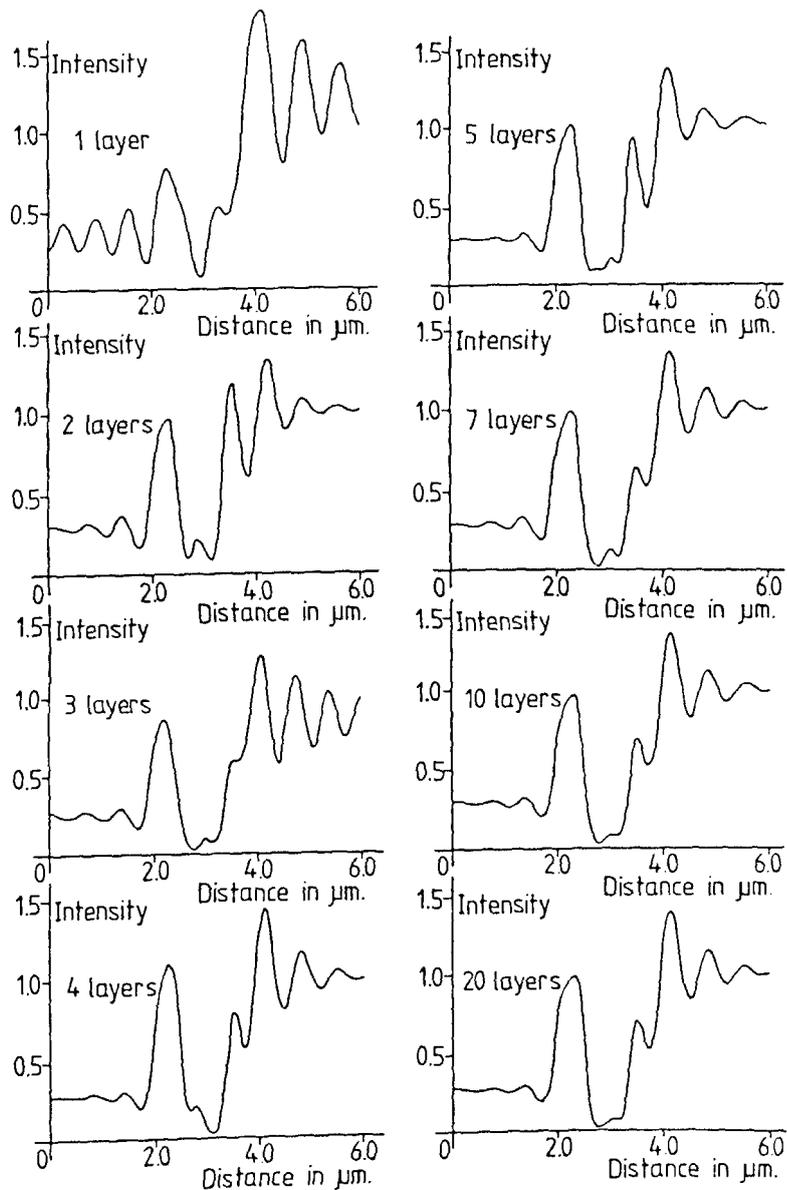
Figure 6.4.1. The effect on the image profile of increasing the number of layers used to define a line with sloping edges.

The first test of this assumption was to generate a set of image profiles for a single layer object, for a range of lateral offsets. It was found that shifting the layer, merely shifted the image profile and did not distort it in any way. Since the object is periodic, there is no point in shifting it more than half a period in either direction as it will simply be repeated.

The second test for modelling asymmetric objects involved generating the image profiles of three line objects. The first object had vertical edge walls, the second had 45° edge walls and the third had one vertical and one 45° edge wall. This third object was made up of a set of shifted layers (figure 6.4.2). If the line object is wide,
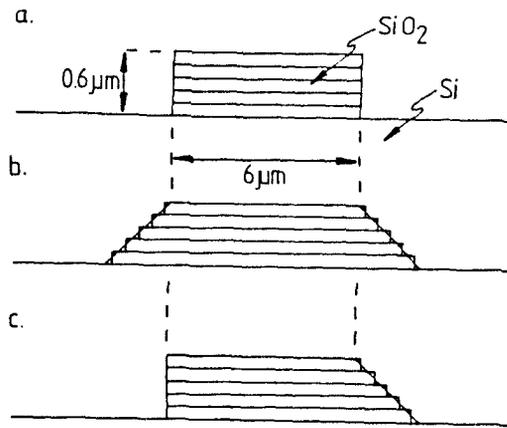
Figure 6.4.2. Layered models for a) vertical edges, b) sloping edges and c) an asymmetric object with sloping and vertical edges.

then the image profile of the asymmetric object should have edges which match the edge images in the respective symmetrical profiles.
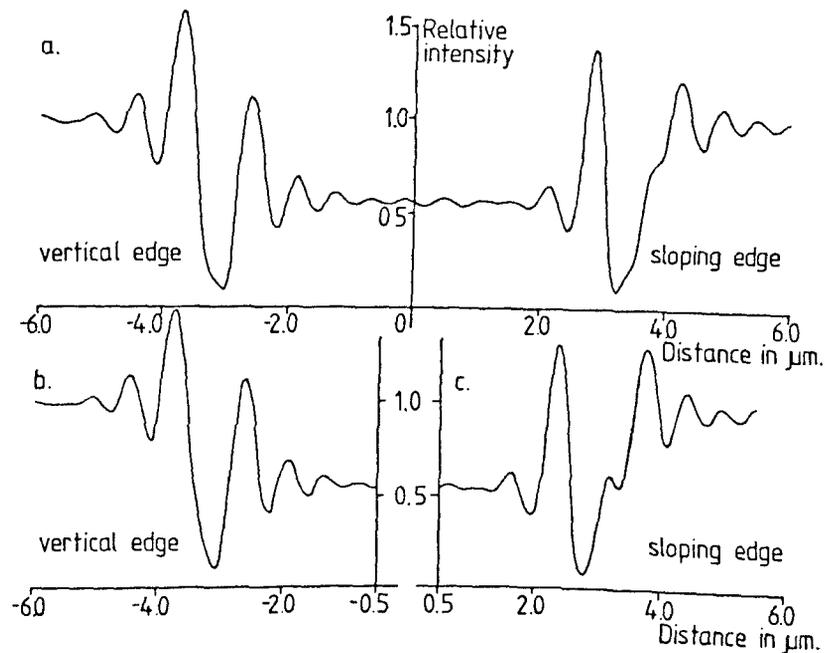
Figure 6.4.3. Image profiles of the, a) asymmetric, b) vertical and c) sloping edge wall objects, for a 600nm thick line.

Figure 6.4.3 shows the image profiles of the line objects defined in figure 6.4.2. Although the profiles in figure 6.4.3 show reasonable agreement, it is not exact and this is possibly due to the interference between the opposing edges. In order to test this, the images of 0.2μm thick line were generated and these showed much better agreement. Thinner layers propagate less energy away from the edge and so the interference is less.

6.4.5 Focus error tests.

Although no thick layer models exist for direct comparison there is
one special case which can be modelled in a different way. A focus
error can be regarded as an object with a layer of air on top of it.
Focus errors are normally modelled by modulating the objective
aperture phase response as described in equation 2.3.20. This
effectively relocates the focal plane of the objective and should be
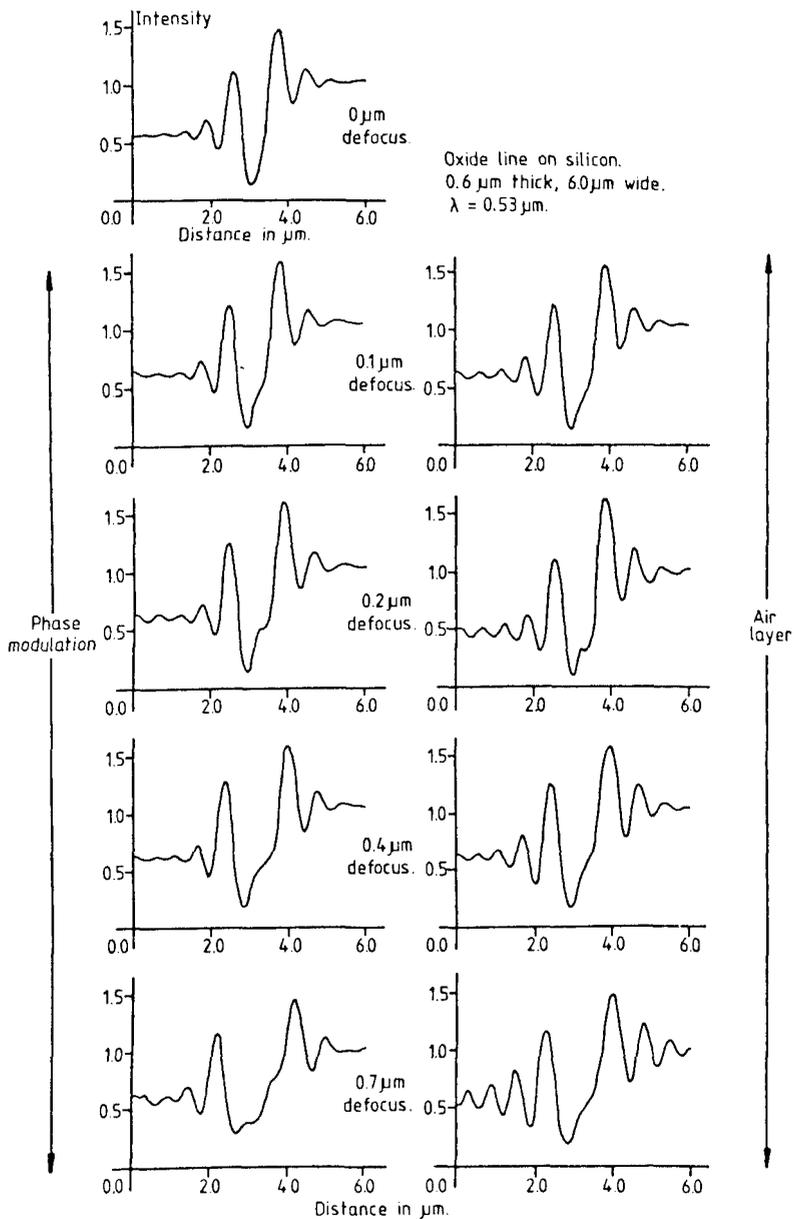equivalent to the layer of air.

