Development of a scanning acousto-optic based differential intensity and phase system for optical metrology

by

K. Roland Appel

Thesis submitted for
The Degree of Doctor of Philosophy
of the
University of London

Department of Electronic and Electrical Engineering,
University College London

November 1989

Abstract

An optical technique is described which simultaneously and independently measures the difference in intensity and phase of light reflected from two close areas on a sample surface. Line scans or images of the object are then formed by scanning the light across the surface. The key element is a Bragg cell which is used to divide a well collimated beam of light into two probing beams amplitude modulated in phase quadrature, with a relative frequency shift.

A comprehensive review is given of current profilometry techniques, and in particular optical methods. Emphasis is given in developing a thorough description of the system. Analysis of the acousto-optic interaction is combined within a mathematical framework which gives a general description of how the differential phase and intensity responses depend on the form of the modulated electrical Bragg cell drive. The analysis is applied to determine the tolerances necessary in the construction of this type of system. The use of a scanning mirror in a differential phase system is discussed and analysis is presented to model the effects of lens aberrations.

Three implementations of the technique have been built and experimental results are shown from prepared specimens which have,

i) predominantly phase contrast – a silicon wafer which has had a series of parallel tracks etched, ranging in depth from 180 to 700Å;

ii) predominantly intensity contrast – a silicon wafer half implanted with As^+ ions;

iii) a mixture of phase and intensity information – a deposited aluminium step on a silicon substrate.

Unlike any previous method, this system can measure true differential phase and intensity simultaneously. The real advantage of this is clearly demonstrated in the case of the third specimen where there is a change in surface conductivity across the sample. The two measurements may then help resolve uncertainties present in interpreting topography from phase data.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Dr. M. Somekh for the many hours of help and guidance that he has given me throughout this project. His enthusiasm and support has been a constant source of encouragement.

I also wish to thank Dr. C. W. See whose initial ideas were the driving force behind this work. His advice and discussions have been an invaluable stimulus underlying much of the research which I have carried out.

I also would like to acknowledge the Science and Engineering Research Council and GEC Hirst for their financial support.

Contents

List of figures

CHAPTER 1

INTRODUCTION

1. Introduction

A true non-destructive surface profilometer is an important measurement tool in any exacting manufacturing process. In the semiconductor industry dimensions of many components are less than a micron, and efforts are continually made to increase performance. This usually involves greater packing densities of components as well as more perfect growth of the constituent layers within devices. There is considerable interest in being able to measure the RMS roughness of surfaces, whether they be laser cavity mirrors or virgin wafers, with accuracies of the order of Ångstroms. There is tremendous art involved in manufacturing components or surfaces with such detailed precision, but this technology can never proceed further than the accuracy of available inspection tools.

The technique still in most common use for measuring surface profiles is the stylus probe, where the vertical movement of a stylus is monitored as it is dragged across the object surface. The measurements correspond exactly to surface topography, and depth resolutions have been reported[1] of up to 1 Å with lateral resolutions of a few tenths of a micron. Such vertical resolutions are not easy or always possible to achieve. As the probe is dragged over the surface it tends to oscillate. Consequently, to increase the depth resolution, a greater stylus weight must be used so as to suppress oscillations. For example, a stylus tip radius of 10μm used with a 1mg stylus load will give a 10nm depth resolution and 0.89μm lateral resolution. The surface pressure under the stylus in this case[2] is 160 kg/cm^2. This is a very high pressure and there is a severe risk of damage occuring to the sample. Even if the sample can withstand such high pressures, the time required for the measurement of a surface profile is limited by how fast the stylus can be dragged across the surface without again causing damage.

There is presently considerable interest in non-contact profilometers which will not

---

[1]J.M. Bennett and J.H. Dancy, Appl. Opt., "Stylus profiling instrument for measuring statistical properties of smooth optical surfaces", Vol. 20, no. 10, p.1785-1802, (1981).

[2]M.J. Downs, W.H. McGivern, and H.J. Ferguson, "Optical system for measuring the profiles of super-smooth surfaces", Precision Engineering, Vol. 7, No. 4, pp.211-215, October 1985.

damage the sample and also enable faster measurement speeds. Techniques with Ångstrom depth discrimination include various optical methods, scanning tunnelling microscopy and atomic force microscopy. The essence of any measurement system is that some type of probe is used to interact with the material under investigation, and the changes in the interaction of probe and material are recorded across the sample.

Exactly what is seen on the sample depends upon the interaction between probe and object. The stylus probe monitors the atom-atom contact of the probe and the surface, and is conceptually the simplest device, having an (electrical) output response proportional to the surface height variation. An optical system will measure the optical parameters of the surface atoms, ie photon-atom interaction; the scanning tunnelling microscope[3] (STM) will map out the work function of the surface atoms; the atomic force microscope (AFM) maps out the atom-atom interactive force over a distance range of about 10Å. The list continues, but the important point to see is that every type of probe will yield *complementary* information. Although it is usually assumed that each measurement type may be directly related to those of the stylus probe, this is not strictly true.

The STM uses effectively a single atom as a probe and under favourable condition, can spatially resolve individual atoms. By virtue of the charged nature of electrons, it may be used only for examining conducting surfaces. The AFM is better suited for examining insulators, and has been demonstrated[4] with a spatial resolution of ~5nm.

The prime attraction of the STM and AFM is the spatial resolution which can be attained. The spatial resolution of far field optical techniques is limited to about half the wavelength of light, which for visible light is about $^1/_4$ μm. This value can be improved

[3]P.K. Hansma, V.B. Elings, O. Marti and C.E. Bracker, "Scanning tunnelling microscopy and atomic force microscopy: Application to biology and technology", Science, Vol. 242, pp.209-216, 14 October 1988.

[4]S. Alexander, L.Hellemans, O.Marti, J.Schneir, V.Elings, P.K.Hansma, M. Longmire and J. Gurley, "An atomic-resolution atomic-force microscope implemented using an optical lever", J. Appl. Phys., Vol. 65, No.1, pp.164-167, 1 January 1989.

somewhat by working in the near field, and Fischer et al [5] have demonstrated a near field scanning microscope working in reflection, using an aperture of ~100nm, achieving a spatial resolution some five times above that of a far field technique. There are two major problems in working in the near field. First the improved lateral resolution is determined by the size of the aperture illuminating the object, and there is an inevitable trade off between resolution and signal to noise of the measured response. The second difficulty is in maintaining the probe at a constant distance, with nanometer accuracy above the sample surface. It is likely that scanning tunnelling microscopy technology will enable the spatial resolution of this technique to be considerably improved.

All three techniques require the probe/ sample distance to be maintained constant with nanometer precision, and this inevitably limits the maximum scan speed. Although considerable advances have been made in the scan technology particularly in the case of the STM, the scan speed is never likely to approach that of a far field optical technique where the probe/ sample separation distance is not nearly so critical. Although the spatial resolution for a visible light system is at best about 0.25 μm, differential phase interferometry has been shown[6] to be sensitive to changes in phase of 3×10^{-3} mrad (corresponding to topography of 1.5×10^{-3}Å), measured in a 1kHz bandwidth. Also unlike the STM, the probe particle is neutral, so enabling both conducting and insulating surfaces to be examined.

It is often assumed that an optical measurement, for example phase, can be directly converted into some unit of distance ie., surface height. This is only true if there is no variation in surface conductivity, which clearly cannot be assumed in general, and Raine and Franks have reported[7] that inhomogeneities in glasses can give rise to apparent

[5]U. Ch. Fischer, U.T. Dürig, and D.W. Pohl, "Near-field optical scanning microscopy in reflection", Appl. Phys. Lett., Vol. 52, No. 4, pp. 249-251, 25 January 1988.

[6]M. Vaez Iravani, and C.W. See, Proceedings of SPIE conference on 'Scanning microscopy technologies and applications', Los Angeles, CA, Editor C. Teague, Vol. 897, p. 43, (1988).

[7]K.W. Raine, and A. Franks, "Effects of variations in refractive index on the interferometric measurement of microtopography in glasses", Optica Acta, Vol. 32, No. 2, pp. 251-254, 1985.

height changes of 1nm or more. This is significantly more than the sensitivity of present interferometers and so considerable care is therefore needed in the physical interpretation of optical results.

There are several ways to approach this problem. It may perhaps be known from the history of a particular specimen that there is no compositional variation across the sample, in which case all optical contrast will be due to topography. Alternatively two measurements could be performed using different wavelengths of light. For example consider an optical system which measures phase. In parts of the sample where there is no compositional variation, the phase response will scale inversely with wavelength, and hence any compositional change will become apparent when studying the difference between the two measurements. This assumes that the phase angle of the complex reflection coefficient is always equal at the two wavelengths of light which again is not generally true.

A different approach is to monitor changes in the intensity reflection coefficient, use this to indicate the presence of compositional change, and so accordingly adjust the interpretation of phase into topography. It is this method which is the subject of the thesis.

The project has involved the development of a scanning optical system which can simultaneously measure the intensity and phase of light reflected from an object. A schematic diagram of the optical configuration is shown in figure 1-1. Incoming light is divided by a Bragg cell into two closely focussed beams on the object surface. The light reflected from each probed area is recombined by the Bragg cell and detected at the photodetector. The Bragg cell performs several necessary functions. It divides the light, introduces a frequency shift between each probe beam and also intensity modulates each beam in anti-phase. This enables two measurements, differential phase and differential intensity to be made, which are respectively the phase and intensity differences between the two reflected beams.

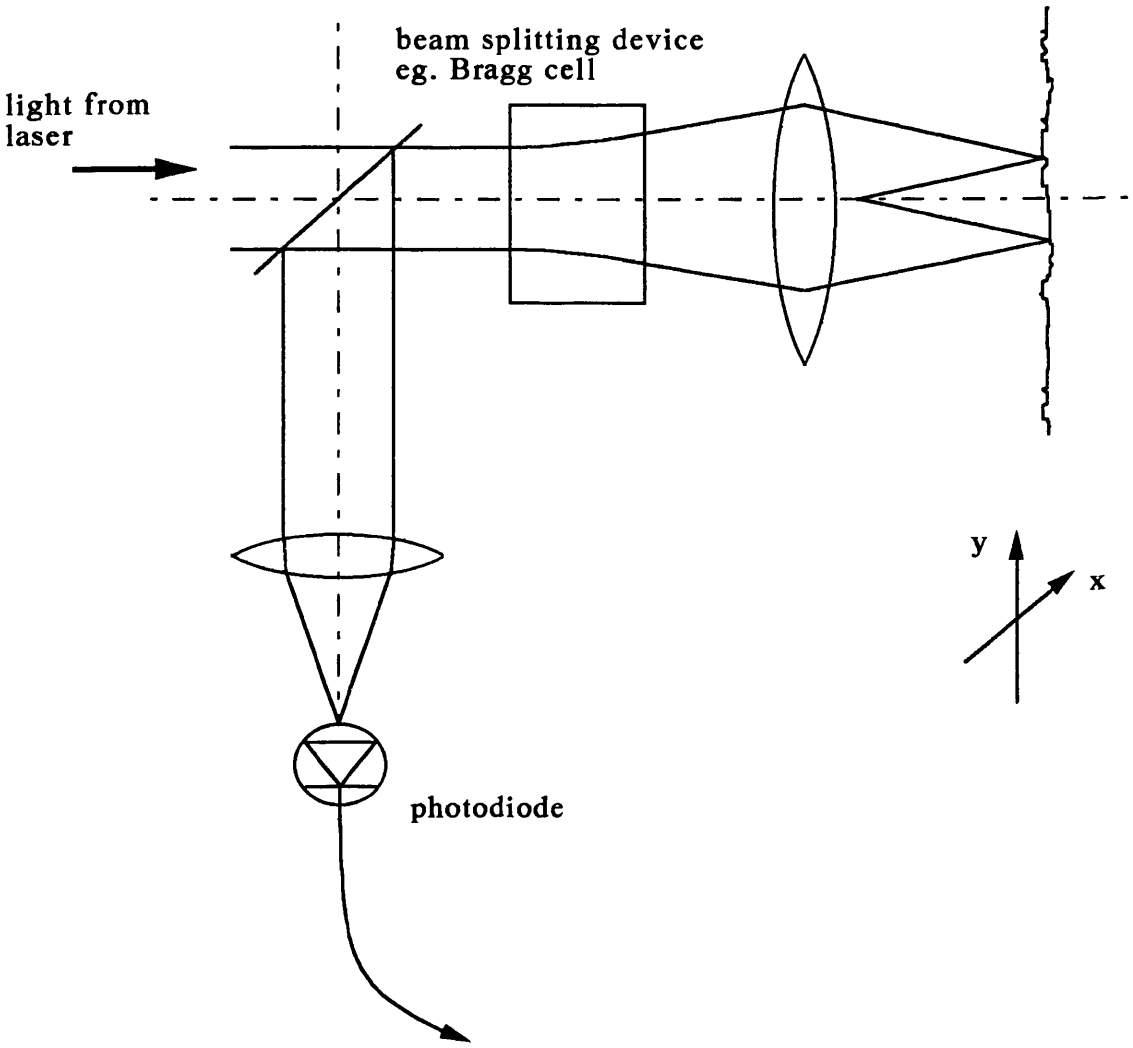


Figure 1-1 Schematic diagram of differential intensity and phase scanning instrument

Line scan measurements or images of the surface are made by scanning the beams in tandem across the surface. The system responds to objects which change the intensity or phase of reflected light and in this sense, the response is termed *differential*. Furthermore, as a common path system (all light follows approximately the same path), the method is very sensitive to weak phase or reflectivity objects.

Phase and reflectivity information alone however is still insufficient to unambiguously calculate topography from the optical measurements. It does nevertheless give twice the information of more conventional techniques, which measure either phase, reflectivity, or a mix of phase and reflectivity. The differential intensity information can be used to readjust the interpretation of the phase (to account for compositional changes), or at the very least, to warn the user of a possible problem in interpreting the phase into topography. Results are presented in chapter 6 where both types of information are used to unravel a complicated structure, in a way which would not be possible with phase information alone.

This system is being presented as an optical profilometer which has high phase sensitivity and the facility to discern small reflectivity variations. Potential applications include measurement of film thickness, reflectivity (refractive index) variation, and surface flatness of samples such as semiconductor wafers. For these applications, the system is operated in 'metrology' mode, with the two probing beams well separated on the sample surface. By changing the Bragg cell drive frequencies, the probing beams can be made to overlap for 'imaging' mode, and the technique may now for example, be used for line width metrology. The technique is ideal for examining low contrast samples and will be well suited in the study of biological tissues where at present stains frequently need to be employed to visualize weak phase features.

The emphasis of this project has been to build up a thorough description of the system. This has involved a combination of practical work in the building of various experimental

implementations, and also theoretical work where several aspects of the system have been analysed in detail.

The arrangement of this thesis is as follows. In the next chapter a general survey is given of various types of optical profilometers. A review is made of full field and scanning optical techniques, and a wide range of homodyne, heterodyne and differential methods are compared.

Chapter three presents the differential phase and intensity system. The underlying principles of it are described and analysis is presented which defines the necessary modulation to be used for driving the Bragg cell.

The Bragg cell is the key element in the system. In order to relate the electrical drive of the Bragg cell to the modulation of the diffracted light, it is necessary to model the acousto-optic interaction. Chapter four shows how the acousto-optic theory is applied to the present system.

Chapter five uses the analysis of the two previous chapters to ascertain the differential intensity and phase accuracies if various aspects of the system are below the ideal specifications. These include the effects of imperfect modulations, the light not at exact Bragg incidence, and the detection of Bragg cell generated intermodulation beams.

Experimental work is presented in the sixth chapter. Three implementation of the technique have been built and results are shown from prepared specimens which have,

i) predominantly phase contrast – a silicon wafer which has had a series of parallel tracks etched, ranging in depth from 180 to 700Å;

ii) predominantly intensity contrast – a silicon wafer half implanted with As^+ ions;

iii) a mixture of phase and intensity information – a deposited aluminium step on a silicon substrate.

The combination of differential phase and intensity results from the third specimen demonstrate the true advantage of collecting both types of information. The optical and electrical configuration of the latest experimental implementation is described.

Chapter seven considers the use of a mirror to scan the light over the object surface. Current differential systems all have stationary optics, and line scans or images are collected by rastering the object under stationary probing beams. The use of a scanning mirror has the potential of much faster frame rates for imaging.

In the last chapter an overall summary of the experimental and theoretical research is given, with suggestions for future work.

CHAPTER 2

REVIEW OF OPTICAL METROLOGY

2. Review of optical metrology

There are numerous optical techniques currently existing which may be applied to surface metrology applications. In this chapter we will look at several of these methods, comparing the performances and highlighting factors which are most likely to prove the limiting factor for measurement resolution. The optical systems are divided into two categories, full field and scanning systems, which are discussed in turn.

2.1 Full field systems

It would have been the natural progression to move from mechanical means of metrology to using optics. The earliest optical measurement systems were full field microscopes which responded best to features with changing reflectivity or transmissivity. These microscopes did not respond to phase objects, and since surface height variations affect solely the phase of the light, surface topography could not be measured by these systems. This is true in a low NA system; using higher NA objectives or introducing some defocus may give a small amount of contrast to such features. Unfortunately, there is no light detector including the human eye which can respond directly to the phase of light. So if light is to be used to measure topography, some means must be found to transform the optical phase into intensity variation, which may then be detected.

The first attempts involved modification of the optical field in the Fourier plane and include the central dark ground microscope, in which a small stop is used to block the zeroth order scattered light; the Schlieren method[1] in which all spectral components on one side of the central zeroth order are excluded (the zeroth order is either partially or totally blocked), and the phase contrast method of F. Zernike where a $\lambda/4$ phase difference is artificially introduced between the zeroth and side orders. Unlike the dark ground and Schlieren methods where quantitative interpretation of the images is difficult, the Zernike method can give a linear relation between observed intensity and phase of the

[1] introduced by A. Toepler in 1864 for examining defects in lenses

reflected light. This is true for small phase variations ($\phi << 2\pi$), and assumes that there is no variation in the object reflectivity across the field of view. None of these types of techniques can discern changes in the reflection coefficient or object height variation. The working principles of these and other established techniques can be found in many articles and books that deal with microscopy and diffraction phenomena[2,3,4,5]

Whereas the dark field and Schlieren methods achieve phase contrast by removing half the Fourier spatial frequency components, the Zernike method relies on interference between the zeroth and side orders. Other interferometric systems were soon to follow, in which instead of interfering light from within the same spatial spectrum, light from the object is superimposed on a second 'reference' beam of constant path length. This enables any spatial phase modulation within the probe beam to become visible as spatial *intensity* modulation in an interferogram.

A modern example is the 'Wyko' system of Wyant et al.[6], shown in figure 2-1. The operation of the system is as follows. Light from an extended source is imaged at the field stop, which is further imaged onto the sample surface[7]. The light is then reflected back through the objective lens, and imaged onto a CCD array. In order to get interference, some of the illuminating light is reflected back from a semi-reflective plate onto a small reflecting disc on a glass plate and again back to the semi-reflective plate where it recombines with light from the object surface. The figure only shows the path of

---

[2]M. Francon, "Progress in Microscopy", chapter 2, Modern Trends in Physiological Sciences, International Series of Monographs on Pure and Applied Biology, Pergamon Press, 1961.

[3]F. Zernike, "Phase Contrast, a New Method for the Microscopic Observation of Transparent Objects", part I , Physica, IX, No. 7, pp.686-693, July 1942 ; part II, Physica, IX, No. 10, pp.974-986, Dec. 1942.

[4]M. Born, and E. Wolf, "Principles of Optics", sixth edition, Pergamon Press, 1980.

[5]J.W. Goodman, "Introduction to Fourier Optics", pp.144-146, Physical and Quantum Electronics Series, McGraw Hill, 1968.

[6]J.C. Wyant, C.L. Koliopoulos, B. Bhushan, and D. Basila, "Development of a Three-Dimensional Noncontact Digital Optical Profiler", ASME J. Tribology, Vol. 108, No. 1, 1986.

[7]this is critical illumination as detailed in M. Francon, "Progress in Microscopy", p.28, Modern Trends in Physiological Sciences, International Series of Monographs on Pure and Applied Biology, Pergamon Press, 1961.

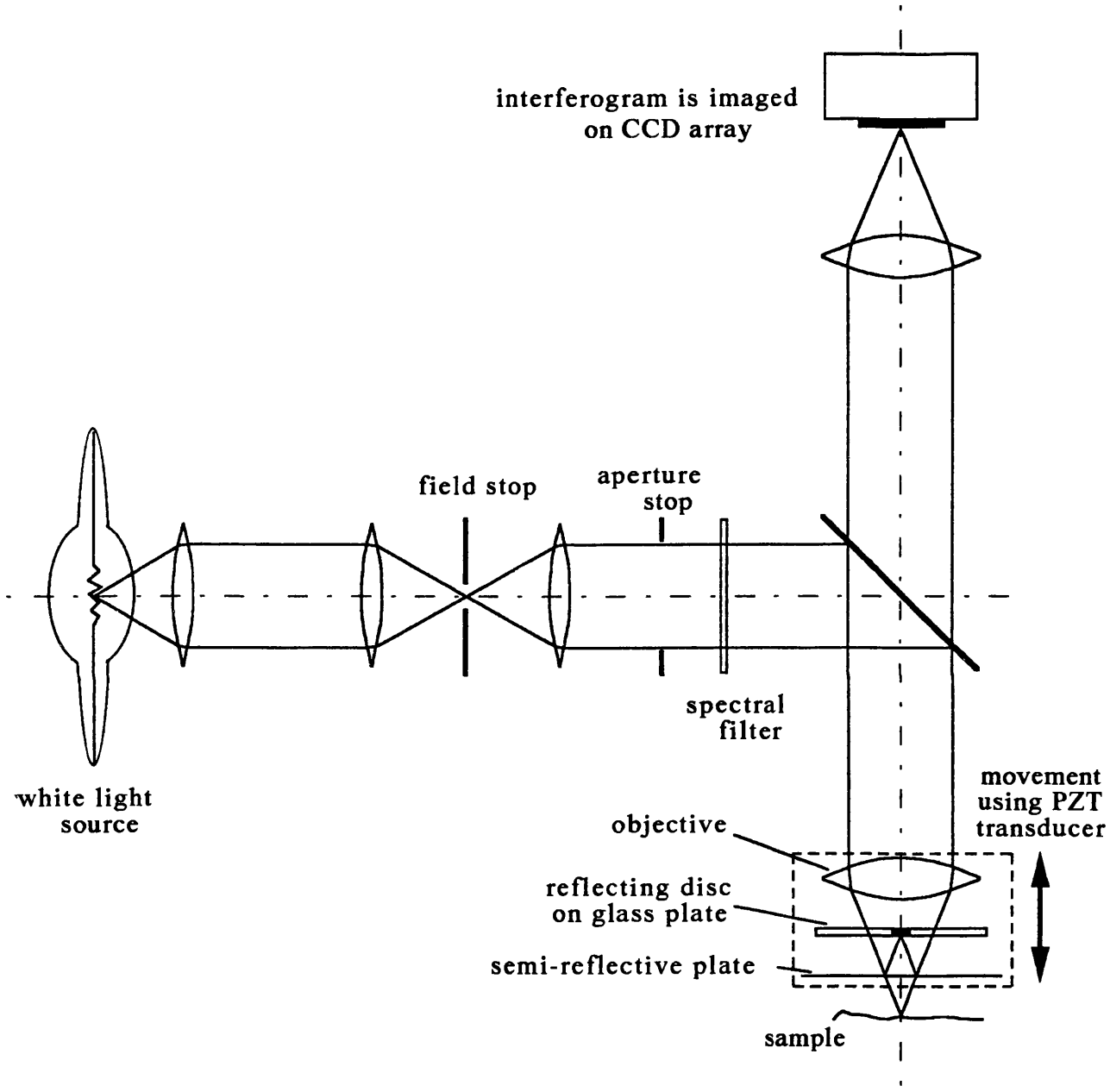


Figure 2-1 The Wyko full field phase stepping homodyne interferometer

light from one point on the source. A bundle of rays starting from another point on the source will be imaged at a displaced position on the sample and reflecting disc, and hence onto a different pixel of the CCD.

The illuminated field area is determined by the field stop, and the reflecting disc must be at least as large as the illuminated area on the sample surface. The presence of this disc will somewhat affect the distribution of the focussed light on the sample surface. However since the light probing all areas on the sample will be affected in a similar way, the resolution of the resulting image will not vary across the field of view. It may be noted that the light at each pixel originates from different locations on the extended source. A spatially coherent source is unnecessary, although from the point of view of lens performance, it is preferable that light emitted from different areas on the source contains similar spatial frequencies.