Figure 6.4.4. A comparison of the thick layer model with
a phase modulated aperture defocus model.

Figure 6.4.4 shows a set of image profiles with focus errors modelled
by both approaches. The profiles show good agreement and demonstrate

the same trends. The agreement starts to break down when the focus error gets up to about 0.7µm. This represents an object thickness of 1.3µm and it may indicate that the model is not entirely accurate for very thick layers.

6.4.6. Finite order solutions - Stability and resolution.

In section 6.2 it was assumed that the number of orders considered in the formulation could be restricted to include only those which were diffracted into the far field and any higher order modes were omitted. If the problem is ill-conditioned then the solutions will not be stable and will be radically different if different truncation limits are chosen. In order to test for stability, a series of image intensity profiles were generated using different numbers of orders and these are shown in figure 6.4.5.

Under the conditions of this experiment, there are only 22 orders on either side of the optical axis which contribute to the far field. The first two figures in figure 6.4.5 show a comparison of the use of 24 and 22 orders. The extra two orders used for the 24 order profile do not leave the object. These higher orders contribute very little to the image except that they make the dark fringe close to the edge a little sharper. It would appear that there is no serious error incurred by neglecting the high orders which do not appear in the diffracted field on the side of the objective.

However reducing the number of orders significantly below the limit defined by equation 6.2.20 does have a noticeable effect on the image profile. The two most obvious effects are the broadening of the profile features and the loss of the bright fringes on either side of the edge. The broadening of the profile can be attributed to loss of resolution due to the restriction on the number of orders. The disappearance of the bright fringes occurs because they are products of the high order modes. Most importantly, the general shape of the profile is retained as the number of orders is reduced and it appears that there is no need to include the very high orders which do not leave the object.
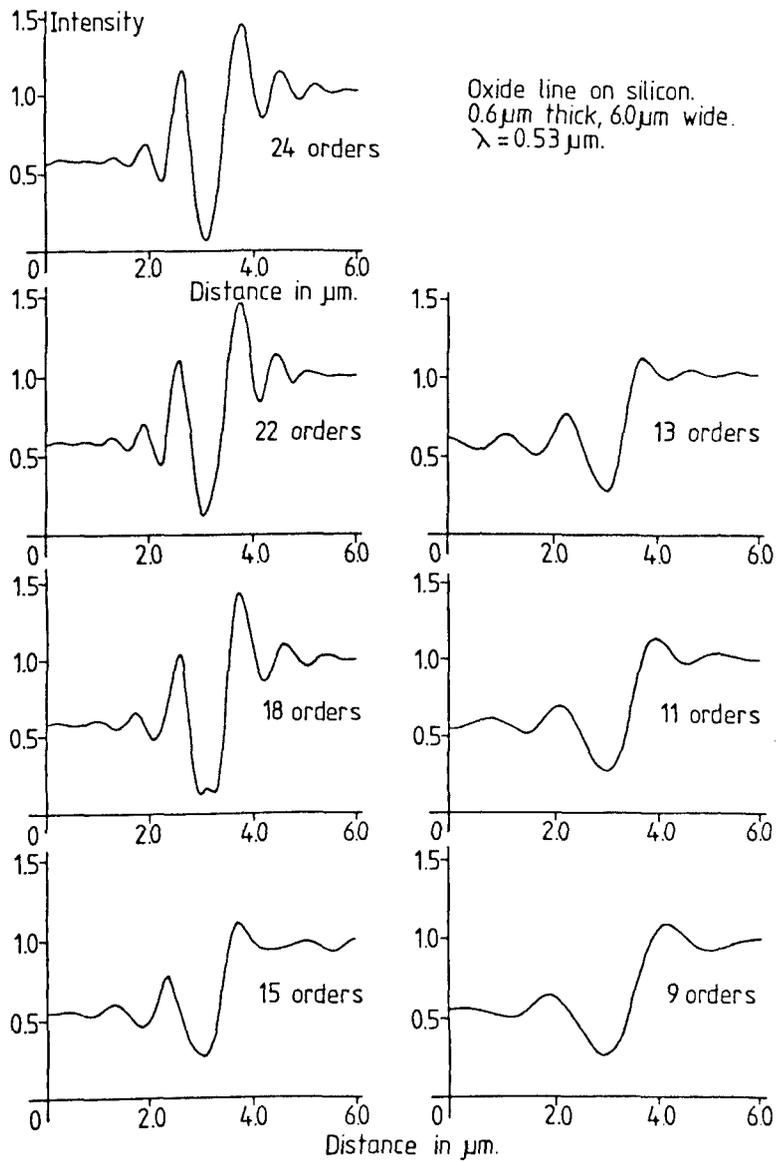
Figure 6.4.5. The effect on the image profile of reducing the number of modes used to model the field profile.

CHAPTER SEVEN

Theoretical and Experimental Studies into Thick Layer Imaging.

In this section the thick layer waveguide model will be used to investigate the image profiles which are produced by different objects and with different optical conditions. This will enable features in the image profiles to be related to the physical structure of the object and will enable accurate linewidth measurements to be made.

Usually it is not necessary to be able to make accurate linewidth measurements provided that it is possible to be able to make measurements which are repeatable between specimens and are not sensitive to parameters such as the layer thickness, substrate thickness, refractive index or edge shape. The model can be used to locate features which are least sensitive to these parameters and allows the performance of different measurement techniques to be assessed.

7.1 The effect of optical parameters on thick layer images.

The image profile is a function of wavelength and refractive index. The sensitivity to wavelength determines the illumination bandwidth which can be assumed to be monochromatic and the sensitivity to refractive index determines how accurately the refractive indices of the different materials need be known.

7.1.1 Wavelength dependence.

On thin layer objects the field profile across the top of the object is a function of the ratio of thickness to wavelength. As the layer becomes thicker and begins to support guided modes, this is no longer true. The relative reflectivity of the patterned layer is still a function of the ratio of thickness to wavelength, but the fringes at the line edge are functions of both thickness and wavelength.

Figure 7.1.1 shows how the image profile of a line edge varies with wavelength for a thin layer. The curves show a gradual contrast change with wavelength and also, the edge slope and edge oscillations which are diffraction effects, change gradually with wavelength. With the longer wavelengths it can be seen that the edge oscillations die away rapidly, but for the shorter wavelengths they are sustained for

quite a considerable distance. The reason for this is that as the illumination wavelength decreases, the object begins to support guided modes and it is these which are sustained a long way from the line edge, rather than diffraction effects. These curves cover the transition from a thin to a thick layer object.
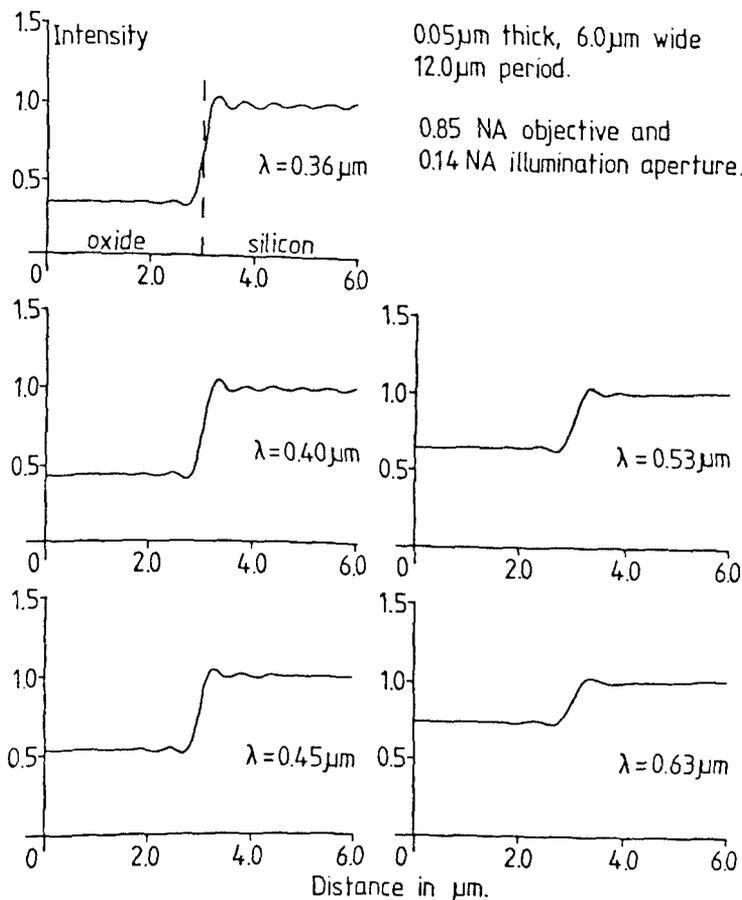
Figure 7.1.1. The effect of illumination wavelength on the image intensity profile of a thin layer of oxide on silicon.

Figure 7.1.2 shows a set of image profiles of a thick line edge for the same range of illumination wavelengths. The edge profile is now a sensitive function of wavelength. Since the thickness has increased, the contrast is more sensitive to changes in wavelength. In all the profiles, the edge is characterised by a dark fringe with a bright fringe on either side. The oscillations away from the edge are very sensitive to wavelength, as is any disturbance at the bottom of the dark fringe.