If the reflecting disc is totally flat, and mounted parallel to the semi-reflective plate, then any phase variations on the sample will become apparent as intensity variations at the CCD. The intensity at each pixel is,

$$I = |E_s|^2.|r_s|^2 + |E_m|^2.|r_m|^2 + 2\,|E_s|.|E_r|.|r_s|.|r_m| \cos(\theta + \phi) \qquad ...(2-1)$$

where r_s and r_m are the amplitude reflection coefficients at corresponding points on the sample and reflecting disc surface, θ is a phase term revealing information about changes in the optical path length imposed by the surface, E_s, E_r are the amplitudes of the light incident at the sample and reference surfaces, and ϕ is a phase constant.

It can be seen that the the detected intensity at each pixel is related to the cosine of θ, the phase structure. There could be severe inaccuracies in this system if the phase structure was deduced from one single full field intensity measurement. Not only will changes in reflectivity affect the intensity, but variations in illuminating light intensity

across the field of view will affect measurements. This is particularly important in this system because as already mentioned, the light for each measurement at the detector array originates from a different location on the source. Both of these problems, as in other homodyne systems, are overcome by a method termed 'phase stepping'. A series of measurements are made, introducing a known phase difference between the arms of the interferometer for each measurement ie., stepping the value of ϕ In this system, a PZT transducer moves the objective, glass plate and semi-reflective plate to four positions, and a simple four step algorithm is used to numerically calculate variations in θ, independent of the object (intensity) reflectivity and illuminating light intensity.

So unlike the Schlieren, dark ground and Zernike techniques, this system can be used to measure object induced phase *independent* of reflectivity variation. Furthermore, quantitative measurements may be performed. Under controlled environmental conditions, a fully optimized version of this system has a quoted[8] depth resolution of 1 Ångstrom (no measurement bandwidth is given). However, despite such a seemingly high accuracy, there are problems associated with this system. The phase measurement is made by interfering light which has been reflected from the sample and reference surfaces. If there are any vibrations in one component through which both beams have travelled, this results in approximately the same random phase variations in each beam, and the effects of such phase noise will cancel out in the interference process. It can be seen however, that the one path which is not common is between the sample and objective. This is the weak link, and necessitates extremely careful design to minimize non-common mode vibrations of the sample and semi-reflective plate.

A better approach would be to use all common path elements. This is no easy achievement since the phase of light reflected from the reference surface should be constant. One system developed by Iwata et al.[9] uses birefringent lenses to do just

[8] refer to the system specification section of the sales brochure for the 'TOPO™-2D', a measurement system made by the Wyko corporation, Tuscon, Arizona, USA. (1991)

[9] K. Iwata and T. Nishikawa, "Profile measurement with a phase shifting common path polarization interferometer", Proc. SPIE conference, 'Laser Interferometry – Quantitative Analysis of Interferograms: Third in a Series', San Diego 7-9 August 1989, Vol. 1162, No. 43 (to be published).

this, and is shown in figure 2–2. Referring to this figure, linearly polarized light from a laser passes through a Soleil compensator[10] which is used to introduce a known phase shift between the two polarizations. The light emerging from the Soleil compensator enters the double focus lens inserted before the microscope objective. This lens contains a birefringent concave element, and it is constructed so that the ordinary ray will pass through the lens undeviated. The double focus lens is positioned so that its focal plane coincides with the front focal plane of the objective. Thus the extra-ordinary ray, which was affected by the birefringent element, becomes parallel after passing through the objective lens. This beam, on reflection, carries the phase information from the object surface, and is termed the probe beam. The ordinary ray passes through the double focus lens unaffected and is focussed to a small spot on the object surface. This ray is used as the interference reference beam.

On reflection, light entering the double focus lens will have passed twice through the quarter wave plate, causing in total, a 90° rotation in polarization. So the probe beam passes through undeviated whereas the reference is deflected by it. The remaining two lenses are necessary to give both beams identical phase fronts, ie., parallel beams. A compensator (eg. Soleil) is used to adjust the polarizations of the probe and reference beams so that they are exactly orthogonal, and after passing through a polarizer, the resulting interference is detected by the CCD camera. As in the Wyko system, the object induced phase is calculated from a series of images, the Soleil compensator being used to introduce a phase shift in one of the interferometer paths.

A problem highlighted by the designers concerns reflections from elements within the system. They suggest using anti-reflection coatings, and also a shorter coherence length light source. In this system, as for the Wyko interferometer, the temporal coherence of a laser is not needed because the path length difference between both beams is of the order of a wavelength. The Iwata interferometer however, unlike that of Wyko, does need a source which has enough spatial coherence to ensure that a well collimated beam of light enters the system.

---

[10] see for example "Optics" by E. Hecht and A Zajac, first edition, Addison-Wesley, 1974.

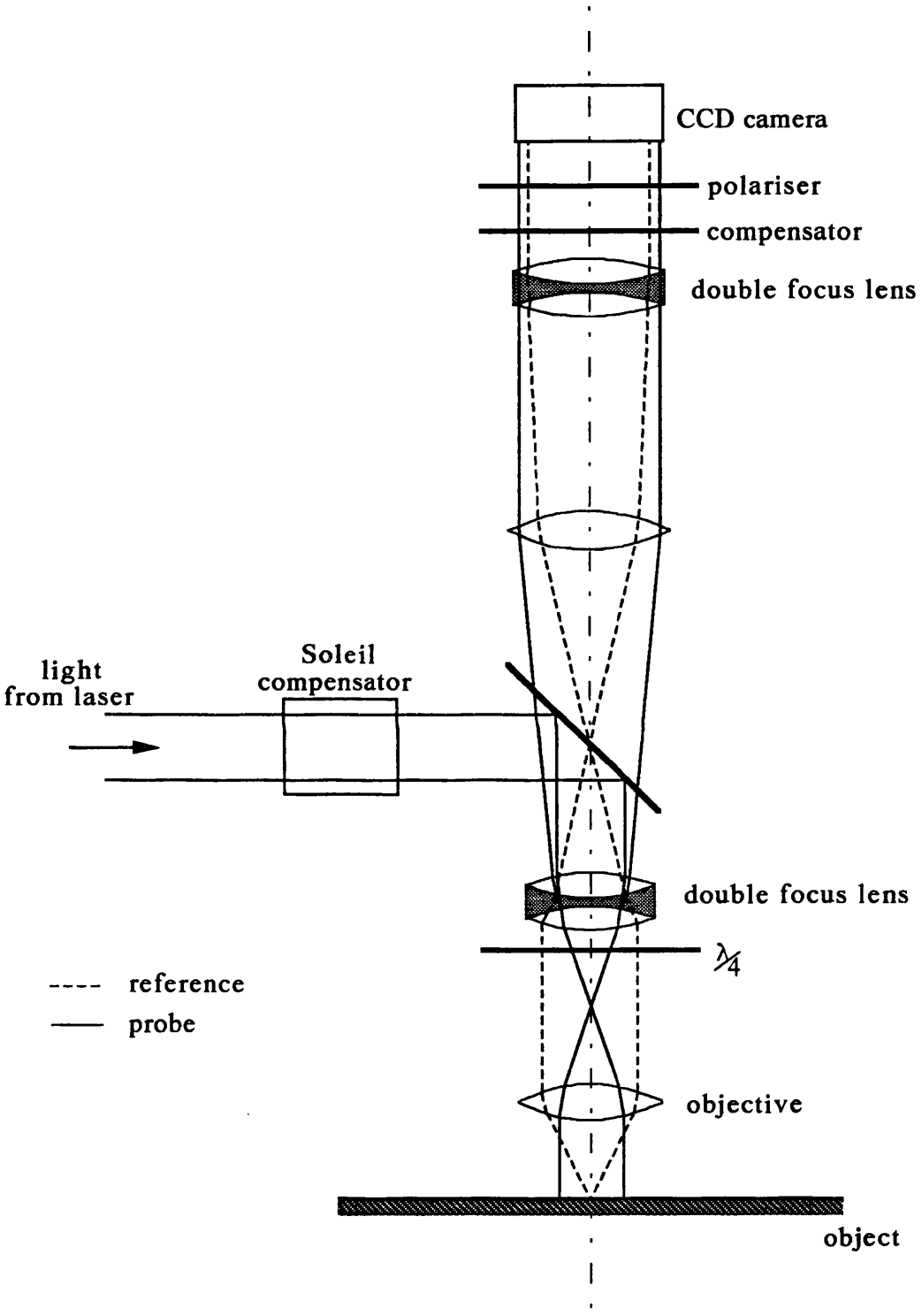
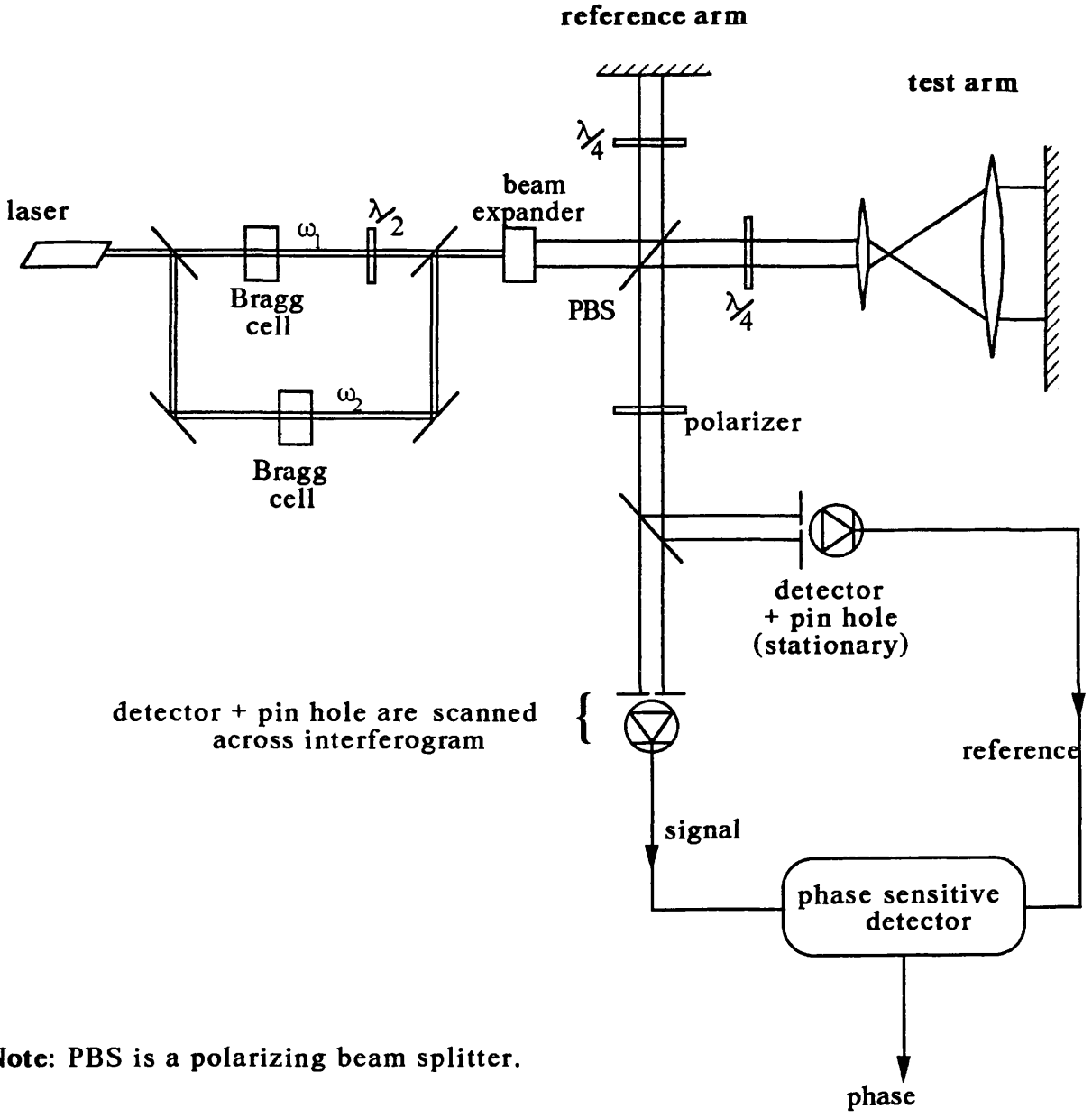


Figure 2-2 Common path full field phase shifting interferometer

The Iwata interferometer, although potentially being able to reduce noise due to microphonics below that of the Wyko system, has many unquantified possible sources of error. The performance of the system hinges on the quality of the double focus lenses : aberrations may be a serious problem. A quarter wave plate will function properly only if the incident light is at normal incidence. As in any phase stepping interferometer, the accuracy of the final phase measurement is dependent on the accuracy of the phase step. In this interferometer, if the phase steps are not exactly 90°, the relative intensities of probe and reference beams will change, and so affect the calculated phase. This problem is not associated with the Wyko system.

Massie[11] et al. have taken a quite different approach to eliminate the problem of microphonics, which also does away with the need for phase stepping. Referring to figure 2-2 which shows the system configured for measuring the quality of an optical flat or lens, a linearly polarized laser is divided into two beams, which are frequency shifted by ω_1 and ω_2, and recombined such that the two frequency components are orthogonally polarized. After passing through a beam expander, a polarizing beam splitter is used to direct one of the frequency shifted components into the reference arm and one into the test arm. Upon reflection, each beam will have passed through a quarter wave plate twice, resulting in a 90° rotation in polarization so that light from each arm will pass through the polarizer. This is set at 45° relative to the two incident polarizations, causing interference fringes in front of both detectors. Since there is a frequency difference of $\Delta\omega = (\omega_1 - \omega_2)$ between the two interfering beams, the fringes will be moving across the detectors at this frequency, or alternatively one may consider that the intensity of any one point on the interferogram to be fluctuating at this frequency.

[11] N.A. Massie, R.D. Nelson, and S. Holly, "High performance Real Time Heterodyne Interferometry", Appl. Opt., Vol. 18, No. 11, pp.1797- 1803, June 1979.



Note: PBS is a polarizing beam splitter.

Figure 2-3 A full field heterodyne interferometer

The intensity response from any one point on the interferogram is,

$$I = |E_s|^2.|r_s|^2 + |E_m|^2.|r_m|^2 + 2\,|E_s|.|E_r|.|r_s|.|r_m| \cos(\Delta\omega t + \theta + \phi) \qquad ...(2-2)$$

where θ, r_s, E_s, r_m, E_r, and ϕ have the same definition as in equation (2 – 1). So to measure the object induced phase variation across the field of view, all that is required is to measure the variation in phase of the $\Delta\omega$ frequency component at all points on the interferogram. This is achieved with phase sensitive detection, using one point on the interferogram as a reference and scanning a second pin hole plus detector across the interferogram. Unlike the two previously discussed homodyne techniques, this heterodyne interferometer requires only one measurement to determine the object induced phase variation of the light. Phase stepping which is necessary for homodyne interferometers, may be viewed as simulating the heterodyne detection scheme of this latter interferometer.

What is most important to notice in this system is how the phase reference signal is generated. Phase noise due to microphonics will be more or less constant across the interferogram. By taking one single point on the interferogram as the reference for measuring the phase of all other points, the object induced phase may be measured in the absence of noise from microphonics. The system can be seen to be similar to that of Iwata in that one point on the sample is used as the phase reference for measurements over the entire field of view.

We have seen that this interferometer will experience low microphonics noise characteristic of a common path interferometer such as that of Iwata. Furthermore inaccuracies introduced in phase stepping will be absent. In these respects, as well as considering other associated problems with Iwata's system, the potential performance of this technique is superior to both of the previous methods. An additional advantage in heterodyne detection is the considerable reduction in the $1/f$ noise which is so prevalent at

dc. Even though the quoted accuracy of the this interferometer is only $\lambda/100$, this is for an initial 'bread board' construction, and if well constructed, will no doubt have a far greater sensitivity.

The disadvantage of this interferometer is that there is at present no cheap method of performing full field phase sensitive detection. Instead one detector is used and scanned across the interferogram, which inevitably increases the image acquisition time. Despite a full field image being formed, only a small amount of the optical power is monitored at any one time, thereby reducing the signal to noise level. To overcome this problem, more recently developed techniques have tended to be true scanning systems where all the available light is focussed onto one point on the specimen. A single measurement is then made before moving the light to another position.

A full field interference technique which should be mentioned is the well established technique of Nomarski [12, 13]. Each point in the image is the result of interference of light from two close points on the object surface and the response is differential, sensitive to changes in phase eg edges. Since this is also a homodyne system, it is not possible to distinguish between real phase objects, and objects where the reflectivity changes. The phase sensitivity of this type of technique is at best about $\lambda/100$, and it is probably best suited for qualitative investigation only.

In certain applications such as for example mirror surface characterization, the surface topography detail is not always wanted. Perhaps all that is required is a figure for the RMS surface roughness or autocovariance. In this case, scatter detection systems may be better suited. One example is the total integrated scattering[14] method (TIS) in which

---

[12] G. Nomarski, "Applications à la Metallographie des Method Interferentielles à Deux Ondes Polarisees", Revue de Metallurgie, pp.121-134, LII, No. 2, 1955.

[13] M. Francon, "Progress in Microscopy", pp. 118-119 and pp. 149-153, Modern Trends in Physiological Sciences, International Series of Monographs on Pure and Applied Biology, Pergamon Press, 1961.

[14] F.D. Correll, C.D. Ferguson, and R.A. Kant , "Optical measurements of the RMS roughness of ion-implanted surfaces", Nuclear Instruments and Methods in Physics Research, Vol. B24/25, pp. 581-583, 1987.

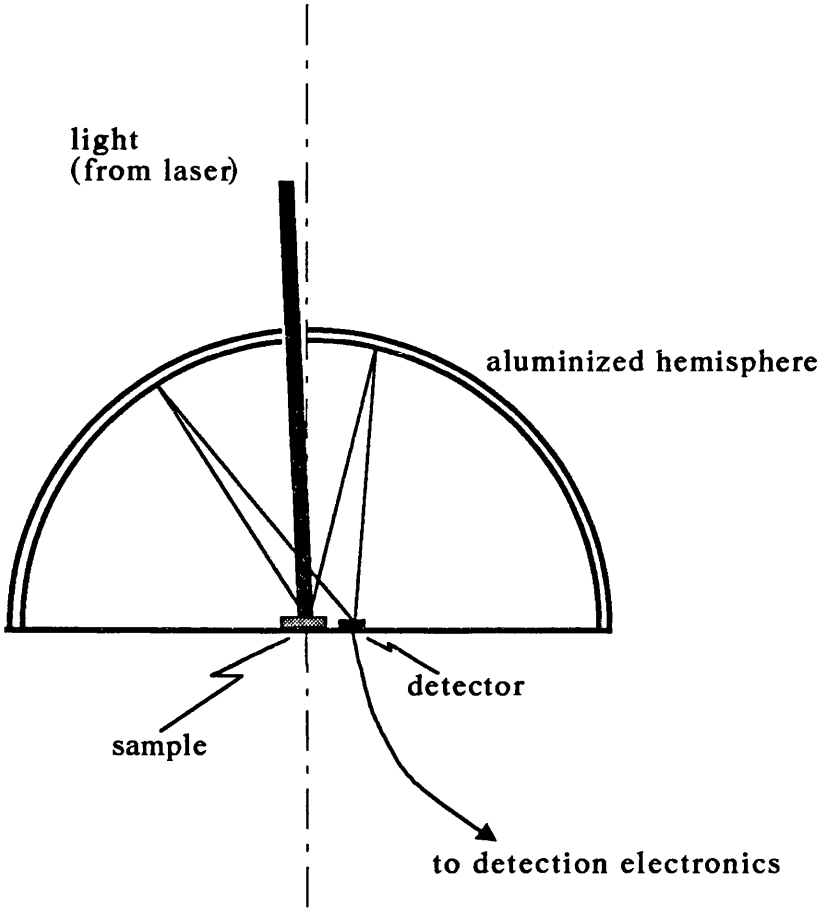


Figure 2-4 Schematic diagram of a total integrated scatter (TIS) instrument

collimated light is directed onto the sample surface, and all the off axis scattered light is collected by one detector (see figure 2-4). Fast statistical measurements may be made over large areas, but spatial resolution is necessarily low since the angular divergence of the light incident on the sample must be small. More details of this and other similar methods may be found in references 15 and 16.

2.2 Scanning systems

Since the early 1970's, a number of workers in the field of microscopy have been working on scanning versions of the interference contrast instruments. Rather than illuminating the entire specimen for all the time, focussed light is rastered across the specimen and sequential measurements are made which can then be used to form a line scan or image on a cathode ray tube. Whereas in Massie's system, only a small amount of the optical power was effectively contributing to each measured point on the interferogram, by scanning the light about the object, all the available light is always detected. This then increases the signal to noise ratio for the same given light source.

There are two ways in which scanning may be performed. Either the light is scanned over the object, or the object is translated with the optical elements remaining stationary. The latter method may not seem so appealing, but it does have the one virtue that all errors due to optical aberrations remain constant, and this is the method chosen by many workers due to its simplicity. The clear advantage of scanning the *light* is that the scan speed is potentially much higher than could be achieved with mechanical scanning. In a scanning differential system however, which is sensitive to extremely small phase variations, considerable care must be taken that lens aberrations in themselves do not effect significant phase variations.

---

[15] K.H. Guenther, P.G. Wierer, and J.M. Bennett, "Surface roughness measurements of low-scatter mirrors and roughness standards", Appl. Opt., Vol. 23, p. 3820, (1984).

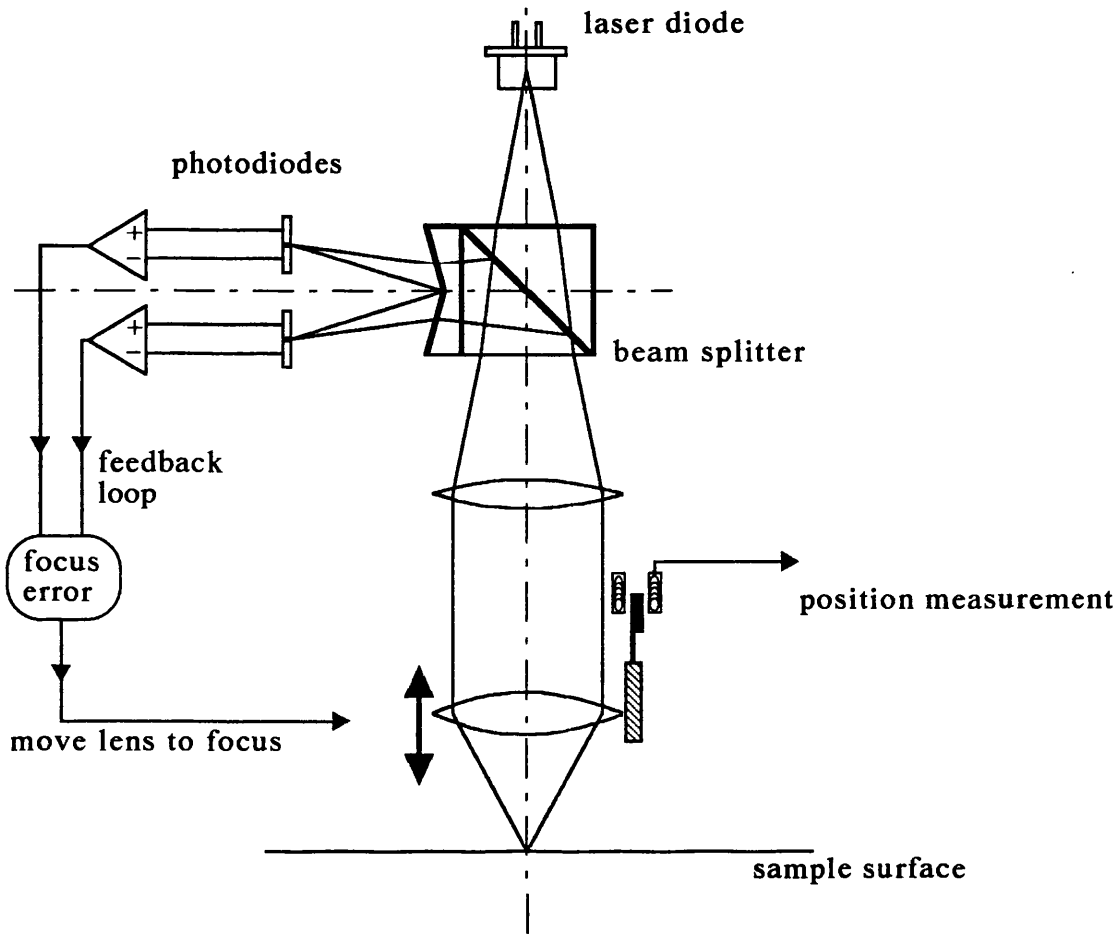
[16] see for example Proc. SPIE conference , 'Surface measurement and characterization', 19-21 September 1988, Vol. 1009.

Scanning optical systems can be divided into two categories, geometrical and interferometric techniques. The interferometric techniques can further be subdivided into homodyne, heterodyne and differential systems. These will be discussed in turn.

2.2.1 Scanning geometrical techniques

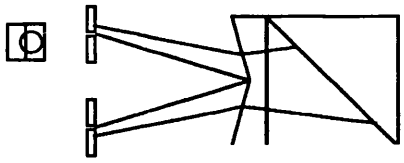
The two systems presented in this section measure the surface profile of an object by monitoring the divergence of reflected light from the surface. The first method (figure 2-5) devised by Brodmann[17] et al. uses the read head of a compact disc player with refined electronics. Light from a laser diode passes through a beam splitter, is collimated and focussed on the object surface. The beam splitter has an additional glass element on one side as shown, so that the reflected light from the object, instead of coming to a focus on axis, is divided into a pair of beams focussed at the centre of two split detectors. If the probing beam is reflected from a feature which is too close or distant from the lens, then the reflected light will no longer come to a tight focus at the detectors, but will form broader spots on each side of the split detectors as shown in cases A and C of figure 2-5. This is detected by the electronics, and using a feedback loop, the lens is moved in a direction which refocuses the light. The position of the lens is measured, and the object then translated so as to probe another small area on the surface. Although as described, the two sets of split detectors duplicate each other, if the feature examined is tilted then the light incident on the split detectors is no longer symmetrical. The effects of tilt are removed by deriving the focus error signal from the average of the two split detector currents.

[17]R. Brodmann, and M. Allgäuer, "Comparison of light scattering from rough surfaces with optical and mechanical profilometry", Proc. SPIE conference , 'Surface measurement and characterization', 19-21 September 1988, Vol. 1009, No. 16.

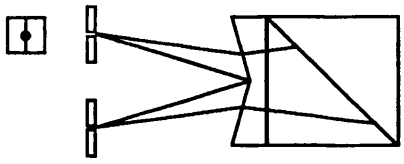


light intensity at photodiodes for different defocus positions of lens

A: lens too close to surface



B: in focus



C: lens too distant from surface

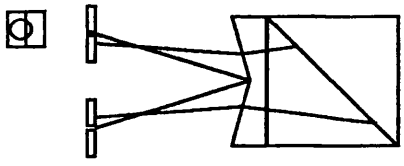


Figure 2-5 Dynamic focussing optical stylus

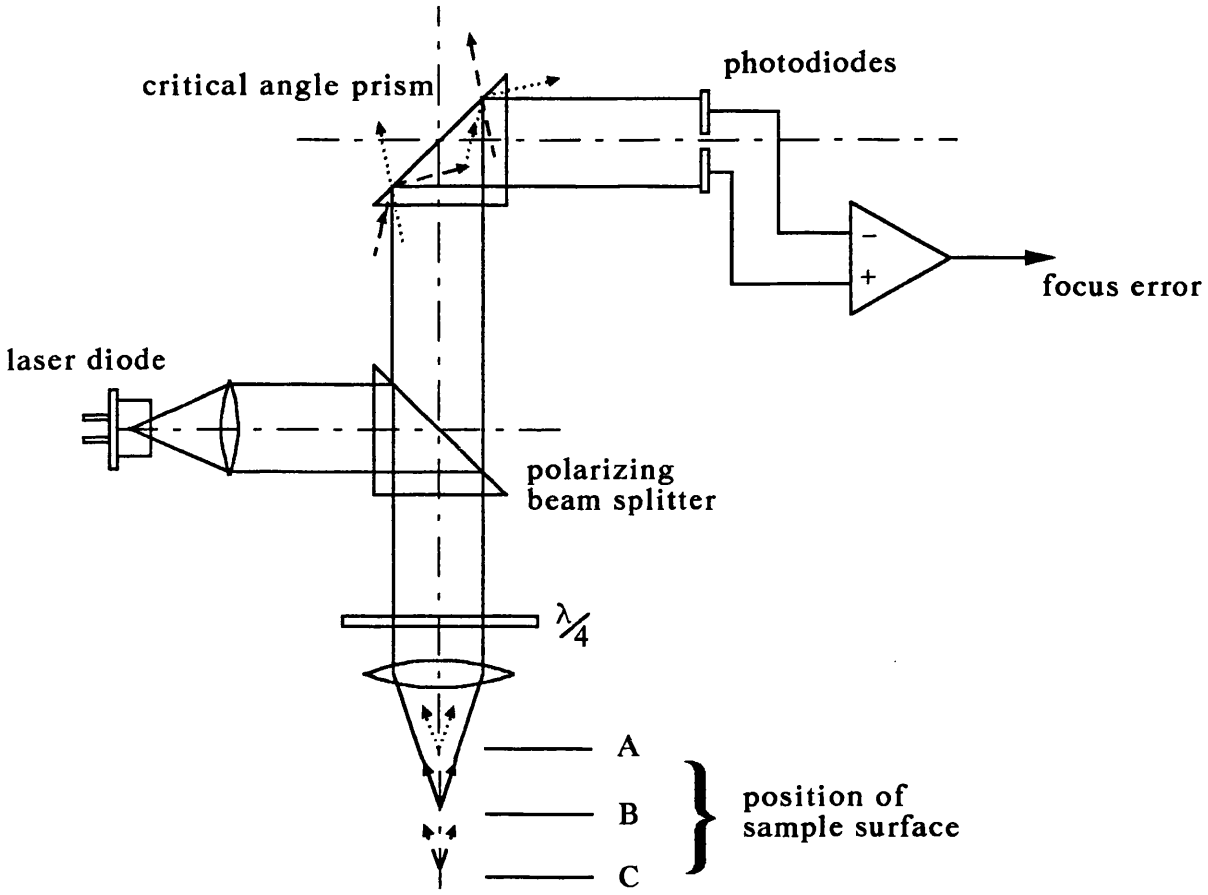
A second technique, the 'High precision optical surface sensor' (HIPOSS)[18] is shown in figure 2-6. This method uses instead of the split focus technique, a prism to differentiate between diverging or converging light reflected from the object. If the object is at focus, then light will enter the prism parallel to the optical axis and be total internally reflected in the direction of the two detectors. When the object is not at focus (case A and C), a proportion of the light entering the prism on one or other side of the optical axis will be above the critical angle and so be refracted *through* the prism. This results in less light incident on one of the two detectors, depending upon whether the feature is above or below focus.

In their paper, Kohno et al. show that the critical angle effect for p-polarized light gives a large fall off of total internally reflected light on one side of the axis, for only small displacements of the sample away from the focal plane of the objective lens. Furthermore, cascading two critical angle prisms (only one is shown) gives even greater contrast. Instead of moving the lens as in the Brodmann technique, the resulting focus error signal is used for direct height measurement. For features less than about a micron in height, this signal is proportional to the actual height. The HIPOSS systems as illustrated in figure 2-6 will suffer from problems of object tilt. This is overcome in a similar way to Brodmann et al., duplicating the critical angle prisms and photodiodes already used.

Both these geometrical techniques can attain surprisingly good depth resolutions, the Brodmann system of less than 1nm, and the HIPOSS system, better than 2Å. Both are nevertheless limited by microphonics, the HIPOSS having a 200Å depth resolution in 20kHz bandwidth[19]. This figure should be compared to that of heterodyne systems presented later which have depth resolutions many orders better, in the same bandwidth. These techniques are thus not well suited to applications such as imaging smooth surfaces with RMS roughness of for example 10Å.

[18]T. Kohno, N. Ozawa, K. Miyamato, and T. Musha, "High precision optical surface sensor", App. Opt., Vol. 27, No. 1, pp. 103-108, 1 January 1988.

[19]this figure was stated in the presentation, 'On a few functions of HIPOSS and their applications', T. Kohno, SPIE Hamburg 19-21 September 1988, (1009-03)



light intensity at photodiodes for different height features

A: close (divergent)

B: in focus (parallel)

C: distant (convergent)

Figure 2-6 The High precision optical surface sensor (HIPOSS)

The critical angle prisms in HIPOSS have a superior angular contrast than the split focus detection of the former which must use extremely low noise electronics in order to approach the sensitivity of HIPOSS. Conversely, HIPOSS will give less reliable results in cases where the sample changes the polarization of the reflected light.

This geometrical optics approach to surface profilometry, although in the present implementations are limited by microphonics, do nonetheless have several interesting properties which cannot be overlooked. Variations in the surface reflectivity will, for low NA objectives (<0.5), have no effect on measurements. In this review, these are the only true homodyne techniques which can eliminate reflectivity variation from the system response without resorting to phase stepping. Second, both methods are insensitive to phase changes in the reflected beams, which are due to variations in surface conductivity. This is perhaps the single most important property as no other technique discussed in this chapter can properly distinguish between changes in surface conductivity or surface height.

2.2.2 Scanning interferometric techniques — homodyne

A homodyne interferometer which shares many similarities with that of Iwata discussed earlier, is that of Downs et al.[20, 21] which is shown in figure 2–7. Referring to this figure, linearly polarized light is incident on a compound birefringent lens which focuses the ordinary rays via an objective lens, to a tight focus on the object surface. This is the probe beam. The focal plane for the extra-ordinary rays is beyond the sample, so they form an out of focus spot on the surface. This is used as the interference reference; the incident power in each of these beams is adjusted to be equal.

---

[20] M.J. Downs, "Surface profile interferometer", UK Patent 2 109 545B, 24 July 1985.

[21] M.J. Downs, W.H. McGivern, and H.J. Ferguson, "Optical system for measuring the profiles of super-smooth surfaces", Precision Engineering, Vol. 7, No. 4, pp.211-215, October 1985.

Before passing through the birefringent lens for a second time, the two beams will have passed twice through the quarter wave plate, reversing the birefringent effects on the two beams. So the rays labelled 'e-polarization' will pass straight through the lens, with the 'o-polarization' rays superimposed on top. The reflected probe and reference beams will have slightly different divergence angles, and although this is not depicted in figure 2–7, it is none-the-less important as it affects the ultimate interference efficiency between the two beams.

After the birefringent lens the light will be elliptically polarized, the degree of ellipticity dependent on the phase difference experienced by each beam as well as the relative intensities of the two beams after reflection. In order to measure the ellipticity, the light passes through a further quarter wave plate orientated so that the two beams are circularly polarized in the clockwise and anti-clockwise directions. The resulting polarization of the two beams together is linear, with an orientation angle proportional to the ellipticity angle. If the analyzer is set such that in the absence of phase structure there is no transmitted light, then for small phase objects, the intensity of light at the photodetector is,

$$I = A|r_r|.|r_r + \delta r| \,.\delta\theta \qquad \ldots(2-3)$$

where r_r is the average reflectivity experienced by the reference beam, $(r_r + \delta r)$ is the reflectivity experienced by the probe, $\delta\theta$ is the phase difference between probe and reference beams and A is a constant (determined by calibration of the equipment).

To increase accuracy, an ac detection scheme is used. A Faraday rotator modulates the polarization angle at frequency ω_F, and the amplitude of this frequency component is detected. Using feedback, a dc offset is applied to the rotator and the polarization is rotated to null the ω_F frequency component. This dc offset is used as a measurement of polarization angle. It may be noted that the detection scheme described here is different to that described in the references because their explanation appears to be in error.

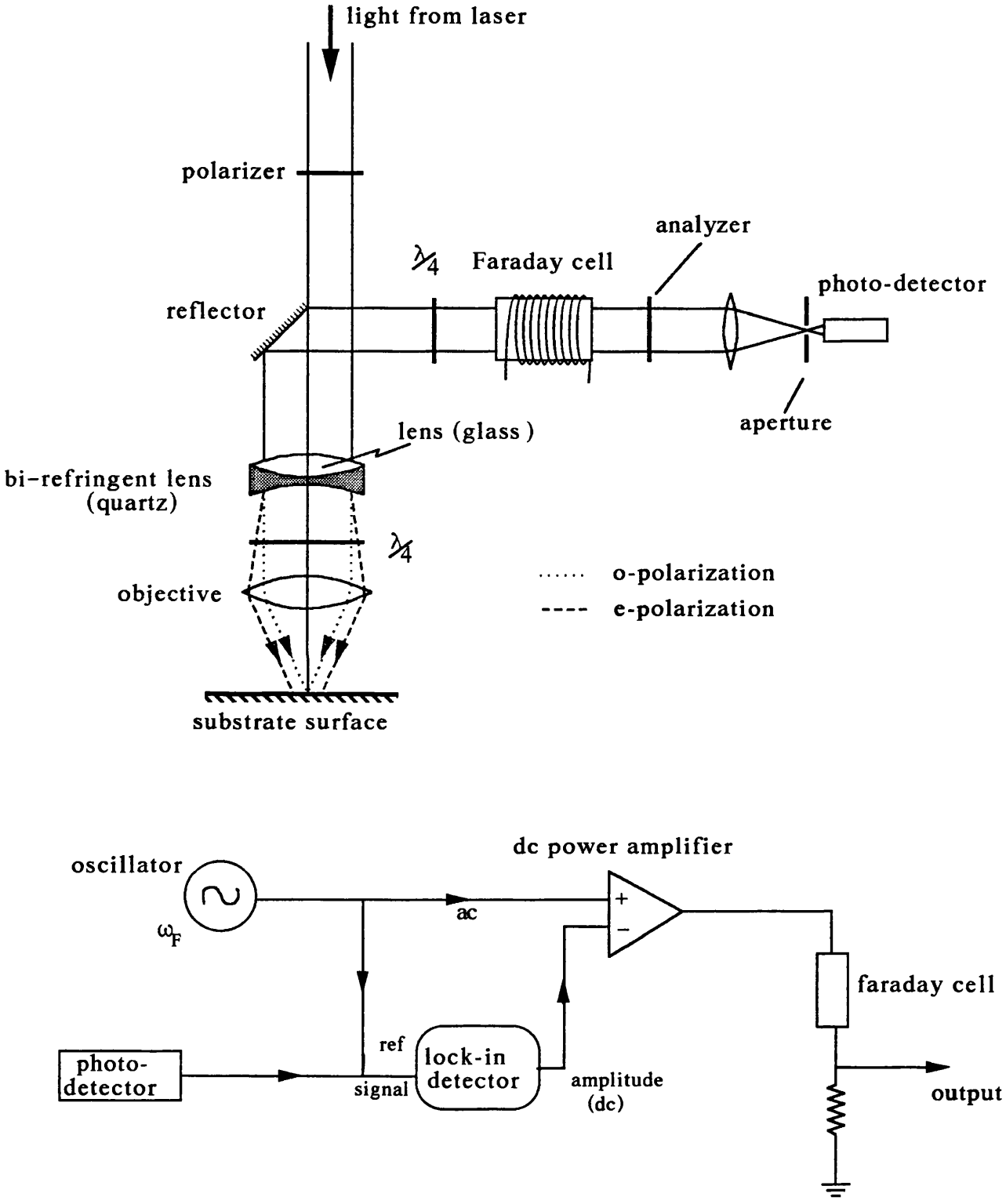


Figure 2–7 The Downs homodyne optical profilometer

This system is important because it is one of the first of its kind which generates a phase reference by averaging the phase of light reflected from a relatively large area on the object surface. As the two beams are scanned over the surface, the mean phase of the reflected reference should stay constant. It is clearly an advantage therefore for the ratio of focussed probe to reference beams sizes to be as large as possible. As this ratio is increased the difference in divergence of both beams after the second pass through the birefringent lens will increase, with a consequent reduction in interference efficiency at the photo-detector. This problem is not unsurmountable, and may be resolved, following the example of Iwata, by using an additional birefringent lens before the second quarter wave plate. A further more serious problem concerns the quarter wave plate immediately above the objective, which will only function correctly if the incident light is collimated. This is not the case, and the problem is exacerbated as the ratio is increased.

To conclude, this is a true common path scanning technique, with angstrom depth resolution (the sensitivity is 0.03Å in a 200Hz bandwidth[22], and is not limited by microphonics). Problems associated with the system are that the response is dependent on the surface reflectivity, and the ratio of reference beam to probe size may not be made large enough to ensure a constant phase reference. Also, like any technique which uses orthogonal polarizations of light to carry two types of information, interpretation of results from specimens exhibiting birefringence will require considerable care.

[22]private communication with M.J. Downs

2.2.3 Scanning interferometric techniques — heterodyne

The heterodyne interferometer is formed by introducing a relative frequency shift between the probe and reference beams in a standard homodyne interferometer. One such example is that designed by Massie et al. which has already been discussed in section 2.1. This system lacked a suitable full field heterodyne detector and had to resort to scanning the detector across the interferogram which wastes optical power and hence results in a lowering in the sensitivity bandwidth.

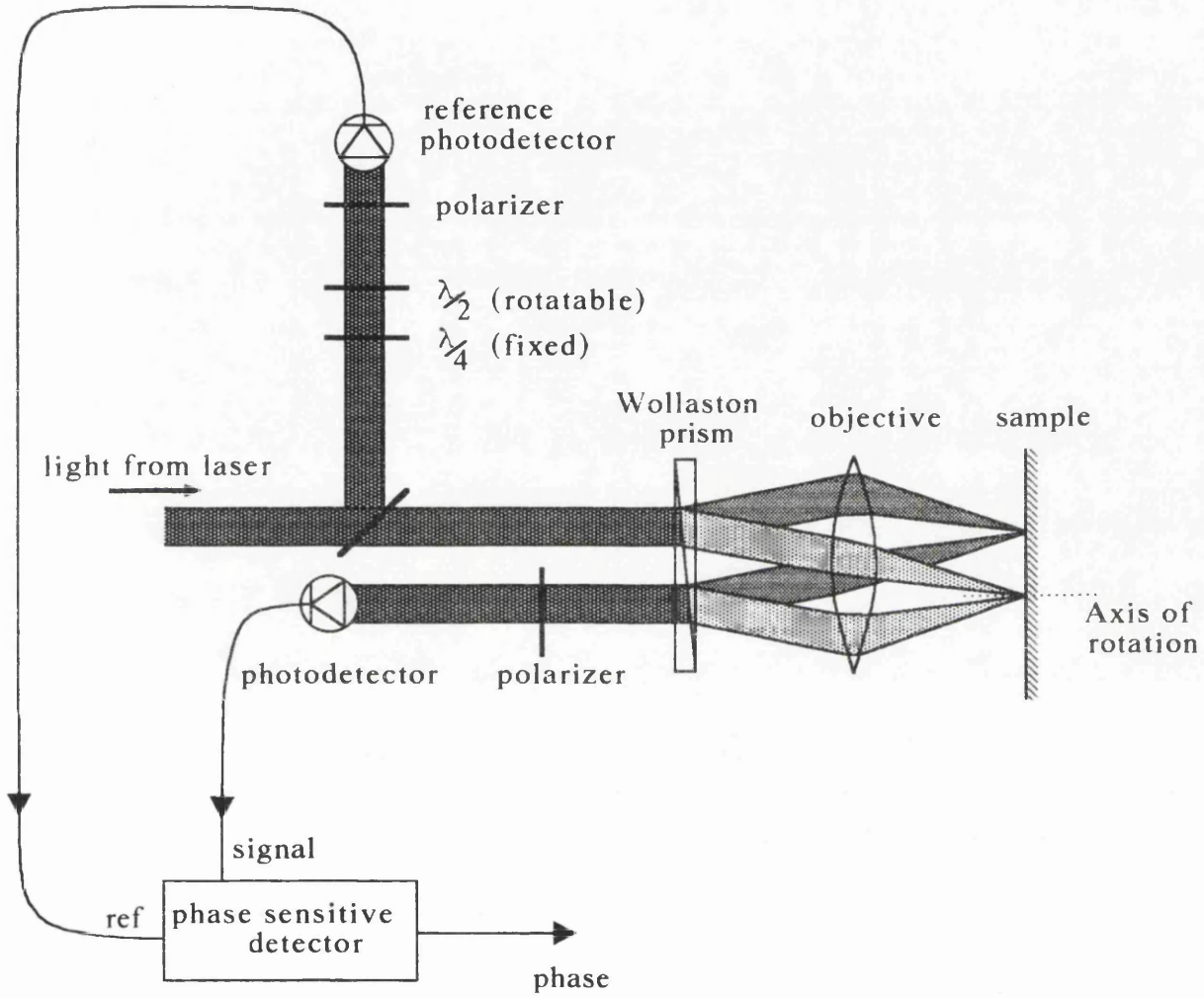
Even though a phase stepping homodyne interferometer is capable of sensing just the phase of reflected light from an object, its practical sensitivity is determined by the accuracy of the phase stepping. For this reason and also the reduction in $1/f$ noise associated with ac detection, a heterodyne interferometer is better suited for the most accurate metrology applications. The biggest remaining hurdle is how to minimize noise due to microphonics, so elegantly demonstrated in the systems of Massie, Iwata, and Downs. This section presents four approaches to tackle the problem.

I G.E. Sommargren [23]

A system developed by Sommargren is illustrated in figure 2-8, to which reference is now made. A Zeeman split He-Ne laser is used to produce two coincident and collinear light beams with orthogonal polarizations and a frequency difference of 2 MHz. A beam splitter directs a small fraction of the light through two wave plates and polarizer. Rotation of the half wave plate affects the phase of the detected reference signal and this serves as an extremely neat non-electronic means of ensuring that the phase sensitive detection electronics operates in mid-range, without the inconvenience of the phase wrapping around from minus to plus 180°.

---

[23] G.E. Sommargren, "Optical heterodyne profilometry", Applied. Optics, Vol. 20, No. 4, pp. 610-618, 15 February 1981.



Note : The laser used is a stabilized Zeeman split He:Ne laser. The two orthogonal polarizations differ in frequency by 2 MHz.

Figure 2-8 The Sommargren heterodyne profilometer

The transmitted light passes into a Wollaston prism which angularly separates the two polarizations, and an off axis objective focuses the light onto two separate areas on the object surface. The light is then reflected from the object back through the Wollaston prism which recombines the two beams, and after passing through the polarizer, is detected. The resulting interference signal is,

$$I = A|E_r|.|E_p|.|r_r|.|r_p| \cos(\Delta\omega t + \theta + \phi) \qquad \ldots(2-4)$$

where θ is the phase due to features on the object ; r_r, r_p are the reflectivities experienced by each beam which have associated electric field strengths E_r, E_p; $\Delta\omega$ is the difference frequency between the two polarizations (ie., 2MHz), and A, ϕ are constants.

As in any heterodyne technique, the object induced phase variation (θ) can be measured directly using phase sensitive detection. This measurement is independent of variations in the intensity reflectivity of the object, and also considerably isolated from the effects of amplitude fluctuations of light from the laser.

This system uses a unique and for most purposes, inflexible scan design. The axis of rotation of the object is as indicated on figure 2-8, so that the beam incident on axis, is used as the reference, whilst the other beam is used to probe the specimen. Thus the apparatus may be used only for circular line scans.

This system is a true common path system with angstrom depth sensitivity but a distinctly awkward scan arrangement. Although not mentioned by Sommargren in his published paper, it should be noted that the system can be readily converted into a differential phase measurement technique. This is achieved by scanning both beams across the object, with constant separation. In order to reduce the beam separation, a 'weaker' Wollaston prism must be used. Differential techniques are discussed in section 2.2.4

II C.C. Huang [24]

This profilometer, unlike Sommargren's, allows a conventional scan arrangement to be used. Referring to figure 2-9, linearly polarized light from a laser is divided by a Bragg cell into two beams with a relative frequency shift of ω_B. The two are passed respectively through a beam reducer and beam expander and are brought back together again to interfere at photodiode A, the output of which is used as a phase reference.

A proportion of each beam is also focussed via a quarter wave plate on the object surface. The focused spot sizes are determined by diffraction, and so are in proportion to the incident beam diameters. The large expanded diameter beam will form a tightly focussed probe on the surface, and the reduced diameter beam, when focussed, covers a large area, and is used as the reference.