Taking λ = 0.53µm as the working wavelength, it can be seen that a ±5% wavelength variation has a considerable effect on the image profile and that the illumination bandwidth needs to be restricted to about 20nm in order to make the assumption of monochromatic illumination.
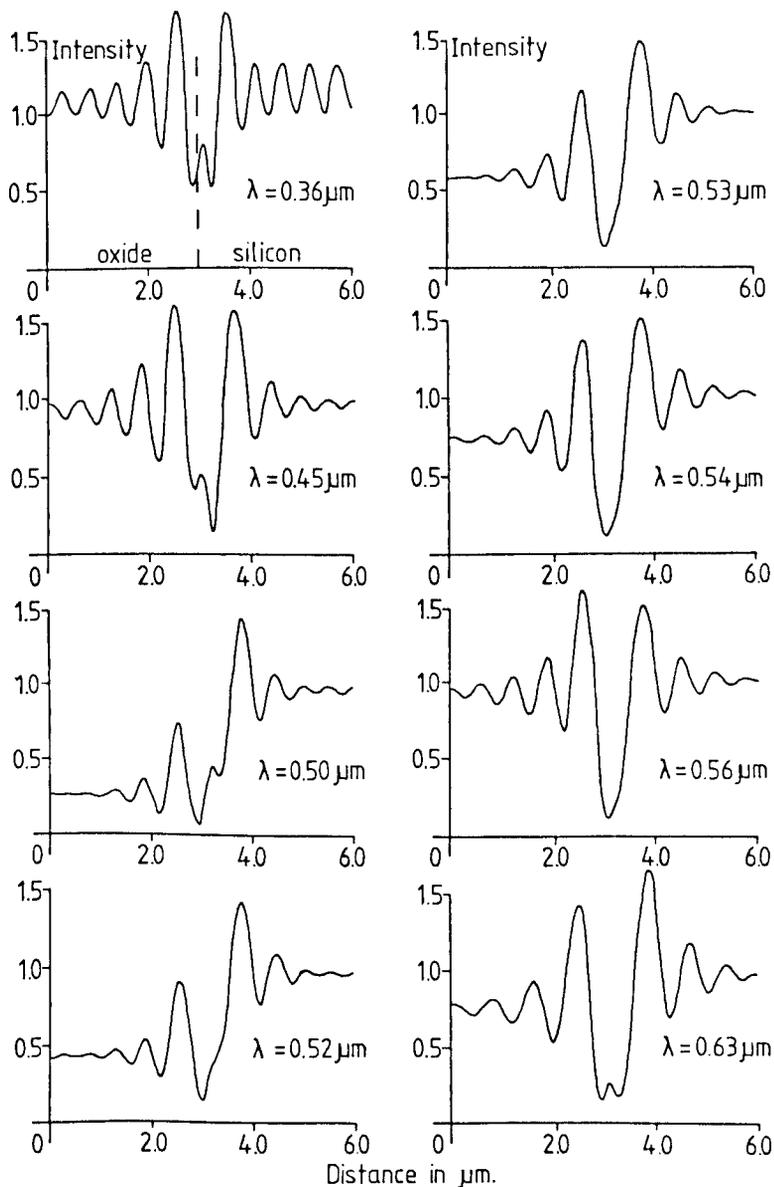
Figure 7.1.2. The effect of illumination wavelength on the image intensity profile of a thick layer of oxide on silicon. Conditions as in fig. 7.1.1 but for a 0.6um thick layer.

7.1.2 The effect of refractive index on the image profile.

The relative reflectivity of the patterned layer is a function of the refractive indices of the material, substrate and the surrounding air. The refractive index of air is extremely stable and may be assumed to be 1.00+0.00i. The crystalline silicon substrate will have a much higher refractive index of about 4.6+0.06i. This will not change much between wafers. The refractive index of the patterned layer is much harder to determine. Some materials such as silicon dioxide have stable and reproducible refractive indices, but others such as polycrystalline silicon exhibit considerable variations in both the

real and imaginary components of the refractive index. Photoresist can be a particular problem as its refractive index changes with time and exposure and also it will be inhomogeneous. The sensitivity of the image profile to refractive index determines how accurately the refractive index need be known when modelling the image.
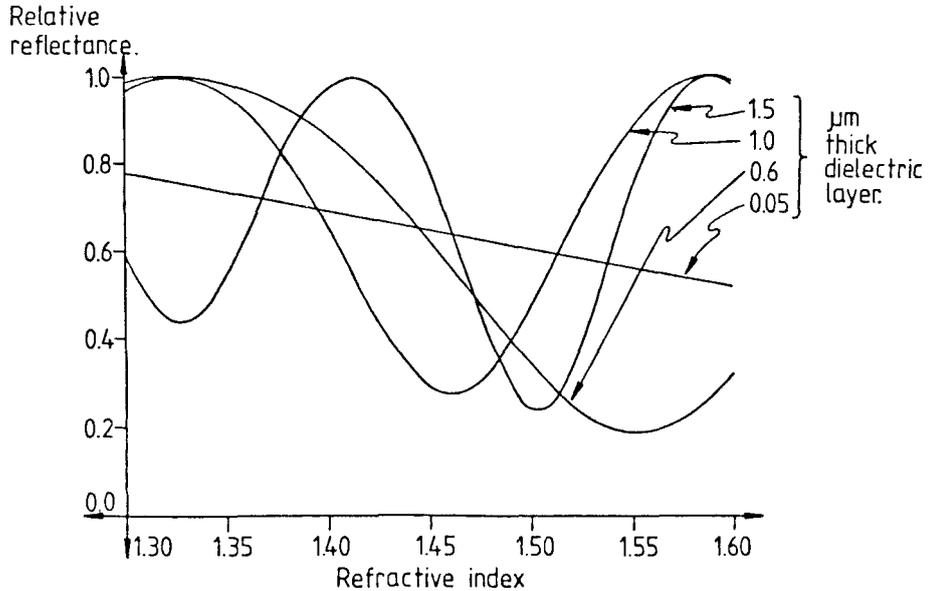

Figure 7.1.3. Relative reflectance as a function of refractive index for different patterned layer thicknesses. λ = 0.53μm.

Using the Fresnel equations it is possible to compute the relative reflectance of a patterned layer for a given angle of incidence and in this case it is assumed that the illumination is normally incident. Figure 7.1.3 shows how the relative reflectances of layers of different thickness on a silicon substrate vary as a function of refractive index. The curves show that as the thickness increases, the sensitivity to refractive index increases. The accuracy with which the refractive index need be known thus increases with thickness. The relative reflectivity is a quasi-periodic function of refractive index and where the curve is steepest the refractive index needs to be determined more accurately.

7.2 The effect of line geometry on the image profile.

The main reason for developing this model was to enable the image profiles of thick lines with shaped edges to be computed. Simple line structures could be adequately defined by the thickness and width of the patterned line object. Thicker layers also require that the edge shape be defined and for this it is convenient to use the polynomial

definition described in section 6.3.1.

7.2.1 Line thickness.

The relative reflectance is a cyclic function of the ratio of the thickness to the wavelength. However the image profiles do not repeat in the same cyclic manner. As the line becomes thicker it is able to support higher order modes and so the shape of the fringes at the line edges changes considerably. The sensitivity of the image profile to the line thickness determines how accurately the thickness need be known in order to model the profile successfully.
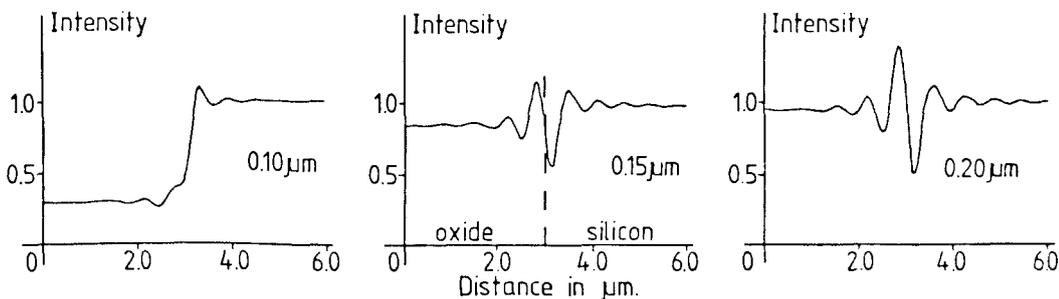
Figure 7.2.1. The effect of thickness on the image profile of a nominally 0.15µm thick line of oxide on silicon.

Figure 7.2.1 shows how the image profile of a nominally 0.15µm thick line changes if the thickness is allowed to vary by ±0.05µm. Figure 7.2.2 shows how the profile of a nominally 0.6µm thick line changes under the same conditions.
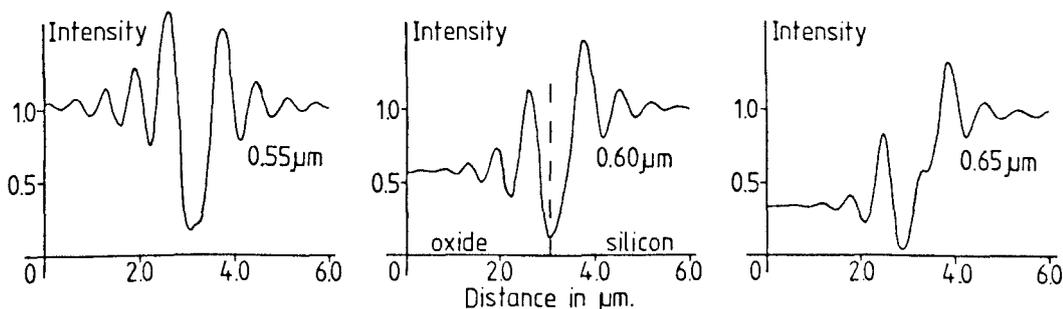
Figure 7.2.2. The effect of thickness on the image profile of a nominally 0.6µm thick line of oxide on silicon.

In both cases it can be seen that this small variation in line thickness causes a considerable change in the image profile.

7.2.2 The effect of edge shape on the image profile.

Using the polynomial definition of the line object, the effects of different geometries were examined. The zero order in the polynomial represents linewidth and the higher orders define the edge shape. Geometries of up to third order were investigated and these are described in figure 6.3.1.

a) Zero order geometries (linewidth).

The image profile of a line edge is much more sensitive to linewidth when the line is thick rather than thin. The width of an edge image for a thin line is determined by diffraction and so for lines which are at least several times wider than the resolution limit, the edge image profile is independent of linewidth. On thick lines this is no longer true as the object supports guided modes and these can propagate a considerable distance away from the edge. This means that even on wide lines, there is considerable interference between the edges.

Figure 7.2.3 shows how the edge image profile of a $0.6\mu m$ thick line changes with linewidth. The image profile changes considerably up to about $4.0\mu m$ and then it maintains roughly the same shape for lines wider than this. Despite the considerable change in profile shape, the edge is characterised by a dark fringe for all the linewidths shown. For lines narrower than $1.0\mu m$ the dark fringe begins to disappear.

b) 1st, 2nd and 3rd order edge shapes.

The non-zero orders in the polynomial expansion for the line geometry define the shape of the edges. First order slopes are rarely encountered but second and third order shapes are quite common. Figure 7.2.4 shows how the edge image profile changes as the line edge becomes less sharp for shapes of up to third order. The object is a nominally $6.0\mu m$ wide, $0.6\mu m$ thick line of oxide on silicon. The X_i parameters have been chosen to give roughly similar edge widths in the three cases.