The two reflected beams are directed via the 'beam reducer' into photodiode B, where the interference is detected. The effect of going in the reverse direction through the beam reducer is to increase the sizes of each beam. The ratio of the beam diameters however remains constant, so results in absolutely no effect on the resulting interference.

The interference signal from both detectors is at frequency ω_B. Measuring the phase of this frequency component from photodiode B will give a direct measure of the object phase structure. Although clearly not a common path system, by using the output of photodiode A as the reference for the phase sensitive detection, the resulting noise due to microphonics can be reduced considerably below that of other noise sources. This is because the light which is used to generate the optical reference at photodiode A has travelled along exactly the same path as that which is used to generate the signal at photodiode B. This scheme is similar to that used in Massie's system, discussed earlier. It can be seen that the noise due to microphonics is not eliminated from each photodiode response, but since the phase noise in both the electrical reference and signal is the same, it is eliminated in the electronic phase sensitive detection.

[24] C.C. Huang, "Optical heterodyne profilometer", Optical Engineering, Vol. 23, No. 4, pp. 365-370, July/August 1984

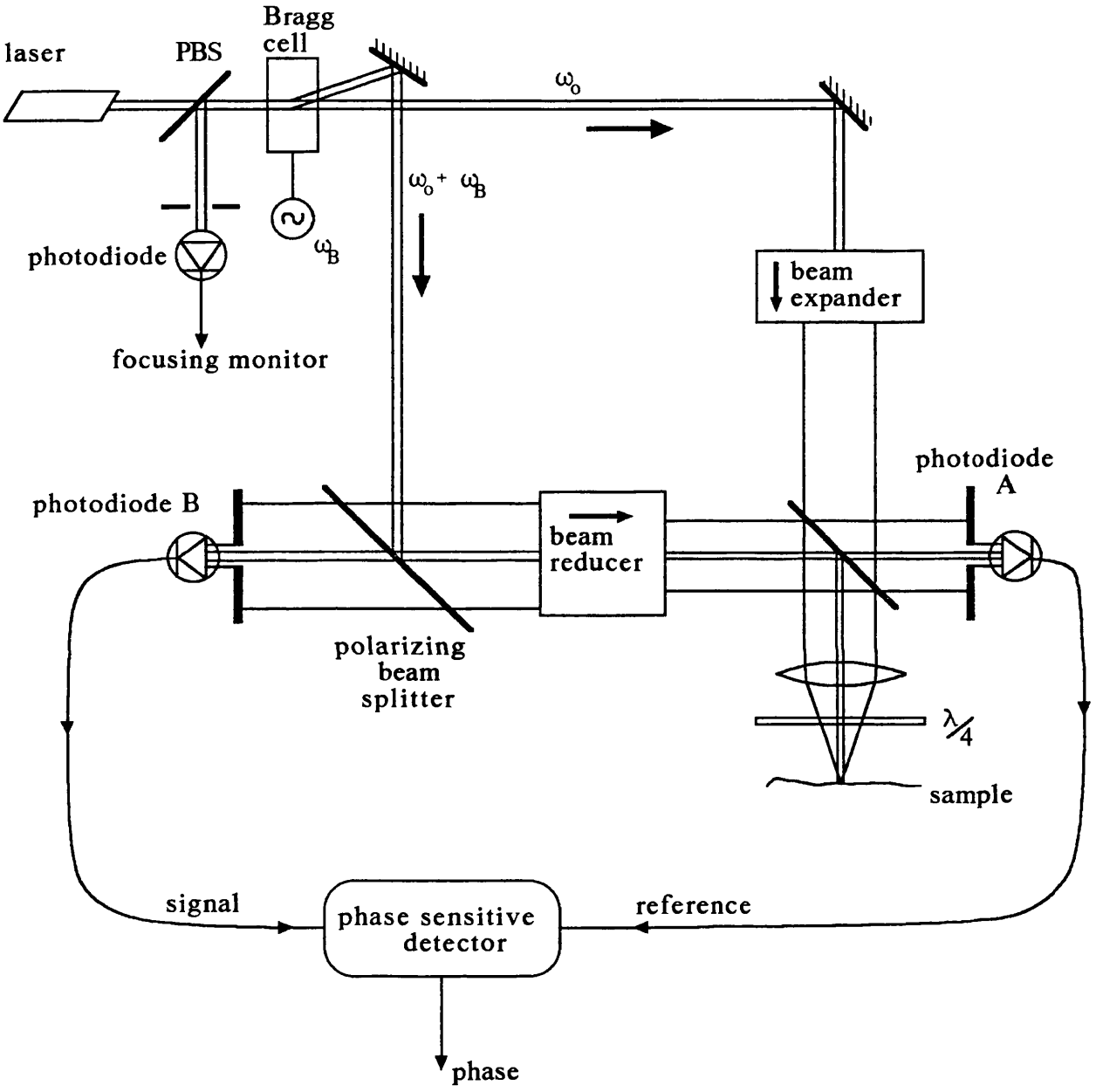


Figure 2-9 The Huang heterodyne profilometer

Note: polarizing beam splitters (PBS) are used only where indicated

Also shown on figure 2–9 is a photodiode for monitoring the focus. Some of the expanded probe beam is reflected back from the object through the Bragg cell and into the focussing photodiode. If the sample is not at focus, then the dc photodiode current drops and this can be monitored.

This system has many positive aspects. The biggest problem is however, as for Downs system, in achieving a large ratio of probe to reference beam diameters. Increasing the ratio reduces the interference efficiency and hence sensitivity bandwidth of the system. Experimental results quoted in the cited reference indicate a sensitivity of around 12Å in a 250Hz bandwidth.

III R.L. Jungerman, P.C.D Hobbs and G.S. Kino [25]

Jungerman et al. use a Bragg cell as the central element (figure 2–10). Light from a laser is split into a zeroth and first order beam which are focussed onto the object. The light is recombined after a second passage through the Bragg cell, and the resulting interference detected. By ramping the Bragg cell drive frequency, f_s, the first order probe beam is scanned across the object, leaving a stationary zeroth order reference beam. The response at the detector will be as given by equation (2 - 4), where $\Delta\omega = 2f_s$.

According to Jungerman, if the electrical reference used for the phase sensitive detection was obtained by simply frequency doubling the Bragg cell drive, the finite acoustic delay in the Bragg cell would cause considerable errors. Instead they derive an optical reference by directing a small amount of laser light into the Bragg cell such that the light subtends an identical angle with the acoustic field as the initial beam from the laser, but remains spatially distinct. The zeroth and first orders are reflected from an area on the sample previously known to be flat, and after a second passage through the Bragg cell, the reflected light is detected in the same way as the first set of beams. The interference signal at $2f_s$ is used as the optically derived reference signal.

[25] R.L. Jungerman, P.C.D. Hobbs, and G.S. Kino, "Phase sensitive optical microscope", Appl. Phys. Lett., Vol. 45, No. 8, pp.846-848, 15 October 1984

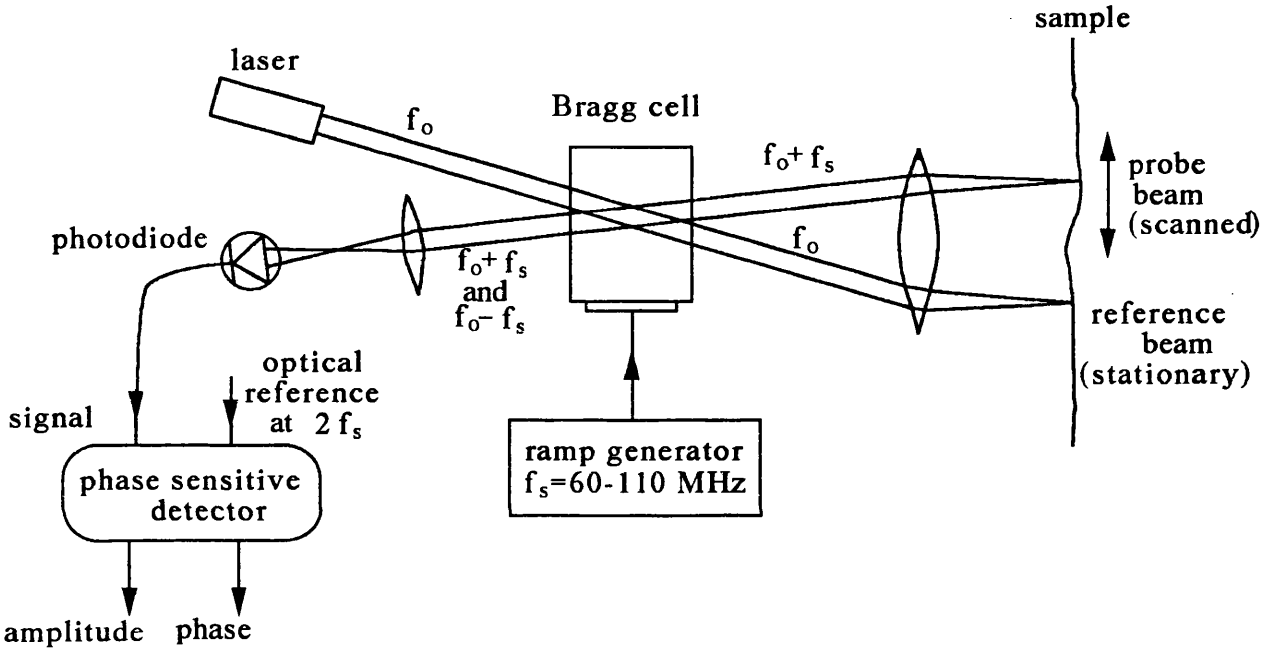


Figure 2-10 The Jungerman scanning phase profilometer

The basic idea for this system is good but certain details make it, as it stands, extremely undesirable in practice. First the accuracy of any measurements are determined by the flatness of the reference area. Large area, high frequency photodiodes are required because the light returning through the Bragg cell moves as the probe beam is scanned. (This may be overcome by instead, detecting the light travelling back in the direction of the laser.) It is also suggested by Jungerman that this system be used as a method for measuring the intensity of the reflected light ie., surface reflectivity. It is unlikely however to give accurate measurements because a Bragg cell used in this manner will cause the intensity of the probe beam to vary as a function of deflection angle.

IV M.J. Offside, M.G. Somekh and C.W. See [26]

This system is shown in figure 2-11. The key elements are an annular quarter wave plate and lens with the central region drilled out. Light from the laser incident on these two components will form a tightly focussed probe and reference beam on the object surface. On reflection, the probe will, after passing back through the annular wave plate, have been rotated in polarization by 90° relative to the reference beam which passes through the wave plate annulus and central region of the lens. The polarizing beam splitter finally directs the reference and probe beams to detectors D1 and D2 respectively. In order to get interference, the light reflected from the object is mixed with a third beam frequency shifted by $\Delta\omega$.

[26]M.J. Offside, M.G. Somekh and C.W. See, 'A scanning heterodyne interferometer with immunity from microphonics', Proc. SPIE conference, 'Surface Characterisation and Testing II', San Diego 7-9 August 1989, Vol. 1164, No. 23 (to be published).

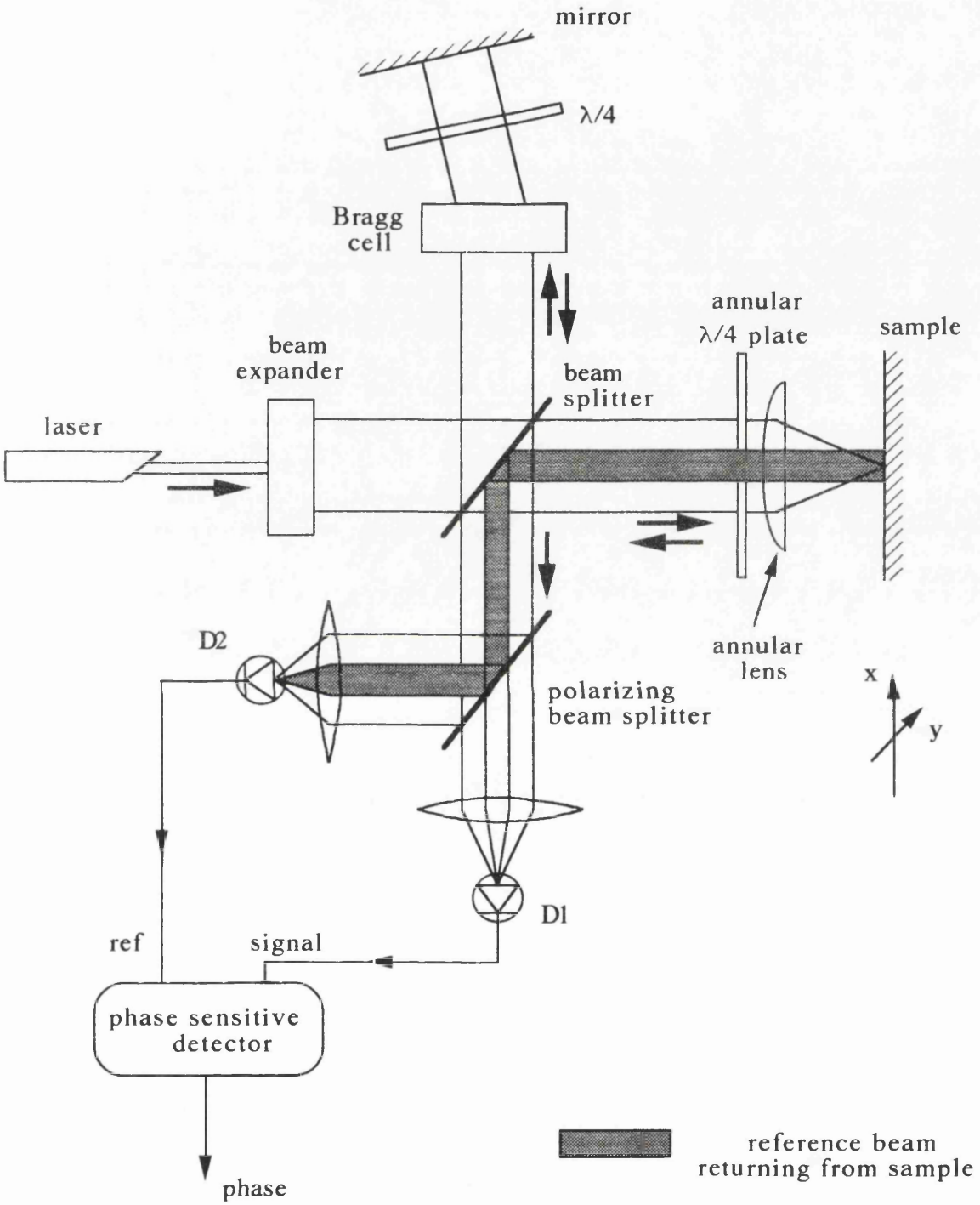


Figure 2-11 The Offside heterodyne phase profilometer

The interference signals from detectors D1 and D2 are,

$$I_1 = A_1 \cos(\Delta\omega t + \theta_1 + \phi) \quad ...(2-5)$$

$$I_2 = A_2 \cos(\Delta\omega t + \theta_r + \phi) \quad ...(2-6)$$

where θ_1 is the phase due to surface features, θ_r is the mean phase of the reference beam, ϕ is phase noise due to microphonics, and A_1 and A_2 are constants. Most important to notice is that the noise due to microphonics in both of the detected signals is the same, and providing the reference beam is large enough in area, θ_r will also remain constant whilst the light is scanned over the specimen. Hence as in the case of Huang's profilometer, even though it is not a true common path system, the phase output after the phase sensitive detection electronics will have the characteristically low noise inherent in a common path interferometer.

The use of three beams in the system allows good interference efficiency which is in contrast to the technique outlined by Huang. The reference beam cannot be made too large as it will degrade the probe beam focus, however this is unlikely to present a real problem. The system has been demonstrated with a ratio of reference to probe beam diameters of 1000. This is a 40 times improvement over Huang's published results. The signal to noise ratio is comparable to that of a true common path system.

Out of the four heterodyne systems so far examined, this clearly has the greatest potential. Care would have to be used in interpreting results from birefringent surfaces.

V L. Laub [27, 28]

This is the last system presented in this section, and the approach taken here bridges the gap between the absolute phase and differential phase techniques presented in the next section. Referring to figure 2–12, a Bragg cell again is used, but this time it is driven so that there are two acoustic waves simultaneously present, at frequencies $f_s + f_m$ and $f_s - f_m$. Two first order beams are generated and these are used as the probes. On returning through the Bragg cell light from each probed area is combined and detected.

The interference response from the detector is,

$$I = A.|r_1|.|r_2| \cos(\Delta\omega_B t + \delta\theta + \phi) \qquad ...(2-7)$$

where $\delta\theta$ is the phase difference imposed on the two reflected beams by object features, $\Delta\omega_B$ is equal to $2f_m$, r_1 and r_2 are the amplitude reflection coefficients of each probed area, and A, ϕ are constants.

The effect of changing the ramp frequency, f_s is to sweep the two probe beams with constant separation across the object surface. The measured phase response, $\delta\theta$ in equation (2 – 4) is therefore the *variation* in phase between the two reflected probe beams, or if the separation of the two beams is very small, the differential phase response. So by integrating the differential phase measurement, the absolute phase can be found.

This heterodyne system offers great versatility combined with a good signal to noise, and also may be applied to the study of birefringent materials.

---

[27] L. Laub, "Apparatus and Methods for Scanning Phase Profilometry", United States Patent No. 3,796,495 , March 12, 1974.

[28] L. Laub, "AC Heterodyne Profilometer", J. Opt. Soc. Am., p.737, Vol. 62, 1973

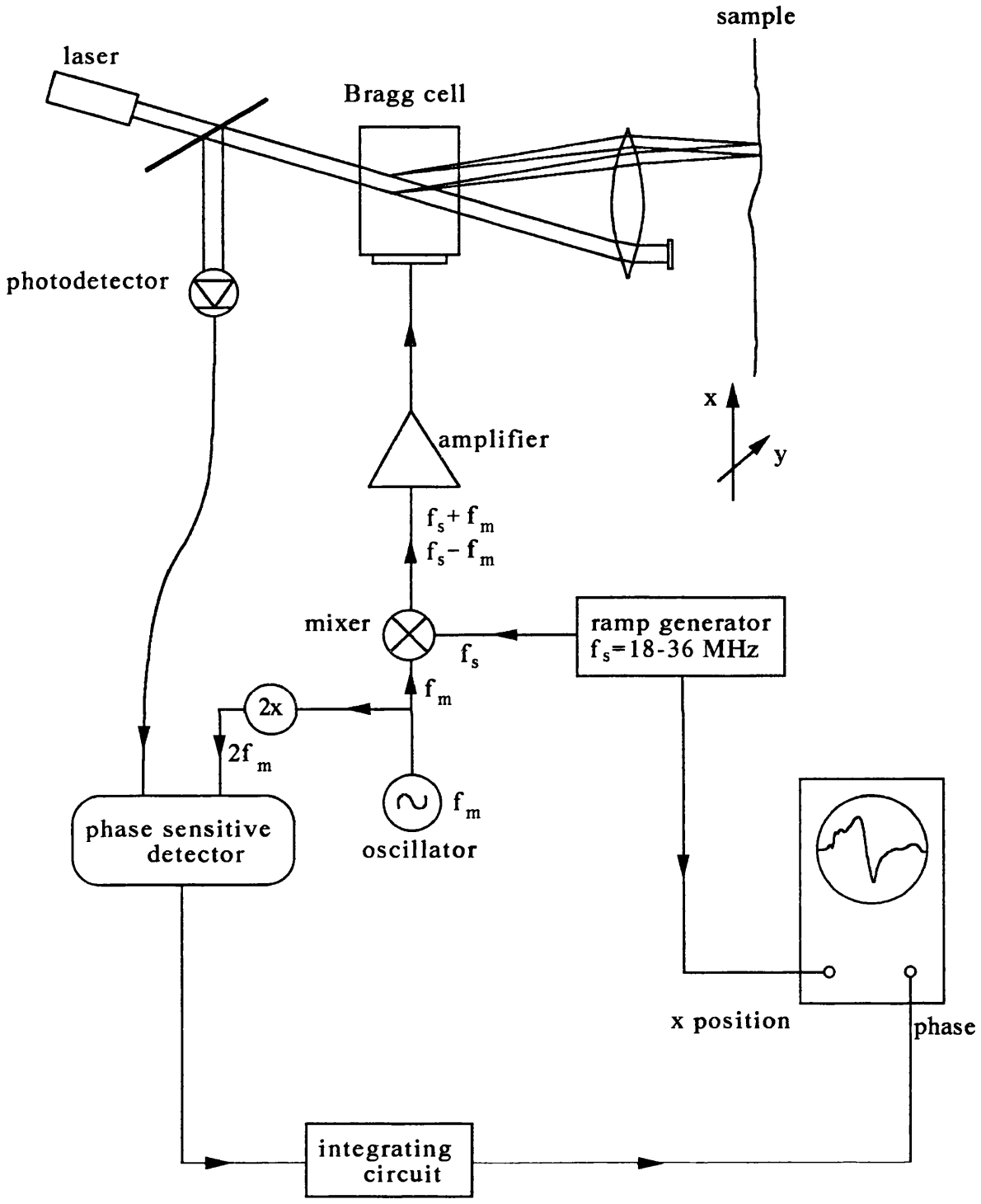


Figure 2-12 The Laub scanning phase profilometer

2.2.4 Scanning interferometric techniques — differential

It has become clear that the major problem for any optical metrology system is the elimination of noise due to microphonics. Mounting the apparatus on a solid foundation in an attempt to vibrationally isolate it is obviously the crudest method, and a common path design is a far more satisfactory solution. The most useful kind of measurement for optical metrology purposes is pure phase, free of reflectivity information. From the discussion in earlier sections, it does appear that some sort of heterodyne arrangement will offer the best chances for a technique with optimum accuracy and sensitivity.

In the previous section, five heterodyne techniques were examined. Of these, probably the last two have the greatest potential. The interferometer described by Laub has the greatest flexibility, as in addition to absolute and differential phase measurements, it can be used for examining birefringent materials. Although not discussed by Laub, this differential response can in some cases be more useful than the absolute phase measurement when it comes to imaging very weak phase structures.

A subject which has only briefly been touched upon is the interpretation of the optical phase. The ideal goal would be if the optical measurements could somehow be transformed directly into a description of surface topography. The only optical techniques which currently approach this are the geometrical methods, but since these are not common path, the poor signal to noise prevents their use for low contrast specimens. The phase measurement deduced from interferometry is governed by both topography and surface conductivity. In the extreme situation of scanning across a topographically smooth interface between conducting and insulating materials, the measured phase will change by π radians. If the phase measurement is to be used to map the surface topography, and variations in surface composition are not known, considerable errors can occur. These problems become more acute when measuring the smoothness of surfaces with roughness figures of nanometers. As was discussed in chapter 1, if the reflectivity

of the surface is monitored at the same time as the phase of the reflected light, then any variation in the reflectivity will at the very least serve as a warning, demanding care in the interpretation of the measured phase in terms of topography.

So in this section we will be looking not only at differential phase techniques, but also at a differential intensity technique which can monitor extremely small changes in surface reflectivity.

I Differential phase and differential intensity contrast using a beam deflector [29]

This system is illustrated in figure 2-13. A beam splitter divides light from the laser into the two interferometer arms. Light in the reference arm is frequency shifted by an amount $2\,\omega_B$. The light in the second arm passes through a standing acoustic wave beam deflector and is focused onto the sample. The action of the beam deflector is to scan the beam back and forth over the surface such that the deflection distance, δx is proportional to $\cos(\omega_s t)$, where ω_s is the acoustic frequency. After reflection from the sample, the light is directed onto the photodiode where interference with light from the reference arm is detected.

There are several frequency components present. If the probe deflection is small such that the phase variation ($\delta\theta$) and reflectivity variation ($\delta r/r_o$) of light at the two extremities of movement is small, then the detected signal from the sidebands at $(2\,\omega_B t \pm \omega_s)$ is,

$$I \propto \sqrt{\left[\frac{\delta r}{r_o}\right]^2 + (\delta\theta)^2} \qquad \ldots(2-8a)$$

A further frequency component which contains information of interest is ω_s, the amplitude of this component being proportional to $\delta r/r_o$. ...(2 - 8 *b*)

The derivation of these responses is given in Appendix A.

[29] C.W. See, M. Vaez-Iravani, and H.K. Wickramasinghe, "Scanning differential phase contrast optical microscope: application to surface studies", Appl. Opt., Vol. 24, No. 15, pp. 2373-2379, 1 August 1985.

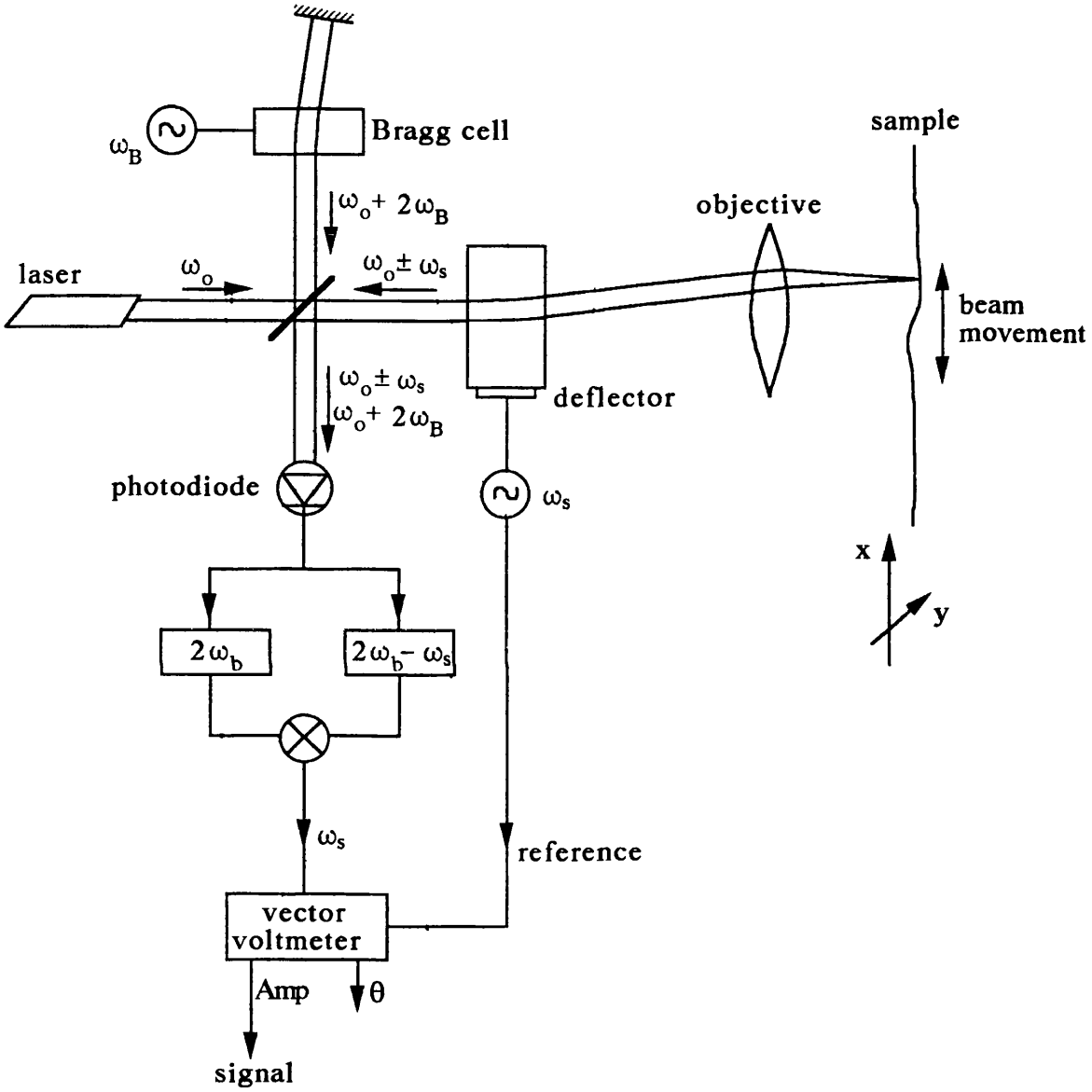


Figure 2-13 Differential contrast optical microscope using beam deflector

Using suitable detection electronics as shown in the figure, the signal corresponding to equation (2 - 8*a*) can be detected. This response, although described as differential phase, is a mixture of phase and reflectivity information which is less desirable than the true differential phase response accessible from Laub's method.

As is clear from equation (2 - 8*b*), by monitoring the ω_s frequency component, the system may be used to measure small variations in object reflectivity. This type of contrast is called differential intensity, and does not rely on any interference phenomenon, and by removing the reference arm, differential intensity can be measured alone[30]. It may be noted here that the term differential amplitude is often used synonymously with differential intensity.

Although not immediately apparent, the measuring of the differential phase signal as shown in figure 2-13 reduces the noise level to that characteristic of a common path system. This is because the $2\omega_B$ and $2\omega_B - \omega_s$ components individually isolated from the photodiode signal both contain common microphonics phase noise, and this is cancelled in the subsequent mixing. A differential intensity measurement made from this system as already mentioned, is not an interference signal and so is not subject to microphonics noise. Further discussion on this subject is in Appendix A.

It has been shown[31] that the noise bandwidth for this type of differential system corresponds to 1.5×10^{-4}nm height variation and 1 part in 10^6 refractive index , in a 1 kHz bandwidth. Such good sensitivities are attributable to the considerable elimination of microphonics noise from the system, and indicate that by increasing the bandwidth to that necessary for imaging, the sensitivities are still sufficient for imaging low contrast features. These sensitivity bandwidth figures are calculated on the basis of a shot noise limited response ie., insignificant noise due to microphonics and Johnson noise.

---

[30]C.W. See, and M. Vaez-Iravani, "Differential amplitude scanning optical microscope: theory and applications", Appl. Opt., Vol. 27,No. 13, pp.2786-2792, 1 July 1988.

[31]M. Vaez Iravani, and C.W. See, "Linear and differential techniques in the scanning optical microscope", Proc. SPIE conference on 'Scanning microscopy technologies and applications', Los Angeles, CA, 1988, Vol. 897, p.43

II Linear differential contrast using Nomarski objective ; M. Vaez Iravani et al. [32]

A scanning implementation of the Nomarski method can be explained with reference to figure 2–14, for the present ignoring the Pockels cell. A Nomarski objective comprises an objective lens and Wollaston prism which is orientated so that polarized light from the laser is divided into two orthogonally polarized beams, focussed onto close areas on the object. Reflected from the object, the two beams are recombined and, after passing through an analyser, the resulting interference is detected. The polarizer and analyser are adjusted so that from a flat area on the object which has no tilt, there is a total extinction of light reaching the detector.

If the two beams experience reflectivities of r and $r+\delta r$ and a phase difference of $\delta\theta$, then in the absence of the Pockels cell, the detected intensity is,

$$I \propto r^2 + (r+\delta r)^2 + 2r(r+\delta r)\cos\delta\theta \qquad ...(2-9)$$

Hence for small phase contrast, the response is proportional to the square of the phase.

In order to get a linear response, the Pockels cell is used to phase modulate the two beams in-phase, with modulation amplitudes of ϕ_1 and ϕ_2 radians respectively. If the modulation frequency is ω_s, then the detected component at ω_s is,

$$I \propto r(r+\delta r)\sin(\delta\theta)\,J_1(\phi_1-\phi_2)\sin\omega_s t \qquad ...(2-10)$$

The depth of modulation of each beam is adjusted so as to maximize $J_1(\phi_1-\phi_2)$. It can be seen that if the phase variation ($\delta\theta$) on the object is small, the resulting intensity measurement is proportional to $\delta\theta$ ie., a linear dependence. However, as in the previous deflector system, the differential phase cannot be measured independent of reflectivity variation.

[32] see reference 31.

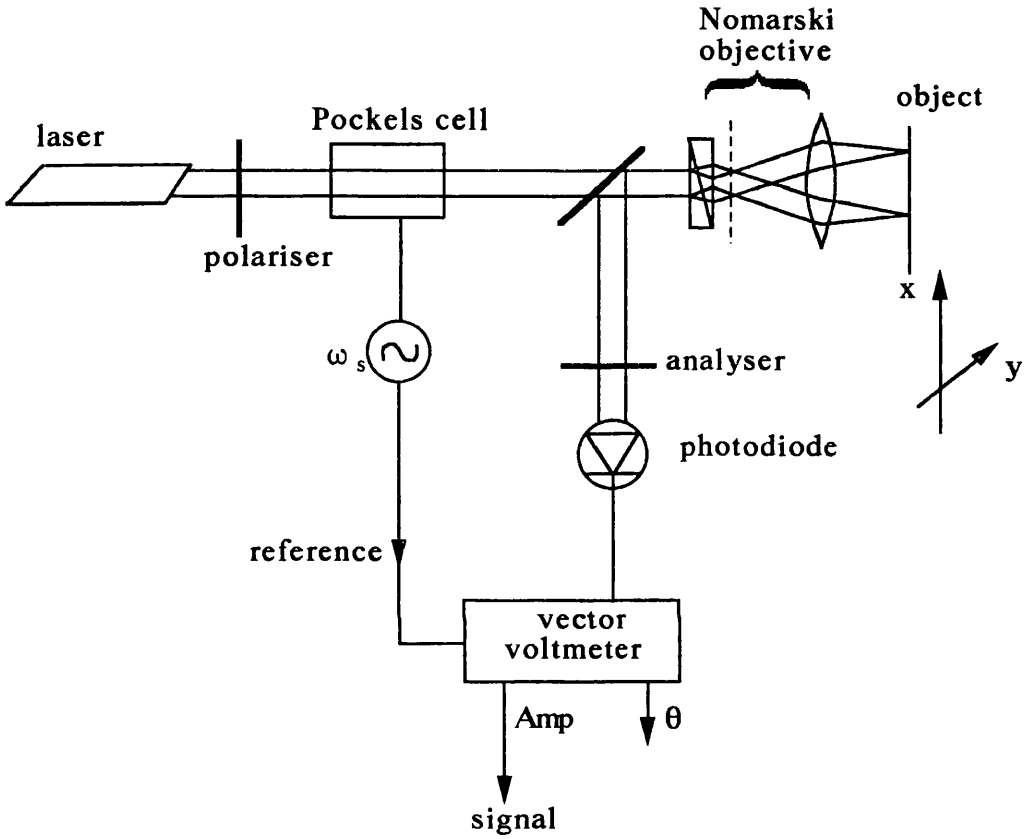


Figure 2-14 Linear differential contrast using Nomarski objective

This technique is common path with the associated low noise. It is possible to measure the phase independent of reflectivity by turning this into a heterodyne differential system, frequency shifting the two polarizations as in Sommargren's system.

III Differential measurements using a split detector ; [33, 34]

This technique takes a completely different approach to all previously outlined methods, and is very satisfying in its innate simplicity. A schematic is shown in figure 2–15. Light is focused by an objective lens onto the object, and the diffracted light is detected by a split detector. A transmission arrangement is illustrated, but the principle is the same as that of a system working in reflection.

If we consider the simple case of an object which is a periodic grating of spatial frequency Ω such that the transmissivity varies as,

$$t(x_0) = \exp\left[\beta\cos\{\Omega(x_0 - x)\}\right] \qquad ...(2-11)$$

where β is complex, and x is the scan coordinate. β is related the magnitude (θ) of the phase variation and absorption (γ) of the grating, such that,

$$\beta = \gamma + i\theta \qquad ...(2-12)$$

In an arrangement of this technique which operates in reflection, γ will relate to the reflection coefficient.

For a weak grating, $|\beta| << 1$, so that we can write,

$$t(x_0) = 1 + \beta\cos\{\Omega(x_0 - x)\} \qquad ...(2-13a)$$

or $$t(x_0) = 1 + \gamma\cos\{\Omega(x_0 - x)\} + i\theta\cos\{\Omega(x_0 - x)\} \qquad ...(2-13b)$$

---

[33]D.K. Hamilton and C.J.R. Sheppard, "Differential phase contrast in scanning optical microscopy", J. of Micrscopy, Vol. 133, Pt. 1, Jan 1984, pp.27-39.

[34]T. Wilson and C. Sheppard, "Theory and practice of scanning optical microscopy", Academic Press, London, 1984.

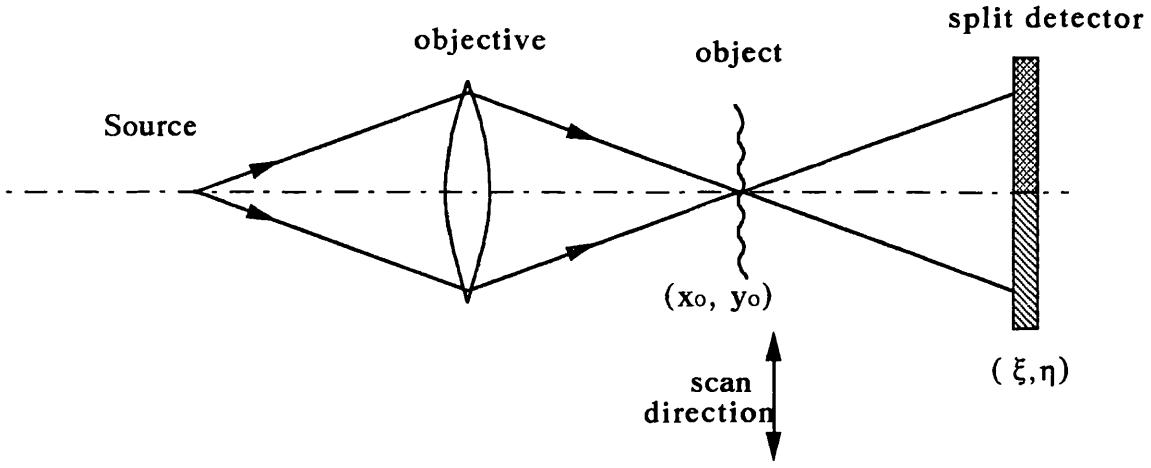


Figure 2–15 Schematic of split detector differential system

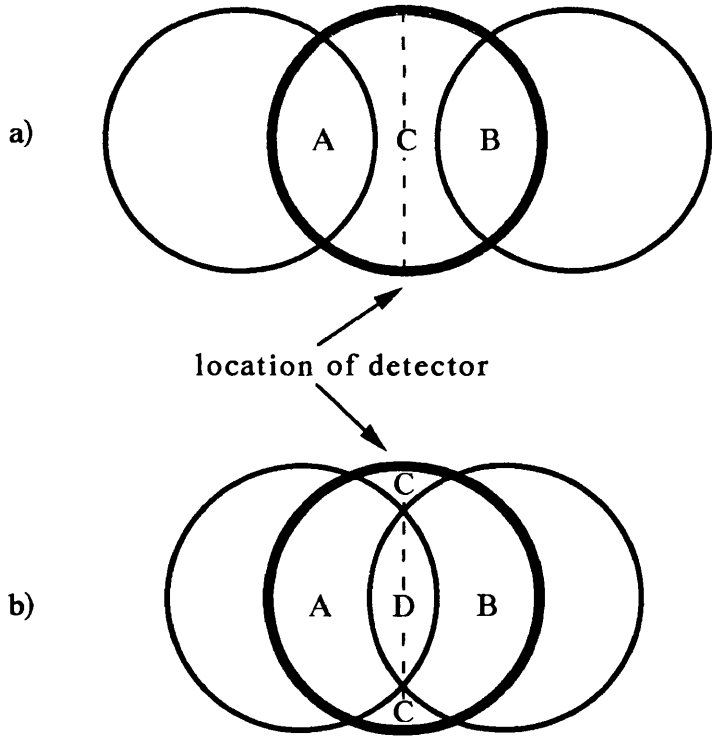


Figure 2–16 diffracted beams at split detector for a weak grating.

The spatial frequency Ω is greater in a) than b).

Using diffraction theory as outlined in reference [35], the field at the detector is,

$$U(\xi, x) = |E_o| \left\{ P(\xi) + \frac{b}{2} P\left(\xi - \frac{\lambda \, d\Omega}{2\pi} \right) \exp(-i\Omega x) + \frac{b}{2} P\left(\xi + \frac{\lambda \, d\Omega}{2\pi} \right) \exp(i\Omega x) \right\} \quad \ldots(2-14)$$

where P is the pupil function of the objective lens.

This field represents three patches of light as depicted in figure 2–16*a*. The first term in equation (2 – 14) is the undeflected zeroth order and the second and third terms are the diffracted plus and minus first orders. For a weak periodic grating, these are the only diffracted orders which exist. Also marked on the figure is the location of the detector which coincides exactly with the central zeroth order beam. The spatial frequency of the grating (Ω) determines the degree of overlap between the different diffracted beams. As Ω decreases, the divergence of the diffracted beams reduces, and all three beams will start to overlap as shown in figure 2–16*b*.

The intensity distribution at the detector is determined by integrating the optical field in equation (2 – 14) :

$$I = |E_o|^2 \int_{detector} \left(\mathrm{P} + \frac{b}{2} \mathrm{P}_+ \, \mathrm{e}^+ + \frac{b}{2} \mathrm{P}_- \, \mathrm{e}^- \right) \left(\mathrm{P} + \frac{b^*}{2} \mathrm{P}_+ \, \mathrm{e}^- + \frac{b^*}{2} \mathrm{P}_- \, \mathrm{e}^+ \right) \, d\xi \quad \ldots(2-15)$$

where the following shorthand is used, P for $P(\xi)$, $P_\pm$ for $P\left(\xi \pm \frac{\lambda \, d\Omega}{2\pi} \right)$, and $e^\pm$ for $\exp(\pm i\Omega x)$.

The detector can be divided into four distinct illuminated areas labelled A to D. Using the definition of the pupil function,

$$P(\xi) = 1 \text{ for } |\xi| < 1,$$
$$= 0 \text{ otherwise,}$$

[35] see reference 34, page 102.

the intensity in each of the detector areas is,

$$I_A = |E_o|^2 (1 + \Re \{b \exp i\Omega x\}) \times \text{(Area A)} \quad ...(2-16)$$

$$I_B = |E_o|^2 (1 + \Re \{b \exp -i\Omega x\}) \times \text{(Area B)} \quad ...(2-17)$$

$$I_C = |E_o|^2 \times \text{(Area C)} \quad ...(2-18)$$

$$I_D = |E_o|^2 \left[1 + \frac{1}{2} |b|^2 \cos 2\Omega x + 2\Re \{b\} \cos \Omega x \right] \times \text{(Area D)} \quad ...(2-19)$$

where (Area A) ... (Area D) are equal to the geometrical overlap of the circular areas as shown on figures 2–16*a,b*.

Taking the situation as shown in figure 2–16*a*) where the spatial frequency of the grating, Ω is large enough that beams form only two separate overlapping areas A and B, if we subtract the responses from each side of the split detector, then the resulting signal is,

$$I_B - I_A = 2 |E_o|^2 \times \text{(Area A)}\ \theta \sin \Omega x \quad ...(2-20)$$

Comparing this with equation (2 – 13 *b*), this can be seen be to a true differential phase response with no absolute or differential intensity dependence (ie., no dependence on γ). It is most important to remember that this result is solely for a weak grating.

Adding the response from areas A and B,

$$I_B + I_A = 2 |E_o|^2 \times \text{(Area A)} (1 + \gamma \cos \Omega x) \quad ...(2-21)$$

this is a non-differential absolute intensity response, characteristic of a non-interference standard scanning optical microscope.

As the spatial frequency Ω of the grating is reduced (figure 2–16*b*), the three beams

gradually coalesce into one large area D. The intensity across this area is evenly spread and is equal to,

$$I_{DD} = |E_o|^2 \times (\text{Area D})\,(1 + 2\gamma \cos \Omega x) \qquad \ldots(2-22)$$

It can be seen that the phase information is lost. As Ω reduces, the illumination on both sides of the split detector becomes equal, and the differential measurement which is the difference in signal from the two sides tends to zero. This demonstrates the high spatial frequency action which is a characteristic of the differential response.

At the other extreme, as Ω becomes large, the plus and minus first orders diverge more and more, until eventually they no longer overlap with the zeroth order beam. This corresponds to the high spatial frequency cut-of point for the system.

It has thus been demonstrated that for a weak grating, the response is true differential phase, independent of absorption (γ) variations. This result can further be shown[36] to be true for an object of arbitrary transmissivity, *providing the magnitude of the variations in θ and γ are small.* In the case of large phase or absorption variations, the response of the system is no longer linear, and the differential phase will depend on a mixture of both θ and γ.

The response of this differential technique to a sinusoidal grating has been discussed in detail. A second structure of interest is a tilted surface ie., a phase gradient. It is shown in reference [37] that the differential phase response to an object of constant slope is,

[36]see reference 34, page 103.

$$I_A - I_B \propto C(\phi/2\pi ; \phi/2\pi) \qquad \ldots(2-23)$$

where ϕ is the phase gradient and C is the partial coherent transfer function. For small ϕ, C is a linear function of ϕ. This is the expected type of differential phase response. However as ϕ increases, the relation is no longer linear, and for large phase gradients, the differential phase response peaks before decreasing back to zero. This can lead to anomalous results as shown in the reference.

The effect of on axis vibrations of the sample (ie., microphonics noise), is to introduce a random phase modulation ($\delta\theta_n$) of the zeroth order relative to the two diffracted orders. The most straightforward way to include this effect is to return to equation (2 - 14) in which the three terms represent respectively the amplitudes of the zeroth, −1 and +1 diffracted orders, and include an extra phase factor in the first term, ie., replace $F(\xi)$ with $F(\xi)$ $\exp(-i\,\delta\theta_n)$. The intensities of the detected areas A and B are then,

$$I_A = |E_o|^2 \,(1 + \Re\,\{b \exp i\,(\Omega x + \delta\theta_n)\}\,)_x \text{ (Area A)} \qquad \ldots(2-24a)$$

$$I_B = |E_o|^2 \,(1 + \Re\,\{b \exp i\,(-\Omega x + \delta\theta_n)\}\,)_x \text{ (Area B)} \qquad \ldots(2-24b)$$

For small noise levels ($\delta\theta_n << 1$), it follows that on subtracting these two responses, the noise terms cancel, and $(I_B - I_A)$ remains the same as that shown in equation (2 - 20). It is therefore clear that the performance of this technique is immune to the effects of small axial vibration. One further point is that the performance of the system relies on accurate balancing of the electronics on each side of the split detector.

[37] see reference 33

IV Simultaneous differential intensity and phase measurement [38, 39, 40, 41]

This system has been developed during the course of the PhD project, and is shown in figure 2–17. In a similar way to Laub, a Bragg cell is driven by two RF frequencies, ω_1 and ω_2, and the two first order beams are used to probe the object. The differential phase response is the same as that given in equation (2 - 7), with $\Delta\omega_B = 2(\omega_1 - \omega_2)$, and is measured using a phase sensitive detector. In addition, each RF frequency is modulated in phase quadrature at frequency, ω_a. This frequency is much lower than ω_1 and ω_2. The result is that the intensity of the two probe beams are modulated in antiphase at frequency $2\omega_a$.

At the detector light from each probed area is detected, and the response at frequency $2\omega_a$ is differential intensity:

$$I = A(R_1 - R_2)\cos 2\omega_a t \qquad \ldots(2-25)$$

where R_1 and R_2 are the intensity reflectivities of the two probed areas, and A is a constant.

---

[38] C.W. See, R.K. Appel, and M.G. Somekh, "Scanning differential optical profilometer for simultaneous measurement of amplitude and phase variation", Appl. Phys. Lett., Vol. 53, No. 1, pp. 10-12, 4 July 1988.

[39] R.K. Appel, C.W. See, and M.G. Somekh, "A scanning differential optical system for simultaneous phase and amplitude measurement", Proc. SPIE conference , 'Scanning Imaging', 21-23 September 1988,Vol. 1028, No. 8.

[40] R.K. Appel, M.G. Somekh and C.W. See, "A scanning differential intensity and phase system for optical metrology", Proc. SPIE conference , 'Surface characterization and testing II', 10–11 August 1989,Vol. 1164, No. 22.

[41] M.G. Somekh and R.K. Appel, "Image formation in common path differential profilometers", Proc. SPIE conference , 'Surface characterization and testing II', 10–11 August 1989,Vol. 1164, No. 16.