The effect of broadening the edge is to broaden the dark fringe associated with the line edge and to increase the intensities of the bright fringes on either side. This has the effect of reducing the
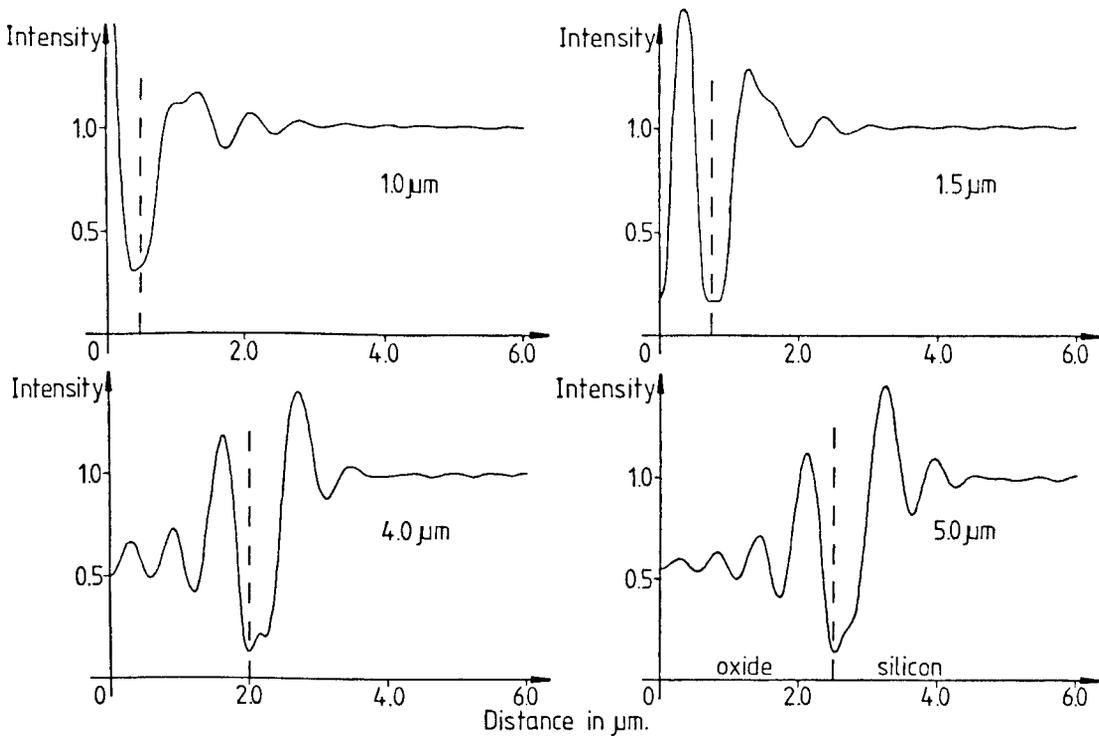
Figure 7.2.3. The effect of linewidth on the image intensity profile of a 0.6µm thick line of oxide on silicon.

field oscillations away from the line edge. One possible explanation for this, could be that the shaped edge acts as a poor waveguide (since it does not have a constant thickness) and so less energy propagates away from the line edge.

Objects such as these which have non-vertical edges raise a problem when defining the linewidth. The width of a line with sloping edges is not obvious and popular definitions of the width have included, the width at the top, the bottom or even the mean width. In most lithography processes, it is the width of the line where it meets the substrate which is of interest but the image formed is a function of the entire line object and so the effects of the edge shape cannot be avoided.

The curves in figure 7.2.4 show an interesting effect. The dark fringe in the profiles for the first and third order geometries remains roughly stationary as the edge broadens, however for the second order profiles it moves slightly to the left. In the cases of the first and third order edges, the physical edge pivots about a point half way down the line object and the mean linewidth remains constant. However according to the definition of the second order line edge, the effect of broadening the edge is to reduce the mean
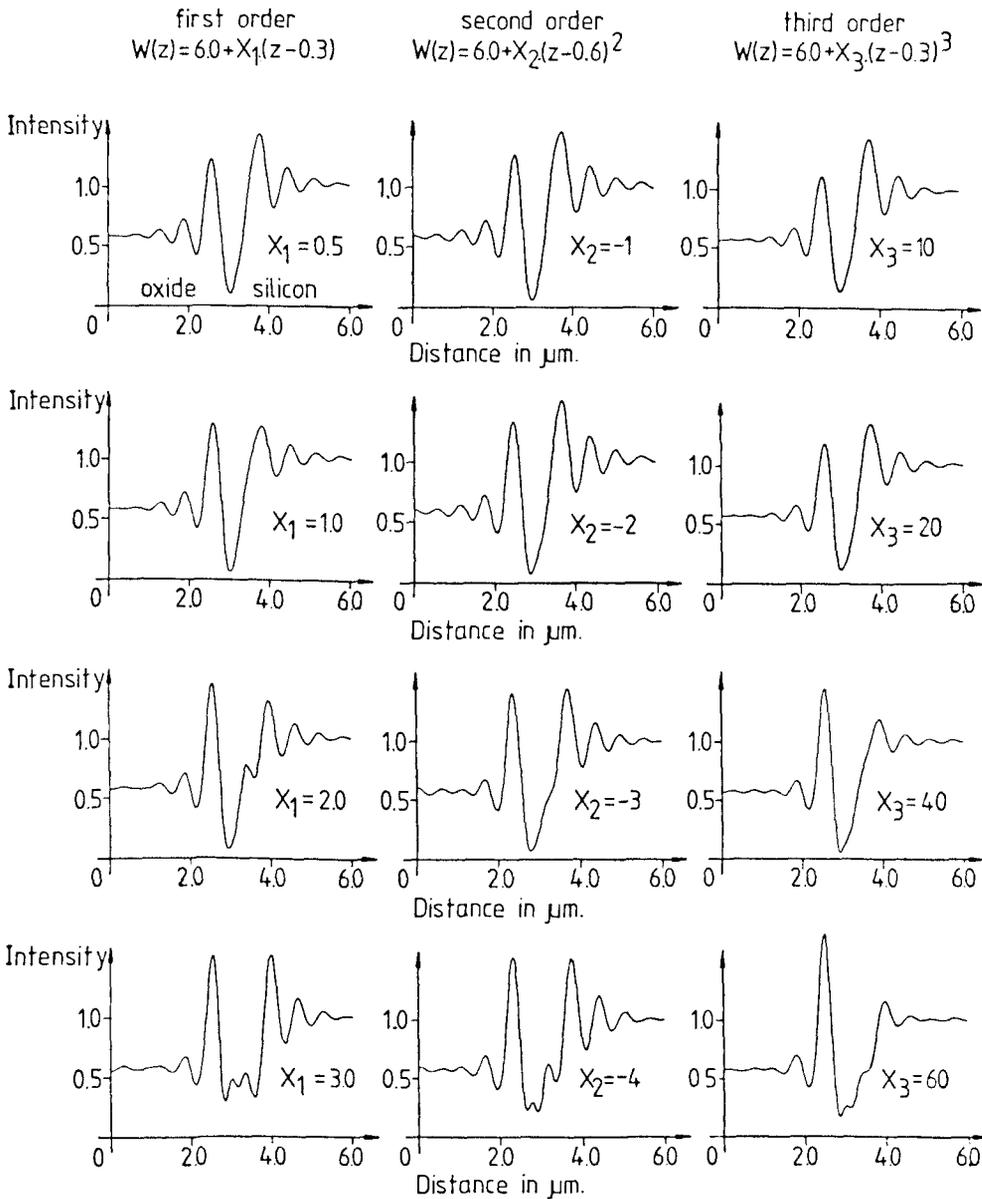
Figure 7.2.4. The effect of edge geometry on the theoretical image
profile of an oxide line on a silicon substrate.

linewidth. Thus it would appear that the dark fringe follows the
location of the mean linewidth rather than the width at either the top
or the bottom of the line.

7.2.3 Modelling granular structures.

So far it has been assumed that the refractive index of the material
is homogeneous. This will generally be true for materials such as
oxide and nitride but it is an unreasonable assumption for polysilicon
and photoresist.

Photoresist is an organic material with large molecules which readily break up when exposed to light and so the refractive index of the material will vary depending on the degree of exposure. In order to make the developed resist more durable for the etching process, it is often baked and this hardens the surface of the resist material. The result of this hardening process is to change the refractive index of the material unevenly throughout the line object.

Polysilicon consists of many small crystals and the boundaries between them produce irregularities in the refractive index profile which perturb the optical image profile. These boundaries can be modelled by adding delta functions to the refractive index profile.

$$\varepsilon'_n(x) = \varepsilon_n(x) + \sum_{j=1}^{J} C_j \delta(x-X_j) \qquad (7.2.1)$$

where,

$\varepsilon_n(x)$ = refractive index profile of layer n.

C_j = amplitude of jth refractive index disturbance, and

X_j = position of the jth disturbance.

The function $\varepsilon'_n(x)$ now describes a noisy layer and by restricting the range of values for X_j, the noise may be confined to different parts of the layer. This enables the effects of granular structures to be investigated.
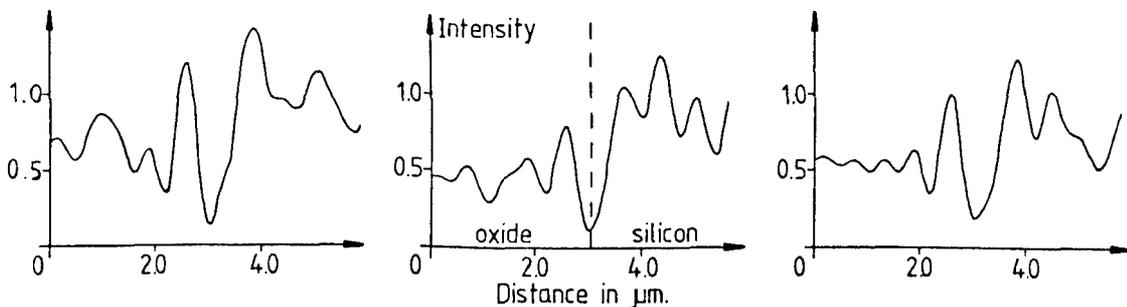

Figure 7.2.5. Theoretical image intensity profiles of a line patterned in a noisy dielectric layer on silicon. A different set of noise data has been used to generate each profile.

Figure 7.2.5 shows three image profiles for a noisy structure. The noise has been added to a 6.0μm wide, 0.6μm thick layer of oxide on silicon and has been introduced into the entire layer. In each of the three cases shown in figure 7.2.5, different sets of noise data have been used corresponding to different sets of C_j and X_j selected from a

random number table. The most striking feature of these curves is
that although the field is perturbed within the area of the object,
the dark fringe at the edge of the line remains well defined and
remains at the line edge.

7.3 Comparison with experimental results.

In order to test the model against practical image profiles a test
specimen was prepared which consisted of lines cut out of photoresist
on a silicon substrate. The same pattern was put down repeatedly but
through a range of focus positions. This produced a series of
patterns with a range of edge slopes. Two patterns were picked out
for their edge properties. One pattern corresponded to the in-focus
exposure and had near vertical edge walls, the second corresponded to
a considerable defocus on exposure and had significantly sloping
edges.

The in-focus exposure was assumed to produce vertical edge walls and
when the structures were examined in a scanning electron microscope
this was found to be very nearly the case. The edge walls had slopes
of about 80° which according to the theory may be effectively assumed
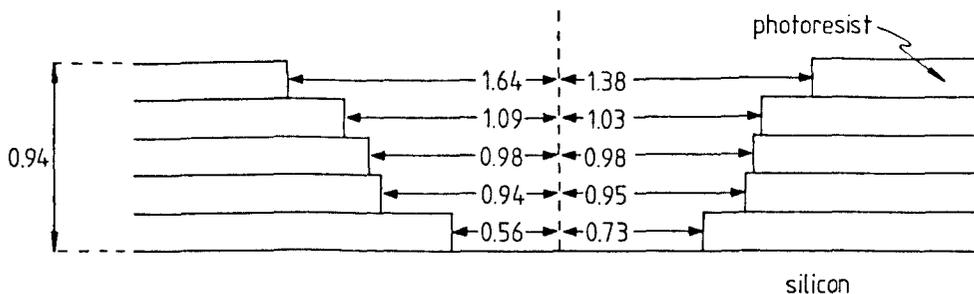to be vertical. The gap in the resist was taken as 2.5μm.

Figure 7.3.1. Physical profile of the asymmetric object window in
photo-resist used to compute the image profiles in
figure 7.3.2b. All dimensions are in micrometres.

The out of focus print had significant edge slopes and the structure
was modelled by the asymmetric third order gap shown in figure 7.3.1.
The photoresist thickness was not known exactly but was nominally
1.0μm. The refractive index of the photoresist was taken to be
1.5+0.02i. The theoretical layer thickness was adjusted until it
produced the same contrast as the practical data, this gave a
thickness for the resist of 0.94μm. The refractive index of the

resist is not known exactly but this is not serious as small changes in the refractive index can be offset by making small adjustments to the layer thickness without making large changes to the image profile.

It is very difficult to determine the location of the focal plane when looking at thick structures such as these and so it is necessary to generate a series of through focus profiles and compare these with the theoretical data. Figure 7.3.2 shows a comparison of the practical and theoretical image profiles produced by these structures. The practical data was generated using a 0.85 NA objective, 0.14 NA illumination aperture and 0.53μm wavelength illumination from a laser. The optical profiling was carried out at the National Bureau of Standards (USA) and the system has been described in the literature (Nyyssonen (1982d and 1984) and Kirk and Nyyssonen (1985))[†].

The profiles for the sloping edge structure are quite different from those for the vertical edge structure but in both cases they show good agreement with the models. The theoretical profiles display the same features as the practical curves but the agreement is not perfect. One significant source of error is that it has been assumed that the resist is homogeneous. This is not strictly true and it would have been better to use a material such as silicon dioxide which is homogeneous and would produce a stable and accurately known refractive index profile.

A second source of error may arise as it is assumed that the illumination is normally incident. The central peak of the sloping edge profile moves as the object moves through focus and also the vertical edge slope object generates slightly asymmetric profiles. The NBS profiling system used to generate these profiles is checked for illumination symmetry when profiling thin layers. However it may be the case that these thick layers are significantly more sensitive to asymmetric illumination than is detectable from the image of a thin layer object.
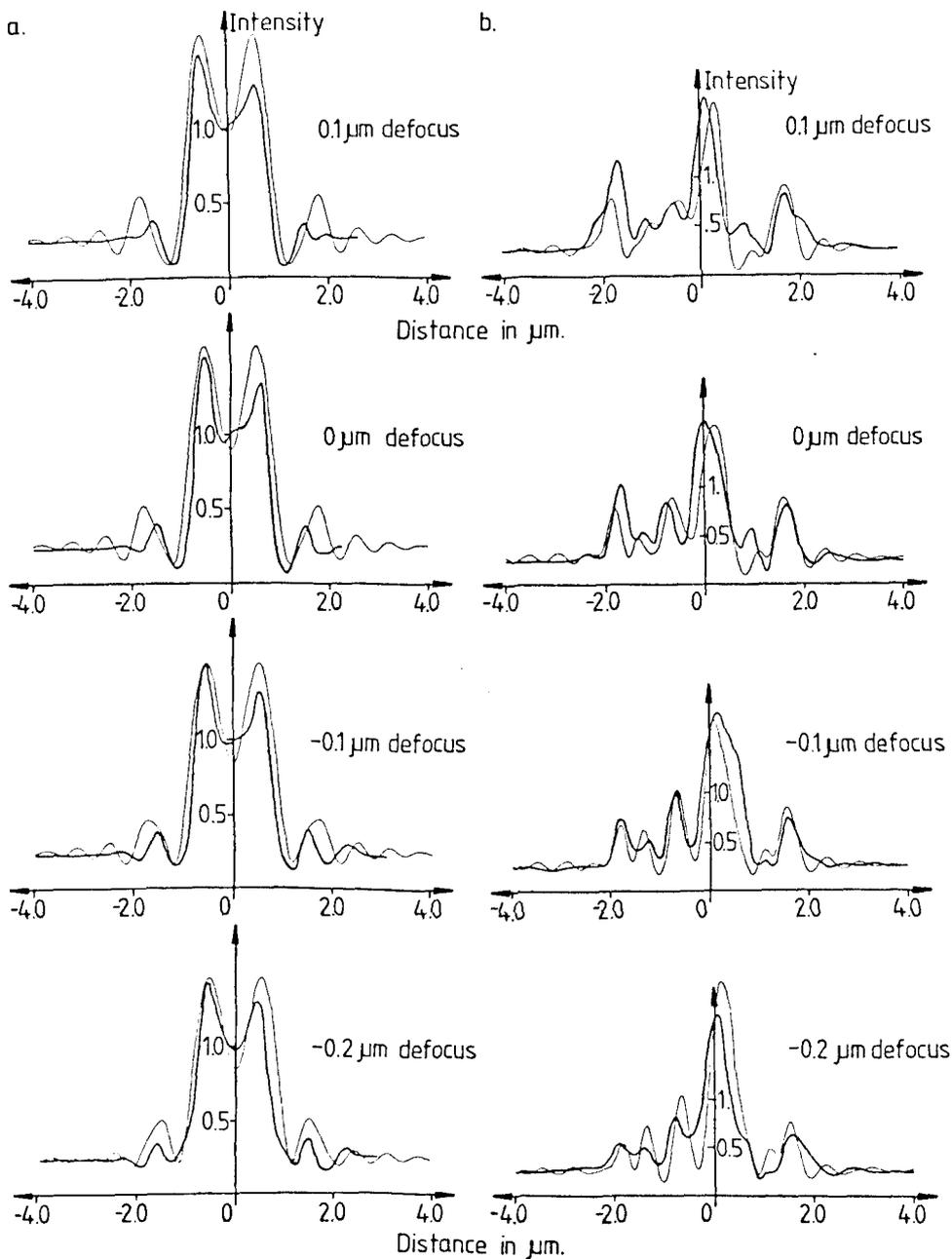
Figure 7.3.2. Comparison of experimental and theoretical image profiles of a window in photo-resist on silicon. The window has either vertical edge walls (a) or sloping edge walls (b). The experimental profiles are shown by the thick lines and the theoretical profiles by the thin lines.

7.4 A study of the stages involved in making an MOS device.

In this section, some typical structures which are encountered in microlithography are considered and how small variations in these structures will affect the optical image profile. The two structures shown in figure 7.4.1 were used as the basis for this study. These
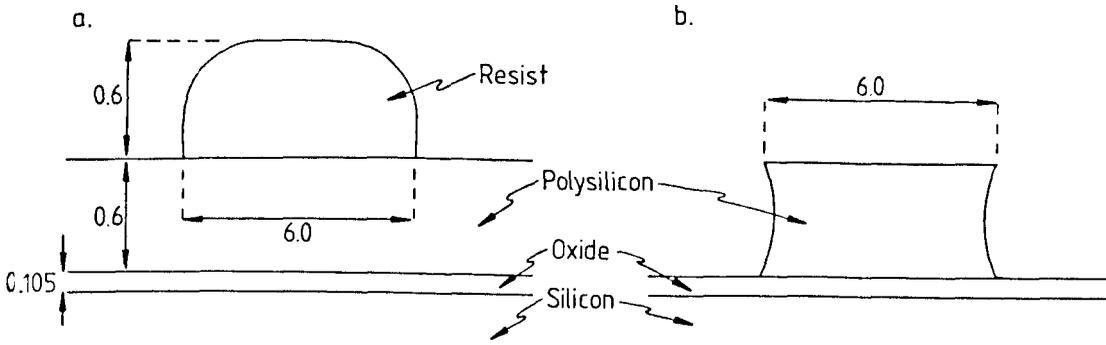
Figure 7.4.1. Cross sections of the shapes used to model the polysilicon patterning stage of making an MOS device. The two shapes are patterned resist in unetched polysilicon (a) and etched polysilicon with the resist stripped (b). All dimensions in μm.

structures represent the key stages of etching the polysilicon layer when making an MOS device. The structure in figure 7.4.1a represents the stage between patterning the resist and etching the polysilicon. Controlling the resist dimensions at this stage will help with controlling the final etched polysilicon linewidth. However measuring the width of the resist line at this stage is difficult as the image profile is a function of the geometry, refractive index and thickness of each of the three layers. The polysilicon was assumed to have a refractive index of 3.8+0.1i.
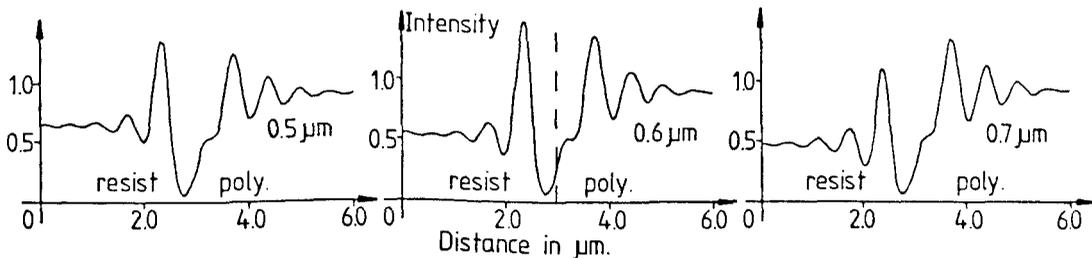

Figure 7.4.2. Theoretical image profiles of a resist line on polysilicon (structure in figure 7.4.1a) for polysilicon thicknesses of 0.5, 0.6 and 0.7μm.