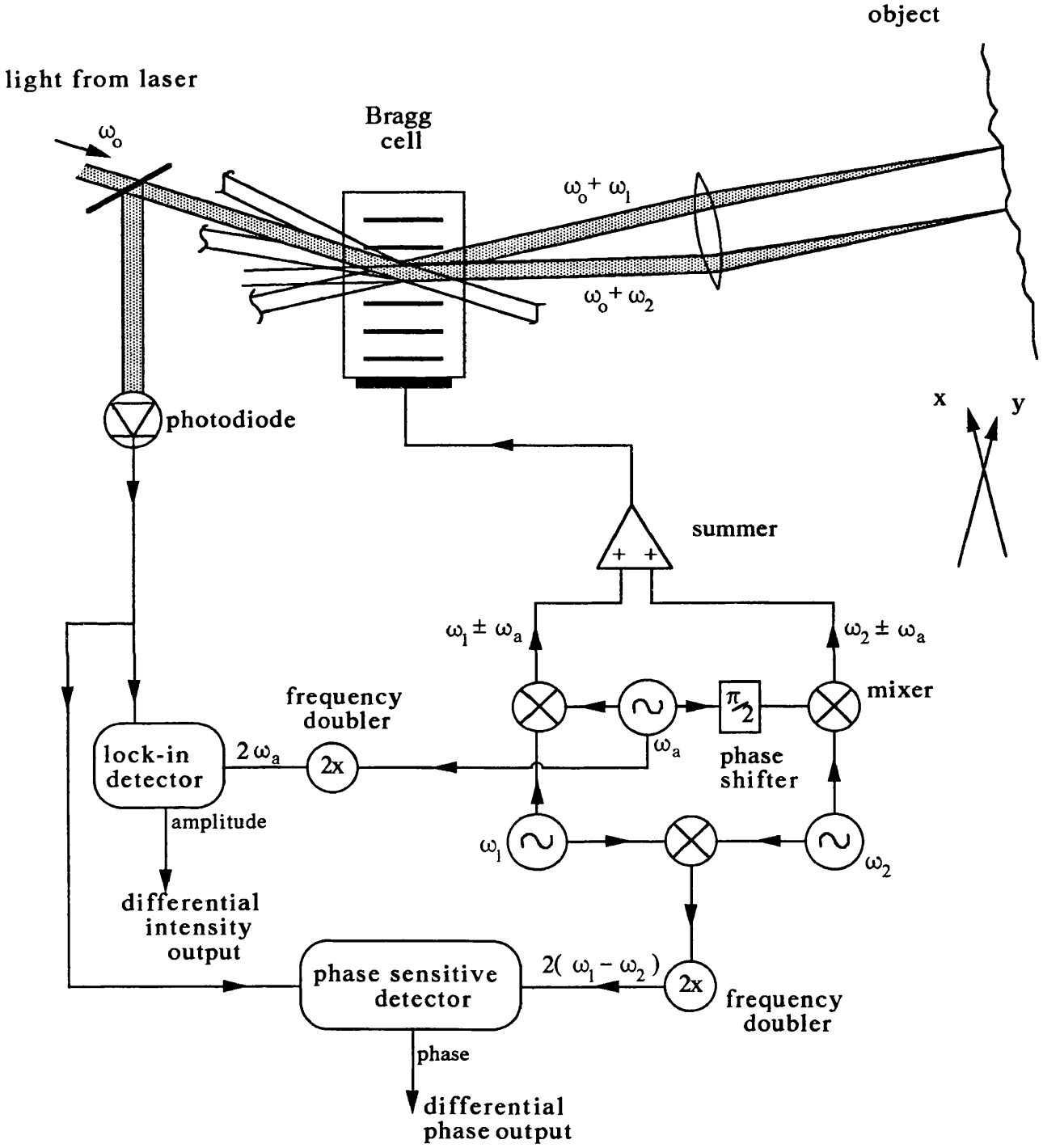


Figure 2-17 The two first order differential phase and intensity profilometer

By varying $\Delta\omega_B$ the separation of the two probing beams may be varied, making the system extremely versatile. For imaging, in order to attain high spatial resolution, the beams must be overlapping, whereas for metrology for example step height measurement, the beams are separated.

The response of the split detector differential techniques previously discussed was shown to be anomalous when measuring a feature of constant slope. In Appendix B it is shown that the equivalent response of this system to such a feature is,

$$I \propto C(\phi/2\pi\,;\,\phi/2\pi)\cos(2\Delta\omega_B t - 2\phi\varepsilon) \qquad \ldots(2-26)$$

where ϕ is the phase gradient, C is the partial coherent transfer function, and ε is the separation of the beams on the object surface. The phase ϕ is measured using phase sensitive detection. What is most important to notice here is that unlike the split detector method, as the tilt of the feature increases, any variations in the partial coherence function have no effect on the absolute value of the differential phase measurement. The sole effect will be on the signal to noise ratio.

2.3 Summary

A broad spectrum of optical techniques have been described which may be applied to metrology, and in particular surface profilometry. By far the most severe problem which must be overcome in any optical technique is microphonics. Further considerations have been that the response should be independent of variations in the intensity of the reflected light, and that the sensitivity bandwidth figure should be as high as possible to enable the technique to be applied to imaging of low contrast features.

The reduction in microphonics is achieved by either making all the light travel through

the same optical elements ie., building a common path interferometer, or by optically generating an electrical reference which has the identical phase noise as a second electrical signal which contains the information. The common phase noise can then be subtracted electrically using phase sensitive detection, and it is believed that the reduction of noise due to microphonics is identical using either method.

The measurement of the optical phase independent of reflectivity variation can be achieved by phase stepping in a homodyne system, or using phase sensitive detection in a heterodyne system. The latter method is very much preferred because it eliminates inaccuracies introduced if the phase step is not exact, and also because of the reduction in the $1/f$ noise.

The signal to noise ratio of the measurement can further be improved by ensuring that there is always good overlap and equal power in the interfering beams.

For measurement of topography, unless a priori knowledge about the composition of the object is available, the phase data from an interferometer is not sufficient. A few methods were outlined which could measure variations in the reflectivity of the surface which can give information concerning composition.

Two techniques in particular, that of Offside et al described in section 2.2.3, and the differential profilometer which is the subject of this thesis, satisfy all the above criteria. The latter technique, having a differential response, can in addition to measuring phase variations, be used in the measurement of extremely small reflectivity variations. Out of all the techniques discussed, it has probably the greatest versatility, being well adapted for metrology or imaging of both homogeneous and birefringent materials.

CHAPTER 3

THE DIFFERENTIAL PHASE AND INTENSITY MEASUREMENT SYSTEM

3. The differential phase and intensity measurement system

This chapter presents two implementations of the differential phase and intensity systems built during the project. In the first section, an overview is given, summarizing the various heterodyne responses. For each setup, there are several positions at which detectors may be located, and the merits of each for imaging and metrology applications are discussed. In the second section, formalism is introduced to describe the frequency heterodyning. Using this framework, expressions are derived which describe the dependence of the differential intensity and phase responses upon the optical modulations performed by the Bragg cell.

3.1 Two implementations of the differential intensity and phase technique : an introduction

In the previous chapter the concept of a differential phase and intensity system was introduced. During the course of the PhD project, two experimental realisations of the technique have been built and results taken. The first implementation is shown schematically in figure 3-1, and it can be seen that the essential feature is a Bragg cell which divides light into a zeroth and first order beam which are then used to interrogate the object. In the current implementation (figure 3-2) a Bragg cell again is used, but this time it is driven with two different RF frequencies generating two first orders beams, and it is these which are used to probe the object. In all further discussion in this thesis, these two systems will be called the zeroth/ first order and two first order systems respectively.

The analysis of both implementations has many similarities. Collimated light entering the Bragg cell is split into two probing beams, κ_1 and κ_2, whose angular separation is proportional to the frequency difference ($\Delta\omega$) between the Bragg cell input signals. In the case of the zeroth/ first order system, $\Delta\omega$ is equal to the single frequency of the

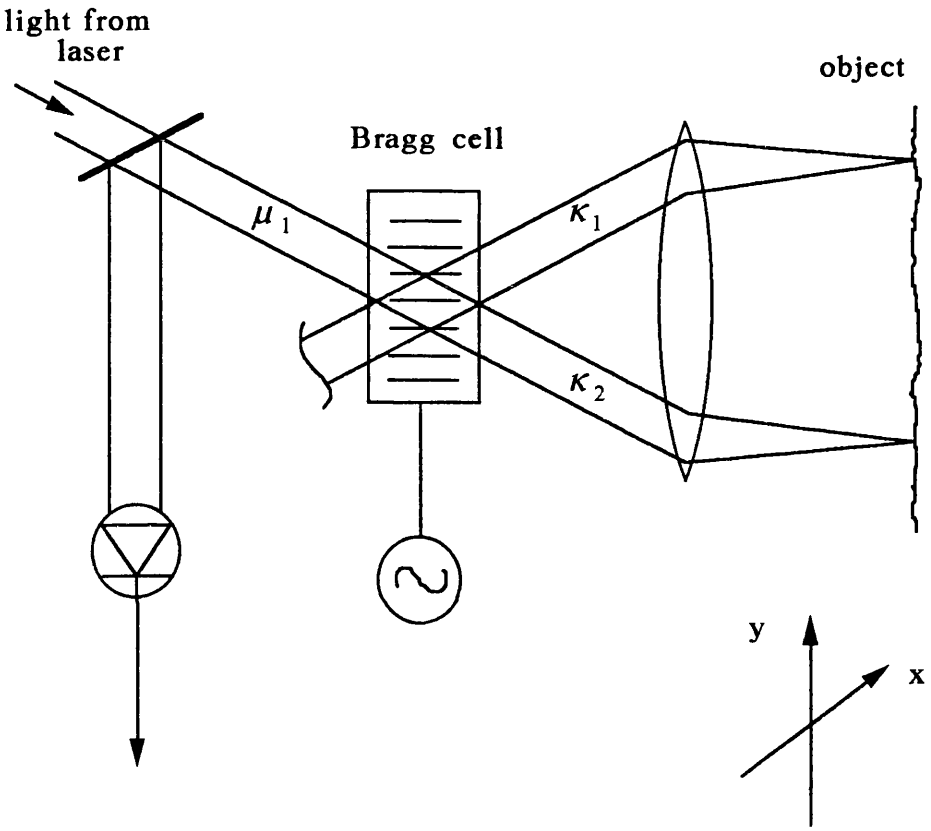


Figure 3-1 The zeroth/ first order system

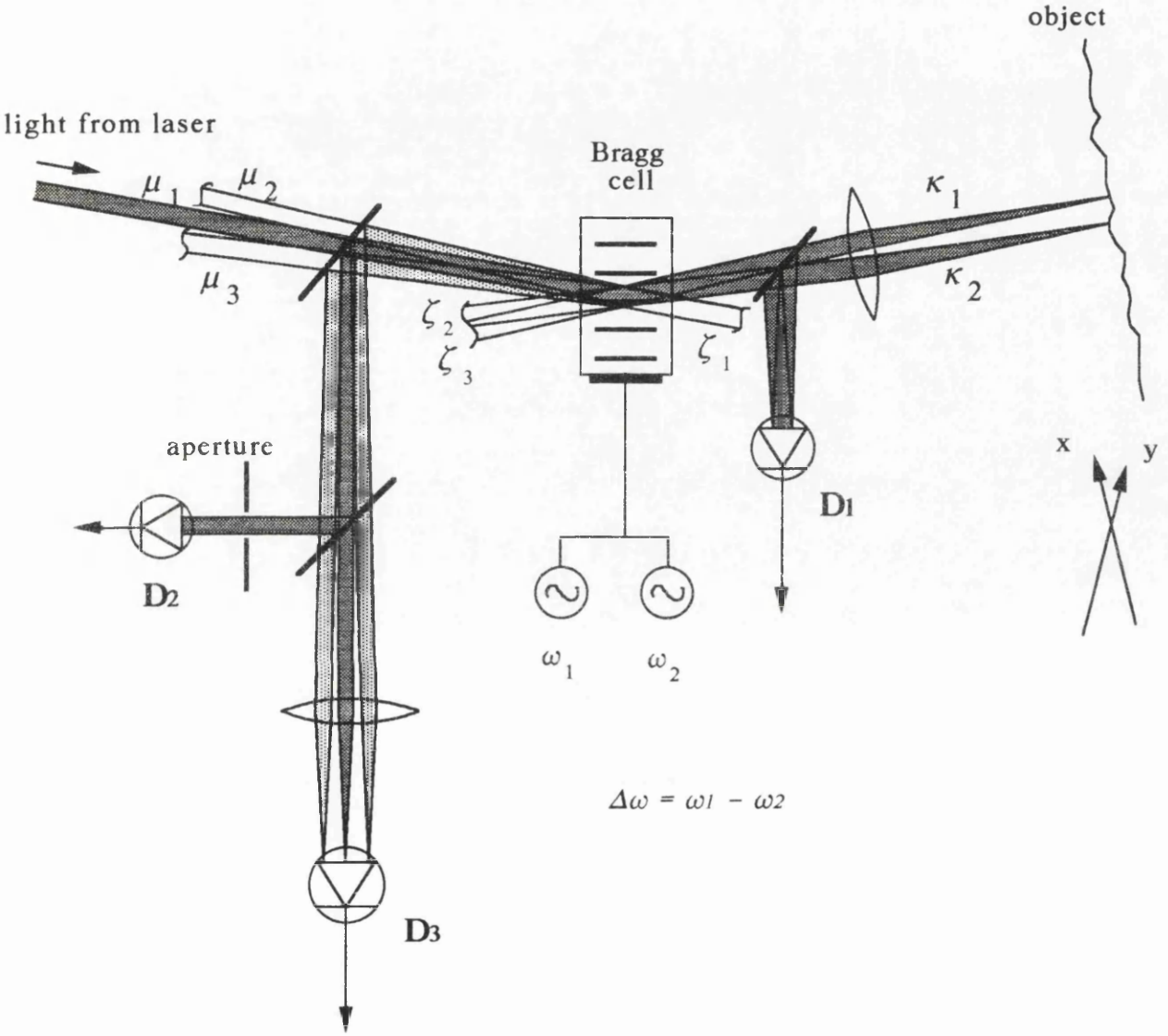


Figure 3-2 The two first order system

electrical drive. For the two first order system, the electrical drive comprises two frequencies, ω_1 and ω_2, and $\Delta\omega$ is equal to ω_1 - ω_2. The objective lens is positioned so that it is one focal length from the point at which the beams appear to diverge and one focal length from the sample. This results in the two beams coming to a focus at the object with their optic axes parallel and normal to the plane of the object ie., telecentrically. The operation of the lens is to transform the divergence of the two beams at the Bragg cell so that the separation of the two interrogating spots is proportional to the difference frequency $\Delta\omega$. Upon reflection, the two beams return through the Bragg cell where they recombine and are detected at the photodiode D2.

Despite this single detector being in principle all that is needed for both differential phase and differential intensity measurements, there are advantages to be gained in positioning detectors at two other positions. Immediately after the sample, detector D1 monitors the light before a second acousto-optic interaction within the Bragg cell. It will become clear as the analysis progresses that this detector is best for differential intensity measurements. Detector D3, which appears only in the two first order system, responds to all three first order beams which are present after the second passage through the Bragg cell. Even though it is the central first beam, μ_1, which holds all the desired information for differential phase and intensity measurement, it is not always possible to spatially isolate this from the two side first orders, μ_2 and μ_3. Analysis of the two first order system must therefore consider the consequences of these two side orders.

It is not anticipated that both detectors D2 and D3 are used simultaneously. Indeed in an experimental setup, D2 and D3 would be replaced by a single detector which, depending on the separation of μ_1 , μ_2 and μ_3 , fulfills the role of either detector.

One essential consequence of the Bragg interaction is that light incident at the Bragg angle is diffracted through an angle which is in direct proportional to the Bragg cell drive frequency. So in the zeroth/ first order system, varying the acoustic drive frequency will

alter the spatial separation of the focussed probe beams on the object. The separation required depends upon the particular application. For differential imaging optimum spatial resolution is required and so the two beams should form two overlapping focussed spots on the object surface. The converse is true for metrology applications such as that of step height measurement where it is definitely not desirable to have any light in the vicinity of the step transition region.

In practice, the zeroth/ first order system has difficulties in satisfying the requirements for imaging. It is easy to form two well separated focussed beams on the object. To decrease this separation, the Bragg cell drive frequency is reduced, but in order to maintain a proper Bragg type interaction, the length of the Bragg cell must be proportionally *increased.* This defines a physical limit on the smallest ratio of focussed beam separation to spot size and it is consequently difficult to reduce the ratio much below[1] about 3 to 1.

The two first order system enables the two beams to be varied from totally overlapping to well separated, giving far greater flexibility in use. Clearly neither system could operate with completely overlapping beams because the differential signal levels would be constantly zero. In practice, the smallest required ratio of beam separation to spot size is likely to be of the order 0.1 . The maximum separation is limited by the acoustic bandwidth of the transducer on the Bragg cell, but in experimental results presented in a later chapter, it is shown that the two first order system can achieve a ratio of 8 to 1 with an 80 MHz Bragg cell. This figure is probably adequate for most purposes but may nevertheless be increased by using for example a Bragg cell with a higher centre frequency.

---

[1] The ratio of focussed beam separation to spot size is equal to $\frac{\alpha d}{1.22\lambda}$, where α is the angular divergence of the beams after the Bragg cell, d is the diameter of each beam, and λ is the optical wavelength. For d= 0.8mm, λ = 632.8nm, and α = 3mrad ($PbMO_4$ Bragg cell at 20 MHz), this ratio is equal to 3.

We will now consider the various measurements which can be made, starting with the differential phase response. This is equal to the phase difference of the two reflected beams and is measured using interference. The beams incident on the object can be thought of as the two arms of an interferometer, where instead of the familiar Michelson geometry, both arms have been bent around so as to be nearly parallel. An essential property of the Bragg cell is that the scattered light is frequency shifted in proportion to the scattering angle. In the zeroth/ first order system (figure 3–1), the beam κ_1 is frequency shifted by $\Delta\omega$ relative to κ_0 which is the unshifted zeroth order beam. Passing back through the Bragg cell, some light from κ_0 and κ_1 is recombined into μ_1 , and results in interference at detector D2. This second Bragg interaction results in a total $2\Delta\omega$ frequency shift, causing the interference signal to be at this 'beating' frequency.

In the case of the two first order system, the two beams (κ_1, κ_2) incident on the object (see figure 3–2) have associated frequency shifts of ω_1 for κ_1 and ω_2 for κ_2 and so a total frequency difference of $\Delta\omega = \omega_1 - \omega_2$. Passing back through the Bragg cell three first order beams (μ_1, μ_2 and μ_3) are generated. The central first order (μ_1), going in the direction of the incoming light from the laser is composed of light from κ_1, now at frequency $2\omega_1$ and from κ_2 at $2\omega_2$. The effect of both beams coincident at detectors D2 or D3 is an interference signal at $2(\omega_2 - \omega_1)$ ie., $2\Delta\omega$. The outer two first orders, μ_2 (top) and μ_3 (bottom) comprise light from κ_2 and κ_1 respectively, and are caused by the beams interacting with alternate acoustic waves on each pass through the Bragg cell. These two beams have the same total frequency shift of $(\omega_1 + \omega_2)$.

It can be seen that for detectors D2 and D3 in both systems, the interference signal carrying the differential phase information is at the frequency $2\Delta\omega$. Each beam reflected from the object has an associated phase θ_1 and θ_2, and the differential phase, $\delta\theta = \theta_1 - \theta_2$ is monitored by detecting the phase of the interference signal. ie., the

response of detectors D2 and D3 is,

$$I_{diff\,phase} \propto \cos(2\Delta\omega t + \delta\theta) \qquad ...(3-1)$$

A line scan or an image is then be formed by moving the light relative to the sample. The phase detection of the electrical signal (3 - 1) can be measured using phase sensitive detection electronics in conjunction with a reference of constant phase derived from the Bragg cell drive.

Detector D3 detects, in addition to μ_1 which contains the desired differential phase information, μ_2 and μ_3. In metrology mode μ_2 and μ_3 will be spatially quite distinct and can indeed be removed, thus allowing detector D2 to be used for differential phase measurement. However even if the differential phase measurement were made from D3, these two side beams will only contribute to a dc background current. This current in no way affects the relative accuracy of the differential phase measurement, but will contribute to noise, eg. shot noise.

When operating in imaging mode, the angular separation of μ_1, μ_2 and μ_3 will be of the same order or smaller than the natural spread within each beam due to diffraction. Consequently the detection of μ_1 is not possible without some light from μ_2 and μ_3. This is still no real problem because there is no combination of any of the three beams which will contribute to an interference signal at $2\Delta\omega$, so the differential phase measurement is unaffected. The consequence of this is important for any practical implementation of the technique.

Differential intensity is the measurement of the difference in reflected light from the two probed areas. This is achieved by modulating the Bragg cell drive so that the intensity of light incident on the object is modulated in anti-phase at frequency ω_m. If the reflectivity of the two probed areas is different, then the total reflected light at detectors D1 or D2 will contain some ω_m frequency component. Conversely, if the reflectivity is

the same, no signal is measured at this frequency : the measurement is effectively dark field. In this way the system can be seen to respond to *changes* in reflectivity. The response can be expressed mathematically as,

$$I_{diff\ intensity} \propto (R_1 - R_2)\cos(\omega_m t) \qquad ...(3-2)$$

where R_1, R_2 are the intensity reflection coefficients of the probed areas.

As already mentioned, in imaging mode it is not possible to isolate μ_1 in the complete absence of μ_2 and μ_3, so detector D2 in the two first order system cannot be used. Depending on the type of modulation used, the accuracy of differential intensity measurements made using D3 may have a considerable relative error, so making it preferable to do all differential intensity measurements from detector D1.

Two further measurements are available from both systems at a frequency above dc. Absolute intensity at frequency $2\omega_m$ relates to the average intensity reflection coefficient R of the two probed areas :

$$I_{abs\ intensity} \propto R\cos(2\omega_m t) \qquad ...(3-3)$$

This response necessitates two passes through the Bragg cell so excluding measurement from detector D1.

Second, the intensity of the interference at frequency $2\Delta\omega$ is proportional to the product of the amplitude reflection coefficients $\underline{r}_1$, $\underline{r}_2$ of the two probed areas :

$$I_{interference} \propto |\underline{r}_1|\,|\underline{r}_2|\cos(2\Delta\omega t + \delta\theta) \qquad ...(3-4)$$

Equation (3 - 4) is a more explicit description of the interference signal, first described in equation (3 - 1).

It is also possible to get an interference response from detector D1, and this will be the same as equation (3 - 4) but at frequency $\Delta\omega$. There are however a number of factors making this detector less attractive for differential phase measurement. First it is clear that the orientation of the detector will have some bearing on the interaction between the two beams whereas for detectors D2 and D3, the interference is contained *within* one beam (μ_1). Furthermore it will be shown in section 3.5 that if the intensities of κ_1 and κ_2 are modulated in exact anti-phase, the interference response at D1 has no $\Delta\omega$ frequency component : the optical power is in sidebands around this frequency.

In both systems, there are a number of diffracted beams which are blocked. The only beam which presents a conceivable problem is the zeroth order (ρ) in the two first order system. Enough distance is required between the Bragg cell and the following lens in order to block this light and, even for a lens of moderate NA, this distance becomes physically too small. This can be easily solved by using two further lenses to project the point of divergence of the light within the Bragg cell further down the system, and positioning the objective lens one focal length from this point. The zeroth order can then be blocked en route.

3.2 Analysis of frequency heterodyning

It was seen in the first section that the Bragg cell performs several important functions. It divides incoming light into two angularly separated beams which have a relative frequency shift and it enables the intensities to be modulated in anti-phase. Many questions remain to be answered such as,

i) the types of modulation that can be used ;

ii) the way in which different modulations affect the signal to noise ratio of both the differential phase and differential intensity measurements ;

iii) the effect on the differential intensity and phase responses if the intensity modulations are not in exact anti-phase ;

iv) It has been said in the previous section that detector D3 may not always be used for accurate differential intensity measurement . We require to know the factors that determine the accuracy of measurement from this detector.

This section is concerned in building up a mathematical framework that can answer such questions. Expressions are derived which relate the differential responses to arbitrary modulation functions. These functions describe the optical modulation caused by an interaction between one particular light beam and acoustic wave in the Bragg cell. It is left until the next chapter to relate these modulation functions to the *electrical* modulations of the RF drives which generate the acoustic waves. This analysis, although done for the two first order arrangement (figure 3–2), may with minor adjustment be adapted for the zeroth/ first order Bragg cell system.