The curves in figure 7.4.2 show how variations in the thickness of the polysilicon affect the image profile of the resist line. Only half the image profile is shown as the line is symmetrical. The two most striking features which emerge from these curves are that the dark and bright fringes associated with the line edge are very sensitive to the thickness of the polysilicon layer.
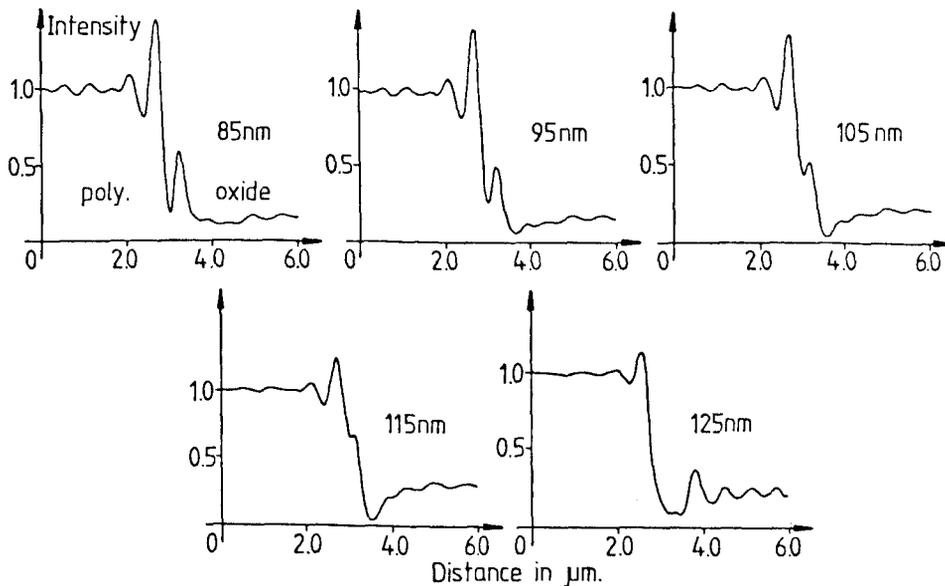
Figure 7.4.3. Theoretical image profiles of a polysilicon line (structure in figure 7.4.1b) for a range of oxide thicknesses from 85nm to 125nm.

The feature in figure 7.4.1b represents a polysilicon line after etching and stripping off the resist. The image profile will be a function of the thickness of the layers and figure 7.4.3 shows how the thickness of the oxide sub-layer affects the image profile of the structure shown in figure 7.4.1b. The polysilicon is assumed to have vertical edge walls and the oxide thickness is varying over a range of ±20nm. As the thickness changes, the edge profile changes considerably. This makes linewidth measurement difficult as there is no feature or threshold which locates the edge independently of oxide thickness.

The image profile is also a function of edge geometry and in figure 7.4.1b it has been assumed that the edge shape can be defined by a second order polynomial centered half way down the polysilicon layer.

$$W(z) = X_2 \cdot (z-0.3)^2 \qquad (7.4.1)$$

This shape is frequently encountered where the etching has been allowed to continue under the resist. The curves in figure 7.4.4 show how variations in the edge curvature of the polysilicon layer affect the image profile. The edge profile is clearly sensitive to the curvature of the physical edge of the polysilicon. It is difficult to define the location of the physical edge of a curved feature but three
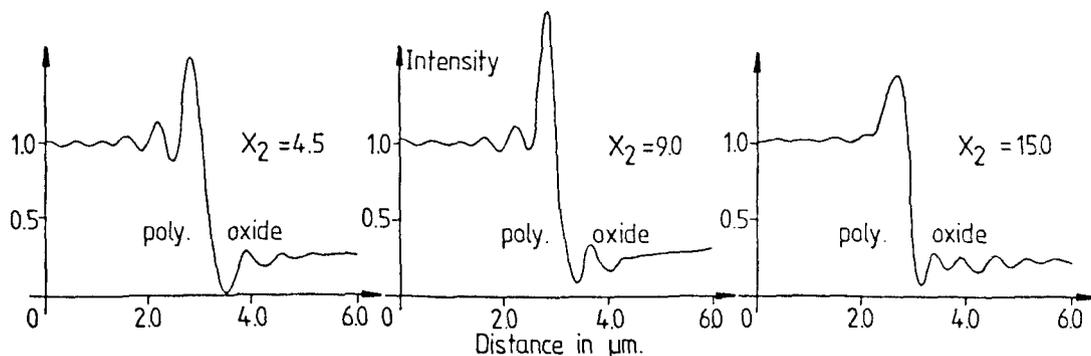
Figure 7.4.4. Theoretical image profiles of a polysilicon line
(structure in figure 7.4.1b) for a range of edge
curvatures defined by equation 7.4.1.

possibilities are the minimum, maximum or mean width of the layer.
For the examples shown in figure 7.4.4 these definitions all give
linewidths which are narrower than if the bottom of the dark fringe
were taken as the location of the line edge. More importantly
however, the offset between the true and measured linewidth varies
over a range of a few tenths of a micron regardless of the definition
used for the line edge.

7.5 Linewidth measurement on thick layer objects.

Linewidth measurement on thick layer objects is much more difficult
than on thin layers as the image profile is much more sensitive to the
object structure. Locating the "true" edge threshold is extremely
unreliable as the edge image profile can vary significantly for only a
small change in any one of the object parameters.

Many thick layer objects produce image profiles with dark fringes near
the line edges and these are often used as reference points when
making measurements. The bottom of this dark fringe is often used as
the line edge or alternatively, an approximation to this point is
taken by calculating the mean of two thresholds on either side of the
dip (Kirk et al. (1984) and Kirk and Gurnell (1985)). Figure 7.5.1
shows how these two measurement techniques compare when applied to a
2.0µm wide, 0.6µm thick line of oxide on silicon (λ=0.53µm, 0.85 NA
objective and 0.14 NA illuminating aperture). Trying to locate the
minimum directly is not a very satisfactory method of measuring the
linewidth. If the 0.6µm thick line varies by ±0.2µm in thickness then
the linewidth measurements may differ by as much as 0.5µm. This
change can occur very rapidly as can be seen by the behaviour of the

curve for thicknesses of around 0.6μm. If however the edge is located by taking the mean of two thresholds then the measurements will differ by less than 0.26μm.
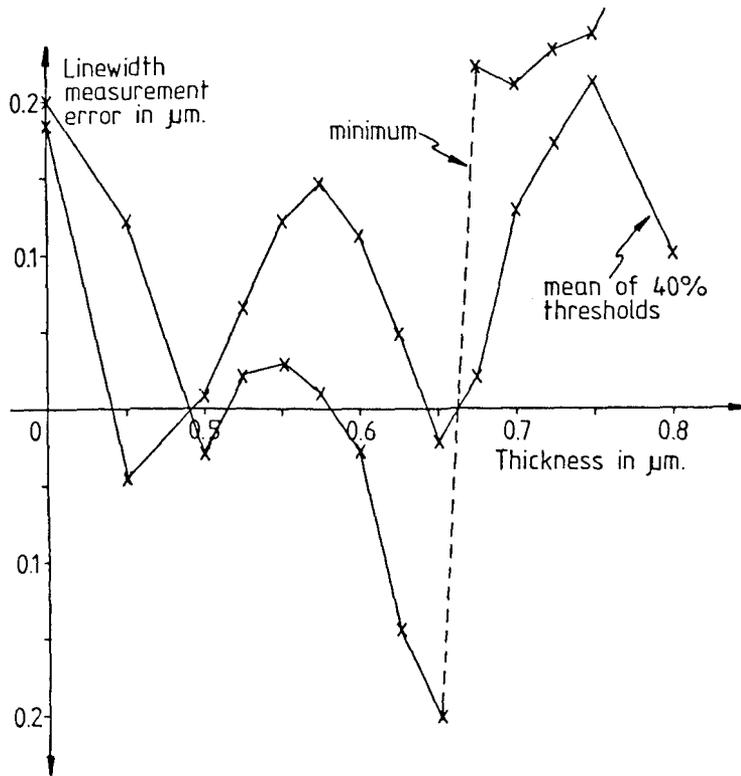

Figure 7.5.1. Linewidth measurement error as a function of thickness.

This still represents too much variation between samples. It can be seen that the measurement error resulting from the mean of two thresholds is quasi-periodic with thickness and it is also quasi-periodic with wavelength. With thin layers, this would be much closer to being periodic but even so, the cyclic variations will be significantly reduced if broad band rather than narrow band illumination is used.

As the thickness changes, the relative reflectivity changes and this can be readily detected when making linewidth measurements. The images used to generate figure 7.5.1 were used in figure 7.5.2 to plot the same measurement errors as a function of the relative reflectivity of the specimen. If a best straight line is fitted to this data then the linewidth error as a function of relative reflectivity (R) is given by,

$$\text{Measurement error in } \mu m \approx 0.21 \times (R - 0.25) \qquad (7.5.1)$$
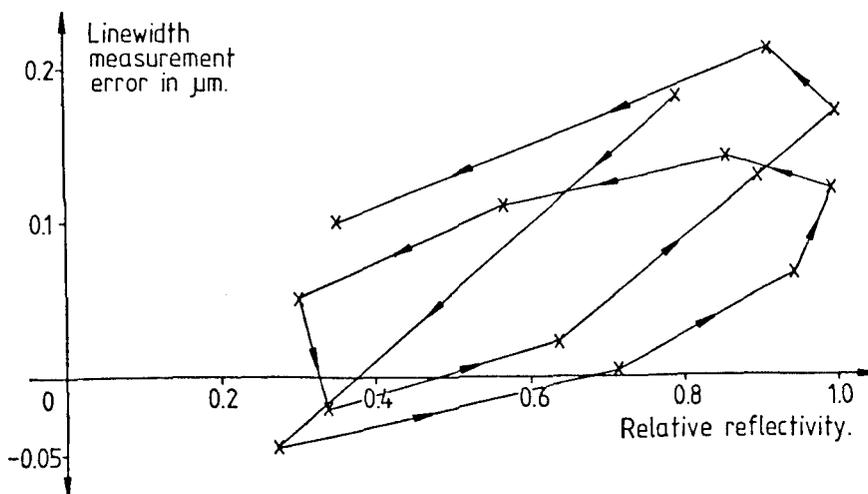
Figure 7.5.2.   Linewidth measurement error as a function of
                reflectivity as the line thickness increases.