The Bragg cell is driven by two independently modulated RF frequencies, ω_1 and ω_2 such that on a single pass, a beam diffracted into a first order is amplitude modulated by $g_1(t)$ or $g_2(t)$, depending upon with which acoustic wave the interaction occurred. These two optical modulation functions are constrained by the relation,

$$|g_1(t)|^2 + |g_2(t)|^2 \leq 1 . \qquad ...(3-5)$$

For the present, $g_1(t)$ and $g_2(t)$ are assumed to be arbitrary real periodic functions ranging between 0 and 1, but as the analysis progresses a number of necessary conditions will be formulated which must be satisfied if true differential phase and intensity measurements are to be made[2]. *Throughout this chapter, it is most important to remember that the g_1 and g_2 functions refer to optical modulation, and that it is left to a later chapter to discuss the required electrical modulation to realise these functions.*

[2]If the differential responses are 'true' then the differential phase and intensity responses are given according to equations (3 – 1) and (3 – 2).

$g_1(t)$ and $g_2(t)$ are defined assuming that the light is incident at the Bragg angle. Deviations from the Bragg angle lower the diffraction efficiency, which is equivalent to reducing $g_1(t)$ or $g_2(t)$ by some constant factor. This effect is absorbed by three sets of constants to cover all the interactions of interest. A_1 and A_2 describe the splitting of the two first order beams immediately after passing through the Bragg cell for the first time ; a_1 and a_2 account for the contributions to μ_1 from each beam returning from the sample, and b_1, b_2 are for the returning beams κ_1, κ_2 diffracted into μ_3, μ_2 respectively. All three pairs of coefficients range between 1 for exact Bragg incidence, to zero for incidence so far off Bragg that no light is diffracted.

It will be seen that a general requirement is that the coefficient pairs, A_1, A_2 and a_1, a_2 are maximized. Providing the frequency difference of the RF drives is not too large, it can be expected that these terms will equal a little less than unity.

We will now look at the effects on the light of the various interactions throughout the system. The light from the laser can be written as,

$$E_i = E_0 \exp(i\omega_0 t) \qquad ...(3-6)$$

where E_0 and ω_0 are the amplitude and frequency of the light entering the system. This is related to the power, P_L of the emitted laser light such that,

$$P_L = \tfrac{1}{2} c \varepsilon |E_o^2| \qquad ...(3-7)$$

where c is the speed of light and ε is the permittivity of air.

Passing through the Bragg cell for the first time, the light is split into two first orders (κ_1 and κ_2) and one zeroth, which is discarded. The two first order beams are modulated as follows,

$$E_1 = E_0 A_1 g_1(t).\exp\left(i[(\omega_0 + \omega_1)t + \phi_1]\right) \quad ...(3-8a)$$

$$E_2 = E_0 A_2 g_2(t).\exp\left(i[(\omega_0 + \omega_2)t + \phi_2]\right) \quad ...(3-8b)$$

where ϕ_1 and ϕ_2 are phase constants.

At the sample, κ_1 and κ_2 are affected by the complex reflection coefficient and topographic detail at the two probed areas ($\underline{r}_1$, $\underline{r}_2$, θ_1, and θ_2), and the resulting amplitude of the reflected light is,

$$E_1' = \underline{r}_1 \exp(2\theta_1 i) \cdot E_1 \quad ...(3-9a)$$

and

$$E_2' = \underline{r}_2 \exp(2\theta_2 i) \cdot E_2 \quad ...(3-9b)$$

The complex reflection coefficient can be expressed in terms of the modulus and phase angle of the reflection coefficient:

$$\underline{r}_1 = r_1 \exp(i\alpha_1) \quad ...(3-10a)$$

and

$$\underline{r}_2 = r_2 \exp(i\alpha_2) \quad ...(3-10b)$$

The detectors are photodiodes and are modelled by assuming that each photon in the incident optical beam (power P_o) will causes on average η electrons to move into the conduction band, where η is the quantum efficiency of the process. The resulting photocurrent is,

$$I_s = \frac{\eta e P_0}{h\nu} \quad ...(3-11)$$

where e is the electronic charge, and $h\nu$ is the photon energy.

A proportion (ρ_1) of the two returning beams is directed onto two overlapping areas, on detector D1. The current response of detector D1 is,

$$I_{d_1} = \frac{1}{2} c\varepsilon\rho_1 \zeta \int_{detector\ D1} (\chi_1 E_1' + \chi_2 E_2') . (\chi_1 E_1' + \chi_2 E_2')^* \ dy \qquad ...(3-12a)$$

where ρ_1, (as well as ρ_2, ρ_3 in the following equations), are constants ranging between 0 and 1, determined by the splitting ratios of the beam-splitters; ζ contains previous constants,

$$\zeta = \eta e/h\nu \ , \qquad ...(3-12b)$$

and where χ_1 , χ_2 are phase factors determined by the orientation β_1 , β_2 of each beam with respect to the optic axis, such that,

$$\chi_1 = \exp\left(-i\left[k_{y1} y + k_{z1} z\right]\right) \qquad ...(3-12c)$$

and

$$\chi_2 = \exp\left(-i\left[k_{y2} y + k_{z2} z\right]\right) \qquad ...(3-12d)$$

with

$$k_{y1} = -|\boldsymbol{k}| \sin\beta_1 \ , \qquad k_{y2} = |\boldsymbol{k}| \sin\beta_2 \qquad ...(3-12e, f)$$

$$k_{z1} = |\boldsymbol{k}| \cos\beta_1 \ , \qquad k_{z2} = |\boldsymbol{k}| \cos\beta_2 \ . \qquad ...(3-12g, h)$$

$|\boldsymbol{k}|$ $(= \frac{2\pi}{\lambda})$ is the wave vector of each beam with components k_{y1} , k_{y2} normal and k_{z1}, k_{z2} parallel to the optical axis. The convergence angle of the two beams is equal to $\beta_1 + \beta_2$, and it is assumed that the two beams overlap completely on the detector surface. The integration is necessary because the two beams are incident with different orientations, resulting in interference fringes across the face of the detector. This is unimportant for measurements which rely on non-interference phenomena, providing *all* the incident light is collected by the detector. The detector will then automatically perform the integration.

Light measured by detectors D2 and D3 interacts in the Bragg cell for a second time. Detector D2 responds to beam μ_1 which comprises light from κ_1 and κ_2, and the wavefronts of these two constituent beams are totally overlapping. This means that the use of a lens in front of the detector will have no effect on any measurements performed, which is clearly not true for the other two detectors. The output from detector D2 is,

$$I_{d_2} = \frac{1}{2} c\varepsilon\rho_2 \zeta (E_a + E_b) . (E_a + E_b)^* \qquad ...(3-13a)$$

where $$E_a = a_1 g_1 (t) \exp(i\omega_1 t) . E_1' \qquad ...(3-13b)$$

and $$E_b = a_2 g_2 (t) \exp(i\omega_2 t) . E_2' \qquad ...(3-13c)$$

Detector D3 responds to the light from all three beams. It was shown in the previous section that the differential phase information is contained in the $2(\omega_2 - \omega_1)$ interference signal, which is only *within* the central beam, μ_1. Moreover, differential intensity does not rely on interference phenomena. The way in which the three beams form on the detector D3 is therefore immaterial for both differential intensity and phase.

So ignoring any interference which may involve the side beams, the signal from detector D3 is,

$$I_{d_3} = \frac{1}{2} c\varepsilon\rho_3 \zeta \{(E_a + E_b) . (E_a + E_b)^* + |E_{\mu_2}|^2 + |E_{\mu_3}|^2\} \qquad ...(3-14a)$$

where $$E_{\mu_2} = b_2 g_1 (t) \exp(i\omega_1 t) . E_2' \qquad ...(3-14b)$$

and $$E_{\mu_3} = b_1 g_2 (t) \exp(i\omega_2 t) . E_1' \qquad ...(3-14c)$$

In the next two sections, explicit forms for the differential intensity and phase responses from each detector will be found.

3.3 The differential and absolute intensity responses

The differential and absolute intensity responses do not rely on any interference phenomena. Therefore the cross terms in equations (3 - 12) to (3 - 14) can be ignored.

Using the Fourier transform pair,

$$S(\omega) \stackrel{\text{df}}{=} \int_{-\infty}^{\infty} I(t) \exp - i\omega t \ \mathrm{dt} \qquad \ldots(3-15a)$$

and

$$I(t) \stackrel{\text{df}}{=} \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) \exp i\omega t \ \mathrm{d}\omega \qquad \ldots(3-15b)$$

and also equation (3 - 7), the spectrum of the intensity response from each detector is,

detector D1:

$$S_1(\omega) = \frac{1}{2} c\varepsilon\rho_1 \zeta \int_{-\infty}^{\infty} \int_{\text{detector D1}} \left[|\chi_1 E_1'|^2 + |\chi_2 E_2'|^2 \right] \mathrm{dy} \ \exp - i\omega t \ \mathrm{dt}$$

ie.,

$$S_1(\omega) = \rho_1 R_L \zeta \int_{-\infty}^{\infty} \left[r_1^2 A_1^2 g_1^2(t) + r_2^2 A_2^2 g_2^2(t) \right] \exp - i\omega t \ \mathrm{dt} \qquad \ldots(3-16)$$

detector D2:

$$S_2(\omega) = \rho_2 R_L \zeta \int_{-\infty}^{\infty} \left[r_1^2 a_1^2 A_1^2 g_1^4(t) + r_2^2 a_2^2 A_2^2 g_2^4(t) \right] \exp - i\omega t \ \mathrm{dt} \qquad \ldots(3-17)$$

detector D3:

$$S_3(\omega) = \rho_3 P_L \zeta \int_{-\infty}^{\infty} \left[r_1^2 a_1^2 A_1^2 g_1^4(t) + r_2^2 a_2^2 A_2^2 g_2^4(t) \right] \exp - i\omega t \; dt$$

$$+ \rho_3 P_L \zeta \int_{-\infty}^{\infty} g_1^2(t)\, g_2^2(t) \left[r_1^2 b_1^2 A_1^2 + r_2^2 b_2^2 A_2^2 \right] \exp - i\omega t \; dt \qquad \ldots(3-18)$$

The modulus of both phase factors, χ_1 , χ_2 are equal to unity, and it can be seen that providing all the light is collected by detector D1, these do not appear in equation (3 - 16). Hence the intensity frequency spectrum from this detector is independent of the propagation direction of each beam ie., β_1 , β_2 . Although the phase factors clearly have no effect on the differential intensity response, they are particularly relevant when considering the differential phase response of this detector, in section 3.4.2 .

Equations (3–16), (3–17) and (3–18) are simplified by expressing g_1 and g_2 in the frequency domain, using the functions $G_1(\omega)$ and $G_2(\omega)$ which are defined as follows:

$$G_1(\omega) \stackrel{\text{df}}{=} \int_{-\infty}^{\infty} g_1(t) \exp - i\omega t \; dt \qquad \ldots(3-19)$$

and

$$G_2(\omega) \stackrel{\text{df}}{=} \int_{-\infty}^{\infty} g_2(t) \exp - i\omega t \; dt \qquad \ldots(3-20)$$

Since $g_1(t)$ and $g_2(t)$ are real and periodic, $G_1(\omega)$ and $G_2(\omega)$ comprise a series of discrete frequencies so that,

$$G_1(\omega) = \frac{1}{2} \sum_{n=-\infty}^{\infty} c_n \, \delta(\omega - n\omega_a) \qquad \ldots(3-21)$$

and

$$G_2(\omega) = \frac{1}{2} \sum_{n=-\infty}^{\infty} d_n \, \delta(\omega - n\omega_a) \qquad \ldots(3-22)$$

where $d_{-n} = d_n^*$, $c_{-n} = c_n^*$ ...(3 – 23)

Using equations (3 – 16) to (3 – 23), the responses from each detector, in terms of discrete frequency components are,

detector D1:

$$S_1 = \frac{1}{4}\rho_1 P_L \zeta \sum_{m,n=-\infty}^{\infty} \{ r_1^2 A_1^2 c_m c_n + r_2^2 A_2^2 d_m d_n \} \delta(\omega - (m+n)\omega_a)$$

...(3 – 24)

detector D2:

$$S_2 = \frac{1}{16}\rho_2 P_L \zeta \sum_{m,n,q,r=-\infty}^{\infty} \left\{ r_1^2 a_1^2 A_1^2 c_m c_n c_q c_r + r_2^2 a_2^2 A_2^2 d_m d_n d_q d_r \right\}$$

$$.\delta(\omega - (m+n+q+r)\omega_a)$$

...(3 – 25)

detector D3:

$$S_3 = \frac{1}{16}\rho_3 P_L \zeta \sum_{m,n,q,r=-\infty}^{\infty} \left\{ r_1^2 a_1^2 A_1^2 c_m c_n c_q c_r + r_2^2 a_2^2 A_2^2 d_m d_n d_q d_r \right\}$$

$$.\delta(\omega - (m+n+q+r)\omega_a)$$

$$+ \frac{1}{16}\rho_3 P_L \zeta \{ r_1^2 b_1^2 A_1^2 + r_2^2 b_2^2 A_2^2 \} \sum_{m,n,q,r=-\infty}^{\infty} c_m c_n d_q d_r \, \delta(\omega - (m+n+q+r)\omega_a)$$

...(3 – 26)

3.3.1 sine/ cosine amplitude modulation

In section 3.1 it was suggested that differential intensity measurements may be made if the intensity of the two probing beams are modulated in anti-phase. This is equivalent to a quadrature phase shift between the amplitude of the beams. G_1 and G_2 are *amplitude* modulation functions, so we will impose a quadrature relation between two coefficients, c_1 and d_1 so that,

$$c_1 = i d_1 \qquad \ldots(3-27)$$

In the time domain, the optical modulation functions are,

$$g_1(t) = u \sin \omega_a t + H(t) \qquad \ldots(3-28)$$

$$g_2(t) = u \cos \omega_a t + K(t) \qquad \ldots(3-29)$$

where u is a constant and H, K are arbitrary functions that contain no frequency components at ω_a.

Thus the response from each detector is,

detector D1:

$$S_1 = \frac{1}{4} \rho_1 P_L \zeta \sum_{\substack{m,n = -\infty \\ for\ m,n \neq \pm 2}}^{\infty} \{ r_1^2 A_1^2 c_m c_n + r_2^2 A_2^2 d_m d_n \} \delta(\omega - (m+n)\omega_a)$$

$$+ \frac{1}{2} \{ r_1^2 A_1^2 - r_2^2 A_2^2 \} . \rho_1 P_L \zeta |c_1|^2 . \mathrm{F} \{ \cos(2\omega_a t - \varepsilon_1) \}$$

$$+ \ c.c. \qquad \ldots(3-30)$$

detector D2:

$$S_2 = \frac{1}{16}\rho_2 P_L \zeta \sum_{\substack{m,n,q,r = -\infty \\ for\ m,n,q,r \neq \pm 2}}^{\infty} \left\{ r_1^2 a_1^2 A_1^2 c_m c_n c_q c_r + r_2^2 a_2^2 A_2^2 d_m d_n d_q d_r \right\}$$

$$. \delta(\omega - (m + n + q + r)\,\omega_a)$$

$$+ \frac{1}{2} \{ r_1^2 a_1^2 A_1^2 - r_2^2 a_2^2 A_2^2 \} . \rho_2 P_L \zeta |c_1|^4 . \mathrm{F}\{\cos(2\omega_a t - \varepsilon_1)\}$$

$$+ \frac{1}{8} \{ r_1^2 a_1^2 A_1^2 + r_2^2 a_2^2 A_2^2 \} . \rho_2 P_L \zeta |c_1|^4 . \mathrm{F}\{\cos(4\omega_a t - \varepsilon_2)\}$$

$$+ \text{ c.c.} \qquad \ldots(3-31)$$

detector D3:

$$S_3 = \frac{\rho_3}{\rho_2} . S_2$$

$$+ \frac{1}{16}\rho_3 P_L \zeta \{ r_1^2 b_1^2 A_1^2 + r_2^2 b_2^2 A_2^2 \} \sum_{\substack{m,n,q,r = -\infty \\ for\ m,n,q,r \neq \pm 2}}^{\infty} c_m c_n d_q d_r \, \delta(\omega - (m + n + q + r)\,\omega_a)$$

$$+ (0) . \mathrm{F}\{\cos(2\omega_a t)\}$$

$$- \frac{1}{8}\rho_3 P_L \zeta \{ r_1^2 b_1^2 A_1^2 + r_2^2 b_2^2 A_2^2 \} |c_1|^4 . \mathrm{F}\{\cos(4\omega_a t - \varepsilon_2)\}$$

$$+ \text{ c.c.} \qquad \ldots(3-32)$$

where i) $\mathrm{F}\{h(t)\}$ is the fourier transform of $h(t)$ as defined in (3 - 14)

ie., $$\mathrm{F}\{h(t)\} = \int_{-\infty}^{\infty} h(t) \exp(-i\omega t)\ dt \qquad \ldots(3-33a)$$

ii) $\sum_{\substack{m,n=-\infty \\ for\ m.n \neq \pm 2}}^{\infty} h(m, n)$ indicates the sum of $h(m, n)$, *not* including terms

when m and n both equal ± 1. ...(3 – 33*b*)

iii) $\cos \varepsilon_1 = \dfrac{\Re (c_1^2)}{|c_1|^2}$ and $\cos \varepsilon_2 = \dfrac{\Re (c_1^4)}{|c_1|^4}$

where $\Re$ signifies the real part ...(3 – 33*c,d*)

iv) and *c.c.* represents frequency components *not* at $2\omega_a$ or $4\omega_a$.

In deriving the above equations (3–30) to (3–32), the following identity was used:

$$2|z|\mathrm{F}\{\cos(n\omega_a t - \varepsilon)\} \equiv z\delta(\omega - n\omega_a) + z^*\delta(\omega + n\omega_a) \qquad ...(3-34)$$

where z is complex, n is real and $\cos \varepsilon = \dfrac{\Re (z)}{|z|}$.

The responses from each detector has been expressed as an infinite series of delta functions plus a $2\omega_a$ and $4\omega_a$ component. The first point to notice is that a frequency component exists at detectors D2 and D3, which is proportional to the absolute reflectivity of the sample. More important however is that a $2\omega_a$ frequency component exists at all three detectors which contains information concerning the *difference* in reflectivities from the two probed areas. In particular, if for example the infinite series contain no frequency components at $2\omega_a$ and,

$$A_1 = A_2 \qquad ...(3-35a)$$

$$a_1 = a_2 \qquad ...(3-35b)$$

then the amplitude of the $2\,\omega_a$ component at each detector is proportional to $\Delta R \cos 2\,\omega_a t$, where,

$$\Delta R = r_1^2 - r_2^2 \qquad \ldots(3-36)$$

It is this quantity which we are trying to obtain from the *differential intensity* measurement, and it is only when the $2\,\omega_a$ component is proportional to ΔR that the differential result is true.

The differential intensity responses above may be compared with equation (3 - 2) derived in section 3.1, using the relation,

$$\omega_m = 2\,\omega_a, \qquad \ldots(3-37)$$

which equates the frequency ω_m at which the intensity of the probe beams are modulated in anti-phase to twice the frequency at which they are amplitude modulated in phase quadrature. It will be seen in chapter 5 that ω_a is the same as the modulation frequency to the Bragg cell.

Examining the conditions required to achieve true differential intensity measurement, equations (3-35*a,b*) correspond to equal splitting and recombination ratios of the light at the Bragg cell. In order that the infinite series contain no $2\,\omega_a$ frequency components, the optical modulation functions g_1 and g_2 must be carefully selected. One possible choice is $g_1 = \sin\ \omega_a t$ and $g_2 = \cos\ \omega_a t$, which is equivalent to all $c_n, d_n = 0$ unless $n = \pm 1$. This type of modulation is realizable for a Bragg cell driven with a single RF frequency which is triangularly modulated, using the zeroth and first order as the probing beams. It is not so easy to achieve in the two first order system.

3.3.2 identical arbitrary modulations in exact phase quadrature

In a second example, we will look at the consequences of the following conditions:

$$c_n = i^n d_n \qquad \ldots(3-38a)$$

$$A_1 = A_2 \qquad \ldots(3-38b)$$

$$a_1 = a_2 \qquad \ldots(3-38c)$$

This describes the situation when the modulation of each beam on the sample is identical, but for a time delay corresponding to g_1 and g_2 in phase quadrature at frequency ω_a.

Substituting the above relations (3–38) into (3–24), the response from detector D1 is,

$$S_1(\omega) = \frac{1}{4}\rho_1 P_L \zeta A_1^2 \sum_{m,n=-\infty}^{\infty} c_m c_n \left(r_1^2 + i^{-(m+n)} r_2^2\right) \delta(\omega - (m+n)\omega_a) \qquad \ldots(3-39)$$

and hence using (3–34), the $2\omega_a$ frequency component is,

$$S_1(2\omega_a) = \frac{1}{2}|e_3|\rho_1 P_L \zeta A_1^2 . \Delta R \; . \; \mathrm{F}\{\cos(2\omega_a t - \varepsilon_3)\} \qquad \ldots(3-40a)$$

where

$$e_3 = \sum_{\substack{m,n=-\infty \\ for\ m+n=+2}}^{\infty} c_m c_n \qquad \ldots(3-40b)$$

and

$$\cos\varepsilon_3 = \frac{\Re(e_3)}{|e_3|} \qquad \ldots(3-40c)$$

In a similar way, the responses from detectors D2 and D3 at frequencies $2\omega_a$ and $4\omega_a$ are,

$$S_2(2\omega_a) = \frac{1}{8}|e_4|\rho_2 P_L \zeta a_1^2 A_1^2 . \Delta R . F\{\cos(2\omega_a t - \varepsilon_4)\} \quad ...(3-41)$$

$$S_2(4\omega_a) = \frac{1}{4}|e_5|\rho_2 P_L \zeta a_1^2 A_1^2 . R . F\{\cos(4\omega_a t - \varepsilon_5)\} \quad ...(3-42)$$

$$S_3(2\omega_a) = \frac{1}{8}|e_4|\rho_3 P_L \zeta a_1^2 A_1^2 . \Delta R . F\{\cos(2\omega_a t - \varepsilon_4)\} + \frac{1}{4}|e_6|\rho_3 P_L \zeta b_1^2 A_1^2 . R . F\{\cos(2\omega_a t - \varepsilon_6)\} \quad ...(3-43)$$

$$S_3(4\omega_a) = \frac{1}{4}|e_5|\rho_3 P_L \zeta a_1^2 A_1^2 . R . F\{\cos(4\omega_a t - \varepsilon_5)\} + \frac{1}{4}|e_7|\rho_3 P_L \zeta b_1^2 A_1^2 . R . F\{\cos(4\omega_a t - \varepsilon_7)\} \quad ...(3-44)$$

where $$R = \frac{1}{2}\left(r_1^2 + r_2^2\right) \quad ...(3-45a)$$

$$e_4 = \sum_{\substack{m,n=-\infty \\ for\ m+n+q+r=+2}}^{\infty} c_m c_n c_q c_r \quad ...(3-45b)$$

$$e_5 = \sum_{\substack{m,n,q,r=-\infty \\ for\ m+n+q+r=+4}}^{\infty} c_m c_n c_q c_r \quad ...(3-45c)$$

$$e_6 = \sum_{\substack{m,n,q,r=-\infty \\ for\ m+n+q+r=+2}}^{\infty} j^{-(q+r)} c_m c_n c_q c_r \quad ...(3-45d)$$

$$e_7 = \sum_{\substack{m,n,q,r=-\infty \\ for\ m+n+q+r=+4}}^{\infty} i^{-(q+r)} c_m c_n c_q c_r \qquad ...(3-45e)$$

and $$\cos \varepsilon_n = \frac{\Re(e_n)}{|e_n|}, \quad n = 4, 5, 6, 7 \qquad ...(3-45f)$$

The important point to see in the above responses is that with the conditions set out in equations (3–38 *a,b,c*), true differential intensity measurements may only be made from detectors D1 and D2. For detector D3, the two 'extra' beams, μ_2 and μ_3 contribute some absolute reflectivity information which will be detected together with the differential reflectivity, the exact amount being determined by the modulus of coefficient e_6. It has already been shown that this term is equal to zero if sine/ cosine amplitude modulation is used, but in general this term is non-zero.

3.4 The differential phase response

Differential phase is the phase variation between the two probing beams, caused by the presence of the object. It is measured at the beat frequency of the two interfering beams, which is $(\omega_1 - \omega_2)$ for detector D1, and $2(\omega_1 - \omega_2)$ for detectors D2 and D3. The response from each set of detectors will be considered in turn.