After correcting the data with this function, the errors have the
spread shown in figure 7.5.3.   This corrected curve has been drawn to
the same scale as figure 7.5.1 in order to show how the correction for
contrast has reduced the spread of measurement errors.   After contrast
correction, the measurements only cover a range of 0.17µm which is an
improvement on the 0.26µm spread of the uncorrected values.

Figure 7.5.4 shows a plot of measurement errors for the same line as
in figure 7.5.1 but this time, it is the refractive index which has
been allowed to vary.   Again contrast correction can be applied to
reduce the spread of the measurements.

Contrast correction functions can be derived from theoretical models
or alternatively they can be derived empirically.   If a range of
specimens is measured optically and in a scanning electron microscope,
then the linewidth measurement error against contrast can be plotted.
A best straight line can be fitted to this data and used as a contrast
correction function.

Virtually all measurement systems capture the image profile and so
determining the contrast is relatively simple.   The measurement
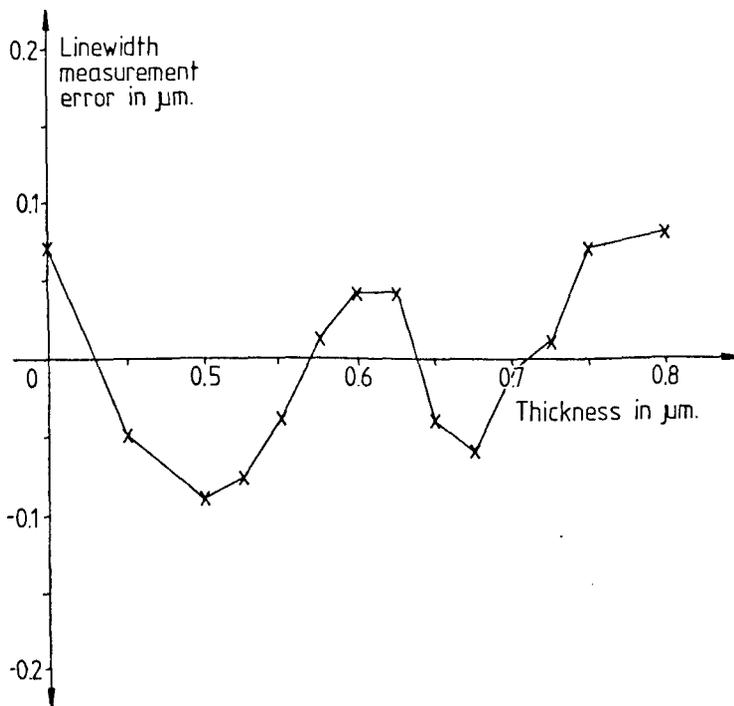correction can then be included in the measurement routine.

Figure 7.5.3. Linewidth measurement error as a function of
thickness for the data in figure 7.5.1 after
correcting with the function in equation 7.5.1.
The measurement uses the mean of two 40% levels.
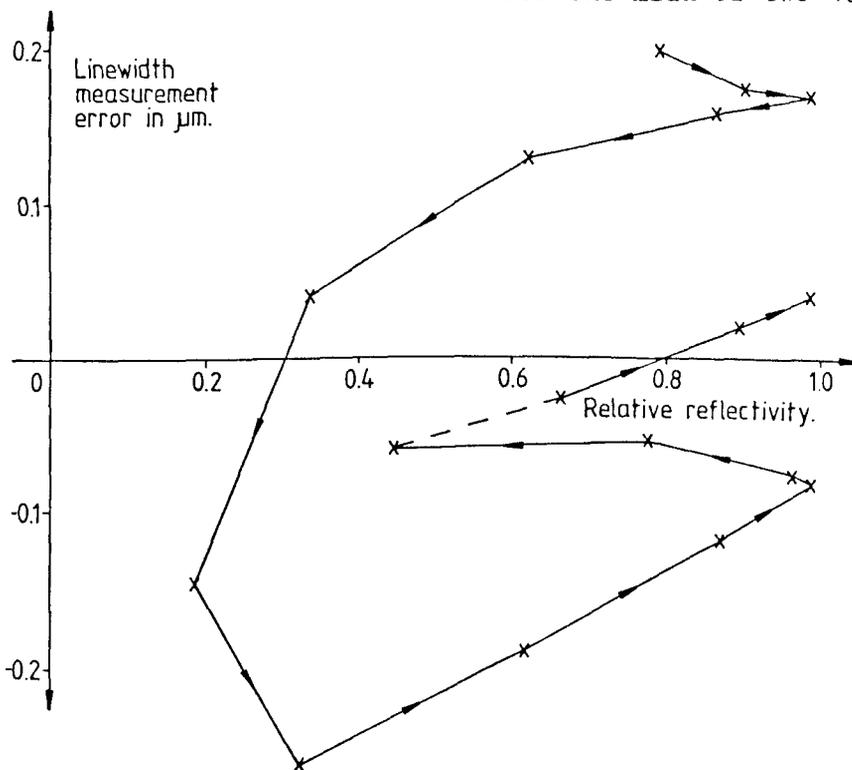

Figure 7.5.4. Linewidth measurement error as a function of
contrast for a 0.6µm thick dielectric line, as
the refractive index varies from 1.1 to 2.2.
The measurement uses the mean of the 40% thresholds
on either side of the dark fringe to define the
edge of the line. The dotted section indicates
where two 40% thresholds could not be located.

CHAPTER EIGHT

Conclusion

This work has divided into three main areas relevant to linewidth measurement; the imaging of thin layers (scalar diffraction theory), thick layers (waveguide theory) and automation techniques.

A model has been developed using scalar diffraction theory, to describe the imaging of thin layer objects. This model has been used to successfully predict the image profiles of line objects for modelling measurement techniques and has enabled the harmonic content of through focus image profiles to be investigated. This has shown that when objects with finite reflectivities are imaged with partially coherent illumination, the amplitudes of the image harmonics do not always go through a maximum when the object lies at the focal plane.

It has been shown that in order to obtain good agreement between practical and theoretical image profiles, it is necessary to restrict the illumination bandwidth when imaging layers with finite reflectivities. However, when attempting to perform linewidth measurements which are reproducible across specimens, it has been shown that using wideband illumination will help to improve reproducibility.

Models have been developed for typical imaging transducers and in the case of the video camera tube, good agreement with practical performance has been demonstrated. The models of optical imaging and transducer response have been used to analyse the performance of image shearing and video based measurement systems. Close agreement between predicted and measured errors using a chrome photomask calibration standard, has enabled systematic measurement errors to be identified in these techniques.

Practical studies of the imaging of chrome on glass photomasks have shown that there can be discrepancies between practical and theoretical image profiles. The imaging of metallic objects is not fully understood and this work has revealed polarisation effects. Further work is necessary to model these effects and to provide a better understanding of this imaging process, particularly at high numerical apertures.

The work has also shown that the performance of typical objective lenses falls short of that predicted by the theoretical diffraction limit. This introduces errors into the measurements and further work is necessary into modelling lens aberrations in order to determine design requirements.

A measurement system based on coincidence setting shear (CSS) and video threshold detection has been presented and it has been reported that this has enabled improved measurement reproducibilities to be obtained compared with the existing manual CSS system.

The design of two dimensional spatial filters has been analysed in order to determine a set of design rules and highly efficient corner detecting filters have been produced. Digital filters were also used to detect critical focus and to automate the edge to edge setting in the image shearing measurement technique.

The digital spatial filter models were combined with scalar diffraction theory and image transducer models to design and analyse a system for automatically measuring the lengths and widths of gaps in magnetic recording heads. The theoretical performance of the system was studied in detail and agreed with the measured performance of the practical system. The work has enabled an automated image shearing based measurement system to be developed which is faster than the manual version and has significantly improved measurement reproducibility. A novel algorithm was presented for locating clean measurement sites on dirty heads and its successful implementation was demonstrated.

Although the system can locate a clean site, focus and measure automatically, it still requires the operator to locate the feature to be measured and to move from one head to the next. Further work in this area could involve the development of algorithms for searching for the heads to be measured in order to fully automate the system.

Scalar diffraction theory has proved to have limited application and so a general waveguide theory for thick layer objects was developed. The novel theoretical model for the imaging of thick layer line objects was tested rigourously and no significant errors were detected. A comparison with practical thick layer images established that the model could successfully predict the optical image profiles

of thick layer line objects.

The model was used to investigate the imaging of thick, shaped line objects, with particular attention to the geometries and process variations encountered in the production of semiconductor integrated circuit wafers. A novel measurement technique was presented which uses the contrast of the image profile to provide a first order correction for measurement errors which are caused by process variations.

Although this work has enabled general thick layer image profiles to be modelled, it is restricted to line objects and narrow illumination angles. Further work is necessary to develop models for wide angle illumination and to include objects other than straight lines.

Bibliography

1. BLACKBURN D.W.J. Private communication, August 1983.

2. BORN M. and WOLF E. "Principles of optics" 6th edition, Pergamon Press, 1980.

3. BULLIS W.M. and NYYSSONEN D. "Optical linewidth measurements on photomasks and wafers" VLSI Electronics: Microstructure science, Vol.3, Academic, New York, 1982. pp.301-346.

4. BURCKHARDT C.B. "Diffraction of a plane wave at a sinusoidally stratified dielectric grating" J. Opt. Soc. Am., Vol.56, No.11, 1966. pp.1502-1509.

5. CHISHOLM J.J. "A new technique for linewidth measurement on 1µm geometry process wafers using fluorescence" Proc. Soc. Photo-Opt. Instrum. Engrs., 480, Integrated Circuit Metrology II, 1984. pp.49-56.

6. CROARKIN C. and VARNER R.N. "Measurement assurance for dimensional measurements on integrated-circuit photomasks" NBS Technical Note, 1164, 1982.

7. DOWNS M.J. and TURNER N.P. "Application of optical microscopy to dimensional measurements in microelectronics" Proc. Soc. Photo-Opt. Instrum. Engrs., 368, Microscopy - Techniques and Capabilities (SIRA), 1982. pp.82-87.

8. DUFFIEUX P.M. "The fourier transform and its application to optics" 2nd edition, Wiley-Interscience, 1983.

9. DYSON J. "Precise measurement by image-splitting" J. Opt. Soc. Am., Vol.50, No.8, 1960. pp.754-757.

10. ERTEZA A. "Sharpness index and its application to focus control" Appl. Opt., Vol.15, No.4, 1976. pp.877-881.

11. ERTEZA A. "Depth of convergence of a sharpness index autofocus system" Appl. Opt., Vol.16, No.8, 1977. pp.2273-2278.

12. FRANKEN A.A.J. "Resolution in camera tubes" Mullard Technical Publication, M80-0027, 1980.

13. GAGNAN R.J. "Image contrast theory of electronic focus" J. Opt. Soc. Am., Vol.68, No.10, 1978. pp.1309-1318.

14. GOODMAN J.W. "Introduction to fourier optics" McGraw-Hill, 1968.

15. GORGUI-NAGUIB R.N., HENEIN K.M. and KING R.A. "2-D filter design techniques" Wireless World, Vol.91, No.1592, 1985. pp.76-77.

16. GUILLAUME M.E., Noailly N.M., Reynaud J.C. and Buevoz J.L. "Fourier transform method for optical linewidth measurement" Proc. Soc. Photo-Opt. Instrum. Engrs., 480, Integrated Circuit Metrology II, 1984. pp.71-77.

17. HECHT E. and ZAJAC A. "Optics" Addison-Wesley, 1974.

18. HOPKINS H.H. and BARHAM P.M. "The influence of the condenser on microscopic resolution" Proc. Phys. Soc., Vol.63, No.10, 1950. pp.737-744.

19. HOPKINS H.H. "On the diffraction theory of optical images" Proc. Roy. Soc., Vol.A217, 1953. pp.408-432.

20. ICHIOKA Y. and SUZUKI T. "Image of a periodic complex object in an optical system under partially coherent illumination" J. Opt. Soc. Am., Vol.66, No.9, 1976. pp.921-932.

21. ICHIOKA Y., YAMAMOTO K. and SUZUKI T. "Defocused image of a periodic complex object in an optical system under partially coherent illumination" J. Opt. Soc. Am., Vol.66, No.9, 1976. pp.932-938.

22. JERKE J.M. "Accurate linewidth measurements on integrated circuit photomasks" NBS Special Publication, 400-43. 1980.

23. JERKE J.M. and WENDELL C.E "Use of the National Bureau of Standards (NBS) antireflective (AR)-chromium optical linewidth standard for measurement on other types of chromium photomasks" Proc. Soc. Photo-Opt. Instrum. Engrs., 342, Integrated Circuit Metrology, 1982. pp.15-26.

24. JERKE J.M., CROARKIN M.C. and VARNER R.N. "Interlaboratory study on linewidth measurements for antireflective chromium photomasks" NBS Special Publication, 400-74, 1982.

25. KASPAR F.G., "Diffraction by thick, periodically stratified gratings with complex dielectric constant" J. Opt. Soc. Am., Vol.63, No.1, 1973. pp.37-45.

26. KERMISCH D. "Principle of equivalence between scanning and conventional optical imaging systems" J. Opt.Soc. Am., Vol.67, No.10, 1977. pp.1357-1360.

27. KINTNER E.C. "Method for the calculation of partially coherent imagery" Appl. Opt., Vol.17, No.17, 1978. pp.2747-2753.

28. KIRK C.P., "The application of electronic techniques to the automatic focusing of optical microscopes" Leeds University, Dept. of Elec. and Elec. Eng., Departmental Project Report, March 1982.

29. KIRK C.P., MOORE D.S. and NELSON J.C.C. "Modelling photomask linewidth measurent systems in order to optimise performance and accuracy" Microcircuit Engineering 83, Academic Press, London, 1983. pp.209-216.

30. KIRK C.P. and MOORE D.S. "Optical metrology II" British Patent Application, No. 8328562, filed 26th October 1983.

31. KIRK C.P. "Multislope integrating alignment detection (MIAD) - A technique for automating an image shearing based critical dimension measurement system" Leeds University, Dept. of Elec. and Elec. Eng., Internal Report, December 1983.

32. KIRK C.P. "Optical Metrology III" British Patent Application, No. 8404664, filed 22nd February 1984 (a).

33. KIRK C.P. "The Design of 2-Dimensional Spatial Filters for Pattern Recognition Systems" Leeds University, Dept. of Elec. and Elec. Eng., Internal Report, May 1984 (b).

34. KIRK C.P. and NYYSSONEN D. "Modeling the optical microscope images of thick layers for the purpose of linewidth measurement" Proc. Soc. Photo-Opt. Instrum. Engrs. 538 Optical Microlithography IV, 1985.

35. KIRK C.P. "Automated Optical Linewidth Measurement" British Patent Application, No. 8507843, filed 26th March 1985.

36. KIRK C.P., MOORE D.S. and NELSON J.C.C. "Analysis of linewidth measurement techniques for the purpose of automation" Opt. Eng., Vol.24, No.4, 1985. pp.650-654.

37. KIRK C.P. and GURNELL A.W. "Modelling optical linewidth measurement techniques in order to improve precision and accuracy", Proc. Soc. Photo-Opt. Instrum. Engrs. 565 Micron and Submicron Integrated Circuit Metrology, 1985.

38. KIRK C.P. and GALE R.W. "Development of an automated image shearing microscope for measuring the critical dimensions of magnetic recording heads", Proc. Soc. Photo-Opt. Instrum. Engrs. 557 Automatic Inspection and Measurement, 1985.

39. LeMASTER, R.J., "AIM - An Automated Inspection Microscope for Measuring Critical Dimensions", Proc. Soc. Photo-Opt. Instrum. Engrs. 480 Integrated Circuit Metrology II, 1984. pp.40-48.

40. MARTIN L.C. "The theory of the microscope" Blackie, London 1966.

41. MASON D.C. and GREEN D.K. "Automatic focusing of a computer controlled microscope" IEEE Trans. Bio. Eng., Vol.BME-22, No.4, 1975. pp.312-317.

42. MOORE D.S. and KIRK C.P. "Optical microscopy applied to metrology in the semiconductor and fibre optic industries" Proc. Elect.-Opt./Laser Int. '84 UK., 1984. pp.139-148.

43. NYYSSONEN D. "Linewidth measurement with an optical microscope: the effect of operating conditions on the image profile" Appl. Opt., Vol.16, 1977. pp.2223-2230.

44. NYYSSONEN D. "Optical linewidth measurement on wafers" Proc. Soc. Photo-Opt. Instrum. Engrs. 135, Developments in Semiconductor Microlithography III, 1978. pp.115-119.

45. NYYSSONEN D. "Spatial coherence: the key to accurate optical micrometrology" Proc. Soc. Photo-Opt. Instrum. Engrs. 194 Applications of Optical Coherence, 1979. pp.34-44.

46. NYYSSONEN D. "Linewidth measurement spotlight" Semiconductor International, March 1980. pp.39-56.

47. NYYSSONEN D. "Calibration of optical systems for linewidth measurement on wafers" Opt. Eng., Vol.21, No.5, 1982 (a). pp.882-887.

48. NYYSSONEN D. "Design of an optical linewidth standard reference material for wafers" Proc. Soc. Photo-Opt. Instrum. Engrs. 342 Integrated Circuit Metrology, 1982 (b). pp.27-34.

49. NYYSSONEN D. "Laser micrometrology for integrated circuits" Proc. Insp., Meas. and Control Symp., ICALEO '82, 1982 (c). pp.24-30.

50. NYYSSONEN D. "Theory of optical edge detection and imaging of thick layers" J. Opt. Soc. Am., Vol.72, No.10, 1982 (d). pp.1425-1436.

51. NYYSSONEN D. "Optical linewidth measurements on patterned metal layers" Proc. Soc. Photo-Opt. Instrum. Engrs. 480 Integrated Circuit Metrology II, 1984. pp.65-70.

52. RICHARDS B. and WOLF E. "Electromagnetic diffraction in optical systems II. Structure of the image field in an aplanatic system" Proc. Roy. Soc., Vol. A253, 1959. pp.358-379.

53. ROWE S.H. "Light distribution in the defocused image of a coherently illuminated edge" J. Opt. Soc. Am., Vol.59, No.6, 1969. pp.711-714.

54. SHEPPARD C.J.R "Imaging modes of scanning optical microscopy" Scanned Image Microscopy, E.A.Ash (editor), Academic, New York 1980.

55. SMITH F.H., "Focus indicator" British Patent Application, No. 8016517, filed 1980.

56. SMITH F.H., "Optical metrology" British Patent No. GB 2027203 (1983).

57. SMITH F.H. "Binocular image-shearing microscopes" Advances in Optical and Electron Microscopy, Vol.9, Academic Press, London and New York, 1984. pp. 223-242.

58. STARK H. "Applications of optical fourier transforms" Academic Press, 1982.

59. STEEL W.H. "Effects of small aberrations on the images of partially coherent objects" J. Opt. Soc. Am., Vol.47, No.5, 1957. pp.405-413.

60. SWYT D.A. and ROSBERRY F.W. "A comparison of some optical microscope measurements of photomask linewidths" Solid State Technology, Vol.20, August 1977.

61. WATRASIEWICZ B.M. "Theoretical calculations of images of straight edges in partially coherent illumination" Opt. Acta, Vol.12, No.2, 1965 (a). pp.391-400.

62. WATRASIEWICZ B.M. "Image formation in microscopy at high numerical apertures" Opt. Acta, Vol.12, No.2, 1965 (b). pp.167-176.

63. WHITTAKER E.T. and WATSON G.N., "A course of modern analysis" 4th edition, Cambridge University Press, New York, 1940.

64. YOUNG M. "Linewidth measurement by high pass filtering: a new look" Appl. Opt., Vol.22, No.13, 1983. pp.2022-2025.