3.4.1 differential phase response from detectors D2 and D3

As discussed in section 3.1, the differential phase information is measured at frequency $2(\omega_1 - \omega_2)$, and this frequency component arises only from the overlapping beams *within* μ_1. Therefore the differential phase response from both detectors will be the same, and is obtained from the cross terms in equations (3 - 13) and (3 - 14) :

$$I_{2(\omega_1 - \omega_2)} = 2r_1 r_2 a_1 A_1 a_2 A_2 \rho_x \zeta P_L \left|g_1^2(t).g_2^2(t)\right|_{dc} \times \cos\{2(\omega_1 - \omega_2)t + 2\delta\theta + \delta\alpha + \phi''\} \quad ...(3-46)$$

where $\left|g_1^2(t).g_2^2(t)\right|_{dc}$ is the dc component of $g_1^2(t).g_2^2(t)$; $\delta\alpha = \alpha_1 - \alpha_2$ and $\delta\theta = \frac{1}{2}(\theta_1 - \theta_2)$ are the phase contrast caused by changes in the object conductivity and topography ; ρ_x is equal to ρ_2 for detector D2 or ρ_3 for detector D3, and ϕ'' is a constant. The optical modulation functions, g_1, g_2 describe the amplitude modulation of the light per passage of the Bragg cell. They are defined at the beginning of section 3.2 .

This response can be compared to the interference response given in section 3.1 (equations (3 - 1) and (3 - 4)). It should be noted that in the equations of section 3 .1, the object dependent phase terms, $\delta\alpha$ and $\delta\theta$ have both been absorbed into $\delta\theta$. Also the factor of 2 before $\delta\theta$ in equation (3 - 46) is because any topography effects a phase variation to the light both on incidence and reflection. The height of a topographical feature is then $\frac{\delta\theta}{2\pi} \times \lambda$, where λ is the optical wavelength.

3.4.2 differential phase response from detectors D1

The interference signal from detector D1 is complicated by the fact that the incident beams are inclined, and so necessitates integrating across the detector. Using equations (3 - 7) to (3 - 12), the ($\omega_1 - \omega_2$) frequency component is,

$$I_{(\omega_1 - \omega_2)} = \frac{1}{2d}\int_{-d}^{d} 2r_1 r_2 A_1 A_2 \rho_1 \zeta P_L \left|g_1(t).g_2(t)\right|_{dc}$$

$$\times \cos\left\{\Delta\omega t + 2\delta\theta + \delta\alpha + \phi' + y(\sin\beta_1 + \sin\beta_2)k + z(\cos\beta_2 - \cos\beta_1)k\right\} dy \qquad ...(3-47)$$

where $\left|g_1(t).g_2(t)\right|_{dc}$ is the dc component of $g_1(t).g_2(t)$; β_1 , β_2 are the inclinations of each beam to the optic axis (the total convergence of the two beams is $(\beta_1 + \beta_2)$; k is the modulus of the optical wave vector ; the width of the beams at the detector is equal to 2d, and ϕ' is a constant. The z-axis is along the optical axis, and it is assumed that the sensitive surface of the detector is parallel to the y-axis.

After performing the integration,

$$I_{(\omega_1 - \omega_2)} = 2r_1 r_2 A_1 A_2 \rho_1 \zeta P_L \left|g_1(t).g_2(t)\right|_{dc}$$

$$\times \xi_1 \cos\left\{\Delta\omega t + 2\delta\theta + \delta\alpha + \phi' + \xi_2\right\} \qquad ...(3-48a)$$

where $$\xi_1 = \frac{\sin\left[dk(\sin\beta_1 + \sin\beta_2)\right]}{dk(\sin\beta_1 + \sin\beta_2)} \qquad ...(3-48b)$$

and $$\xi_2 = zk(\cos\beta_2 - \cos\beta_1) \qquad ...(3-48c)$$

The effect of the inclined beams can thus be described by two constants, ξ_1 which is a function of the beam orientations and illuminated area of the detector, and ξ_2 which is

dependent on the beam orientations only. These two constants play quite different roles in the differential response. Variations in ξ_2 whilst taking measurements will clearly have adverse effects on the relative accuracy of the differential phase, whereas ξ_1 affects the signal to noise of measurements, and so the absolute accuracy.

For small orientation angles, ξ_1 may be expressed in terms of the ratio, R_{ss} of focussed beam separation distance to focused beam diameter, such that,

$$\xi_1 = \frac{\sin(2\pi R_{ss})}{2\pi R_{ss}} \qquad \ldots(3-48d)$$

This is plotted in figure 3–3. For imaging $R_{ss} < 1$, and for metrology, $R_{ss} > 1$, and it may be observed that $\xi_1 = 1$ only when R_{ss} is zero. This corresponds to no separation between the beams on the object surface, and is of no use in practice. For a value of the ratio, $R_{ss} = 0.25$, ξ_1 is equal to 0.6, which implies a 2dB drop in the detected signal (voltage), or 4dB in electrical power. The signal drops off fast however, so that $R_{ss} = 0.5$ corresponds to the first null ie., ξ_1 is equal to zero.

By far the biggest concern is in possible variations in ξ_2 . Providing the orientations remain fixed during measurements then neither ξ_1 or ξ_2 will vary in value. In order to get some appreciation for any foreseeable problems, the effects of object tilt and defocus will be considered separately.

Looking first at tilt of the object surface, providing the tilt is small ie., within the depth of focus of the objective, then an angular disturbance at the object transforms into a linear movement at the back focal plane, that is, the detector surface. So both beams move, across the detector, the orientations remaining fixed and hence no change in ξ_1 or ξ_2 .

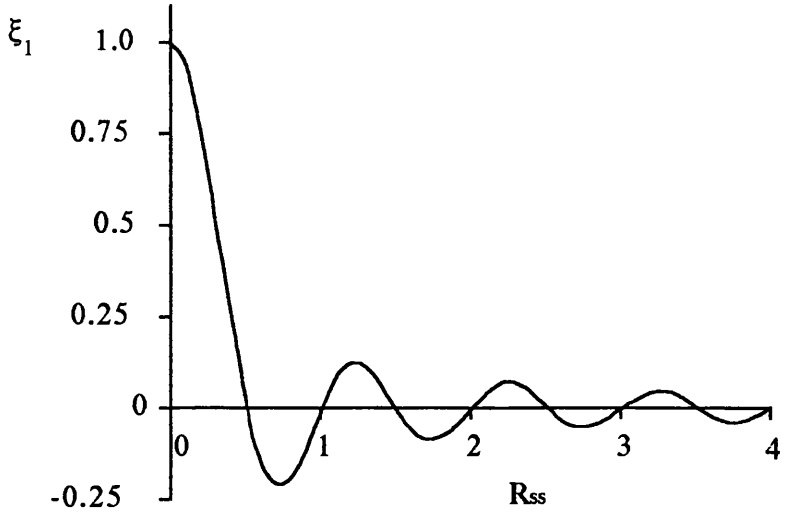


Figure 3-3 Graph of ξ_{1y} versus R_{ss} (equation (3 - 48*d*))

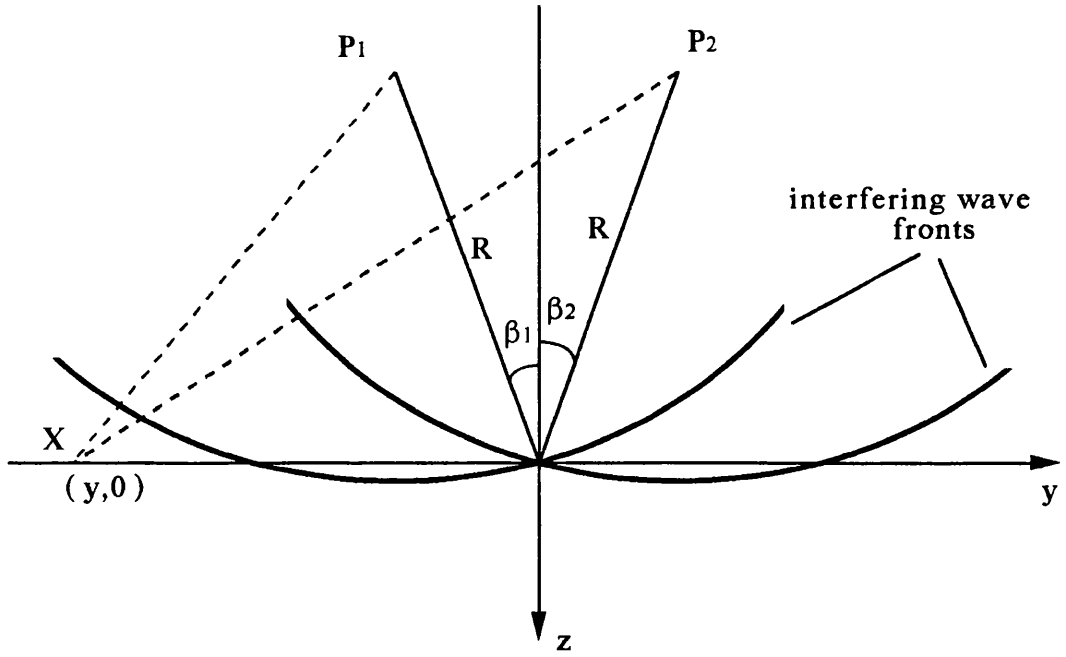


Figure 3-4 The interfering wavefronts at detector D1 with object defocus

The effect of object defocus is that the light *within* each beam will now be converging or diverging. This means that the wavefronts of each beam at the detector are curved, as shown in figure 3–4, where y is the detector plane and R is the distance from this plane at which each beam appears to diverge. If the object is at focus, R is equal to infinity. The phase associated with each wavefront along P_1X and P_2X is,

$$\gamma_1 (y, 0) = \left(R^2 + y^2 - 2Ry \sin \beta_1\right)^{\frac{1}{2}} k \qquad ...(3-49a)$$

$$\gamma_2 (y, 0) = \left(R^2 + y^2 + 2Ry \sin \beta_2\right)^{\frac{1}{2}} k \qquad ...(3-49b)$$

The phase difference of each wave, $\delta\gamma(y, 0) = \gamma_2 - \gamma_1$ follows :

$$\delta\gamma(y, 0) = \frac{Ry}{\sqrt{R^2 + y^2}} (\sin \beta_1 + \sin \beta_2) k \qquad ...(3-50)$$

where terms in $\sin^2 \beta_1$, $\sin^2 \beta_2$ and higher, (or if $\beta_1 = \beta_2$, then terms in $\sin^3 \beta_1$, $\sin^3 \beta_2$ and higher) have been neglected.

This may be further simplified by assuming small defocus, neglecting terms of order $\frac{y^4}{R^4}$ and higher,

$$\delta\gamma(y, 0) = y\left(1 - \frac{y^2}{2R^2}\right)(\sin \beta_1 + \sin \beta_2) k \qquad ...(3-51)$$

Finally the effects of this defocus can be included in the integrand of equation (3 – 47) so that,

$$I_{(\omega_1 \cdot \omega_2)} = \frac{1}{2d} \int_{-d}^{d} 2r_1 r_2 A_1 A_2 \rho_1 \zeta P_L \left|g_1 (t) . g_2 (t)\right|_{dc} \times \cos\{\Delta\omega t + 2\delta\theta + \delta\alpha + \phi' - \delta\gamma(y, z)\}\, dy \qquad ...(3-52)$$

In the above analysis the phase term $\delta\gamma$ has been expressed at $z = 0$, that is, the detector surface is normal to the optical axis, at $z = 0$. Comparing equations (3 - 52) and (3 - 47) it can be seen that the effect of defocus is to add small a second order phase correction term into the integrand, which will have only a very small effect on the measured value of the differential phase.

It may therefore be concluded that object tilt and defocus have very small effects on the differential phase measurement. For more complicated variations, for example where each beam experiences different tilts, this is likely to cause more significant errors. Moreover, if the incidence of the two beams at the detector becomes unsymmetrical or the detector is not exactly perpendicular to the optical axis, movement of the detector in a direction *along* the optical axis is critical. This is different from the situation with detectors D2 and D3 where all that is necessary is that the detectors collect all of the light within μ_1 : exact details on how the detectors is positioned is irrelevant.

3.5 The effect of modulation on the differential phase signal level

The signal level at $\Delta\omega$ and $2\Delta\omega$ from both detectors is dependent upon the optical modulation functions, g_1 and g_2 , which are necessary for differential intensity. Without any modulation, for maximum light in the two probing beams, both g_1 and g_2 are equal to $1/\sqrt{2}$ (this figure is arrived at using equation 3 - 5). This results in $|g_1(t).g_2(t)|_{dc}$ and $|g_1^2(t).g_2^2(t)|_{dc}$ equalling 0.5 and 0.25 respectively. Turning on the modulation will reduce the dc values. As an illustrative example, consider the particularly simple type of modulation, $g_1 = \sin \omega_a t$ and $g_2 = \cos \omega_a t$. After some algebra,

$$g_1(t).g_2(t) = \frac{1}{2}\sin 2\omega_a t \qquad \ldots(3-53)$$

$$g_1^2(t).g_2^2(t) = \frac{1}{8} - \frac{1}{8}\cos 4\omega_a t \qquad \ldots(3-54)$$

It can be seen that there is no dc component in equation (3 – 53), which means that there will be no interference signal from detector D1 at frequency $\Delta\omega$. The electrical power instead is shifted into the sidebands at frequency $\Delta\omega \pm 2\omega_a$ which may be used for the differential phase measurement. However since a sideband is now used, the response will be half of that when measurement is performed on the carrier;

ie.,

$$I_{(\omega_1 - \omega_2 \pm 2\omega_a)} = \frac{1}{2} \times \left[2 r_1 r_2 A_1 A_2 \rho_1 \zeta P_L \left| g_1(t) . g_2(t) \right|_{\omega_a} \right.$$

$$\left. \times \xi_1 \cos\left\{ \left(\Delta\omega \pm 2\omega_a \right) t + 2\delta\theta + \delta\alpha + \phi' + \xi_2 \right\} \right] \quad \ldots(3-55)$$

Table 3–I shows the effects on the interference response at each detector for the modulation switched on and off. It is assumed that the coefficients , A_1, A_2, a_1, a_2 are all equal to 1 ; ρ_1 , ρ_2 and ρ_3 are equal, and phase cancellation is ignored ie., $\xi_1 \approx 1$.

	modulation off		modulation on	
	D 1 $\Delta\omega$	D 2/D3 $2\Delta\omega$	D 1 $\Delta\omega \pm 2\omega a$	D 2/D3 $2\Delta\omega$
relative intensity	$1/2$	$1/4$	$1/4$	$1/8$
electrical power	0dB	6dB	6 dB	12 dB

Table 3–I Comparative interference from detectors D1, D2/ D3 with and without sine/ cosine optical modulation.

It can be seen that when the modulation is switched on the interference signal level falls by 6dB for both detectors. This is because the energy is redistributed into other frequency components. So as would be expected, introducing extra modulation which is necessary for an extra independent measurement naturally reduces the signal level of the first response. Another way of understanding this fall-off is that in order to obtain interference between light from each probed area, both areas must be illuminated simultaneously. This is contrary to the requirements for differential intensity where each area must be illuminated sequentially. So there is a natural trade off between differential phase and intensity signal levels.

The signal from detector D2 appears to be consistently 6dB lower than that from D1. However it should be remembered that the effects of phase cancellation have not been included in the table and this will considerably reduce the signal level from D1. This, combined with the reduction in relative and absolute phase accuracy from D1 due to the detected light not being collinear makes D2 the preferred detector for differential phase measurement.

3.5.1 Examining $|g_1(t).g_2(t)|_{dc}$ for arbitrary quadrature modulation

It was seen that for sine/cosine modulations, the dc component of $g_1\ g_2$ is zero. It is interesting to briefly investigate whether this is the general rule if more arbitrary modulations are used.

Consider the second example given in the section 3.3, of two identical arbitrary modulations which are phase shifted by 90° at frequency ω_a. The c_n and d_n coefficients are then related according to equation (3 – 38) ie., $c_n = i^n d_n$.

The product of g_1 g_2 in the Fourier domain, (using the Fourier transform relation (3 – 15*a*)) is,

$$\Gamma_1 = \int_{-\infty}^{\infty} g_1(t).g_2(t) \exp -i\omega t \quad \mathrm{dt} \qquad \ldots(3-56)$$

Using equations (3 – 19) to (3 – 23), and the expression (ie., $c_n = i^n d_n$) which relates the c_n and d_n coefficients,

$$\Gamma_1 = \frac{1}{4} \sum_{m=-\infty}^{\infty} i^m d_m \left\{ \sum_{n=-\infty}^{\infty} d_n \delta(\omega - (m+n)\,\omega_a) \right\} \qquad \ldots(3-57)$$

Expanding the second summation gives,

$$\Gamma_1 = \frac{1}{4} \mathrm{x}$$

$$\sum_{m=-\infty}^{\infty} \left\{ \sum_{n=-\infty}^{-1} d_n \delta(\omega-(m+n)\,\omega_a) + d_o \delta(\omega - m\omega_a) + \sum_{n=1}^{\infty} d_n \delta(\omega-(m+n)\,\omega_a) \right\} i^m d_m$$

$$\ldots(3-58)$$

It is only the dc component of Γ_1 which is of interest, so ignoring all terms where $m+n \neq 0$:

$$\Gamma_1 \big|_{dc} = \frac{1}{4} \left\{ d_o^2 + \sum_{m=1}^{\infty} i^m d_m d_{-m} + \sum_{m=1}^{\infty} i^{-m} d_m d_{-m} \right\}$$

and simplifying,

$$\Gamma_1 \big|_{dc} = \frac{1}{4} \left\{ d_o^2 + \sum_{m=1}^{\infty} (i^m + i^{-m})\ |d_m|^2 \right\}$$

$$\therefore \qquad \Gamma_1 \big|_{dc} = \frac{1}{4} \left\{ d_o^2 + 2 \sum_{m=1}^{\infty} i^{2m}\ |d_{2m}|^2 \right\} \qquad \ldots(3-59)$$

This is a very interesting result because it shows that the interference at frequency $\Delta\omega$ is not at all reliant on the fundamental frequency of the amplitude modulation. This of

course was already demonstrated in the simple case of the sine/cos modulations above. Also the interference at $\Delta\omega$ is dependent on *even* harmonics and so amplitude modulations such as triangular or square wave which are composed exclusively of odd harmonics, would yield no signal at $\Delta\omega$.

3.6 Summary

Two heterodyne techniques have been presented, the zeroth/ first order and two first order systems. The latter technique has the greatest flexibility and most of the analysis has been discussed with reference to this particular setup.

It was seen that this technique has two modes of operation for imaging or metrology applications, achieved by varying the separation of the probing beams on the object surface. There are three possible detector positions, but in imaging mode, detector D2 cannot be used. Several heterodyne measurements may be made, the most important being differential phase and differential intensity.

The differential intensity response is sensitive to *changes* in object reflectivity, and is termed 'true' differential intensity if the response is proportional to the difference in the intensity reflection coefficients at the two probed areas (equation 3-36). Differential intensity necessitates the two beams to be amplitude modulated in phase quadrature at frequency ω_a. The measurement is performed at frequency $2\omega_a$. A second heterodyne measurement at $4\omega_a$ yields absolute intensity information (this is not available from detector D1). Absolute intensity is proportional to a weighted average of the intensity coefficients of the two interrogated areas.

A set of sufficient conditions were established which will give true differential intensity responses from each detector. These are as follows:

detector D1

i) equal splitting ratio ie., $A_1 = A_2$;

ii) identical amplitude modulation of each beam but for a constant time delay; corresponding to a $\pi/2$ phase shift at frequency ω_a.

detector D2

i) same conditions as detector D1;

ii) equal recombination ratio $a_1 = a_2$.

detector D3

i) same conditions as detector D2;

ii) modulation which satisfies $|e_6| = 0$ (equation 3-45*d*).

The splitting and recombination ratio conditions are controlled by adjusting the relative orientation of the Bragg cell and the light, and should not be difficult to achieve experimentally. The modulation conditions for detectors D1 and D2 can be seen to be the same, and what is most important to appreciate is that modulation corresponding to any arbitrary form is acceptable providing that there is a constant time delay.

Detector D3 requires much more stringent conditions. Although for one particular limiting case (sine/ cosine modulation) true differential intensity measurement is achieved, this is by no means the general rule. If the modulation is not exactly correct, the $2\omega_a$ frequency component comprises the sum of differential *and* absolute reflectivity information. In the two first order configuration, it is extremely difficult to achieve sine/ cosine modulation of the two beams. Moreover, it is unlikely that any other modulation may be produced that will result in the e_6 coefficient being zero, and for this reason, detector D3 is not recommended to be used for true differential intensity measurement.

Differential phase is measured at frequency $2\Delta\omega$ from detectors D2 and D3, and $\Delta\omega$ from detector D1. There are several problems associated with interference measurements made from detector D1. Most importantly the interfering beams are not collinear. This causes phase cancellation and also means that movement of the detector influences measurements, with a consequent reduction in sensitivity. This is in stark contrast to detectors D2 and D3 where the sole requirement is to collect all the required light ie., beam μ_1 . Measurement from detector D1 is further complicated because it must be performed on a sideband at frequency $\Delta\omega \pm \omega_a$. After phase cancellation is taken into account, there is no advantage as far as signal level is concerned in using detector D1 instead of D2 or D3.

The difficulties foreseen with detector D1 are so severe, making detectors D2 or D3 the clear choice for differential phase measurements. Exactly which of the two should be chosen is not so important, but it is preferable to block unneeded light so as to minimize electrical noise, so for this reason detector D2 should be used in preference to detector D3 whenever possible.

CHAPTER 4

THE ACOUSTO-OPTIC INTERACTION

4. The acousto-optic interaction

In the previous chapter the modulation and frequency shifting of each beam by a Bragg cell was seen to enable differential phase and intensity measurements to be independently made. The analysis defined the necessary criteria for a true differential intensity response, and these criteria were specified in terms of the g_1 and g_2 functions, describing the optical modulation from each pass through the Bragg cell. Experimentally it is not possible to have direct control over g_1 and g_2 . The modulation functions are determined by the acoustic field in the Bragg cell, which in turn is generated by an electrical drive. It is this drive that provides the sole means for effecting changes to g_1 and g_2 .

The acousto-optic interaction is not easy to model and for all but the simplest of cases, can only be done numerically. This chapter reviews the current literature on the subject. Much of the theory discussed is described in detail by Korpel in reference [1]. The theory is extremely complex, and in many ways, the purpose of this chapter is to clarify obscurities which exist in this and other original material. The aim is also to present the theory in a form which may be applied to the specific optical system under study.

The next section outlines the past and present work which has been done on modelling the acousto-optic-interaction. Section 4.2 considers a thin acoustic grating which causes Raman-Nath scattering. Bragg scattering requires a thick grating, and this is modelled in section 4.3 by propagating light through many adjacent thin cells. Two parameters, the Klein-Cook and Raman-Nath parameters are used to define the necessary criteria for Raman-Nath and Bragg scattering, and these relate to the interaction length of the Bragg cell and the electrical drive power. Modelling an acousto-optic cell with a multifrequency drive is discussed in section 4.4.

The analysis of section 4.3 showed that the acousto-optic interaction in a thick grating may be represented by an infinite series of integral expressions. In section 4.5, Feynman

[1] 'Acousto-optics', by A. Korpel, published by Marcel Dekker 1988

diagrams are introduced as a means for interpreting these integral expressions, and eliminating those which give a negligible contribution. Section 4.6 applies this method to a case of specific interest, a Bragg cell driven by a two frequency source. An analytic series solution is presented in section 4.7, and its application to the two first order system is discussed in the following section.

4.1 Introduction

The type of device which will be considered in this chapter is shown in figure 4-1. A piezoelectric transducer at one end of a cell is driven by an ac electrical signal which causes a travelling acoustic wave to propagate across the cell. The material used in the cell may be liquid eg. water, or solid (crystalline or amorphous). The propagating acoustic wave gives rise to density variations, and results in refractive index fluctuations which for an acoustic disturbance of low power, is usually linearly related to the electrical drive voltage. This spatial variation in the refractive index is manifest as a phase grating to the incident light, which is diffracted into several orders. Moreover the acoustic wave is travelling, so each order is Doppler shifted in proportion to the angle of diffraction.

The distinction between a phase and amplitude grating should be emphasized. The latter affects the amplitude of the light and is the most familiar type, for example in a ruled grating. A sinusoidal amplitude grating will give only three diffracted orders, -1, 0 and +1, whereas a sinusoidal phase grating results in many more diffracted orders. It is for this reason that an acousto-optic cell, when driven by a single ac frequency, will generate many diffracted orders as shown in figure 4-1.

As the grating becomes thick, the interaction length is increased and light diffracted from different depths within the material destructively interferes so that only a zeroth order

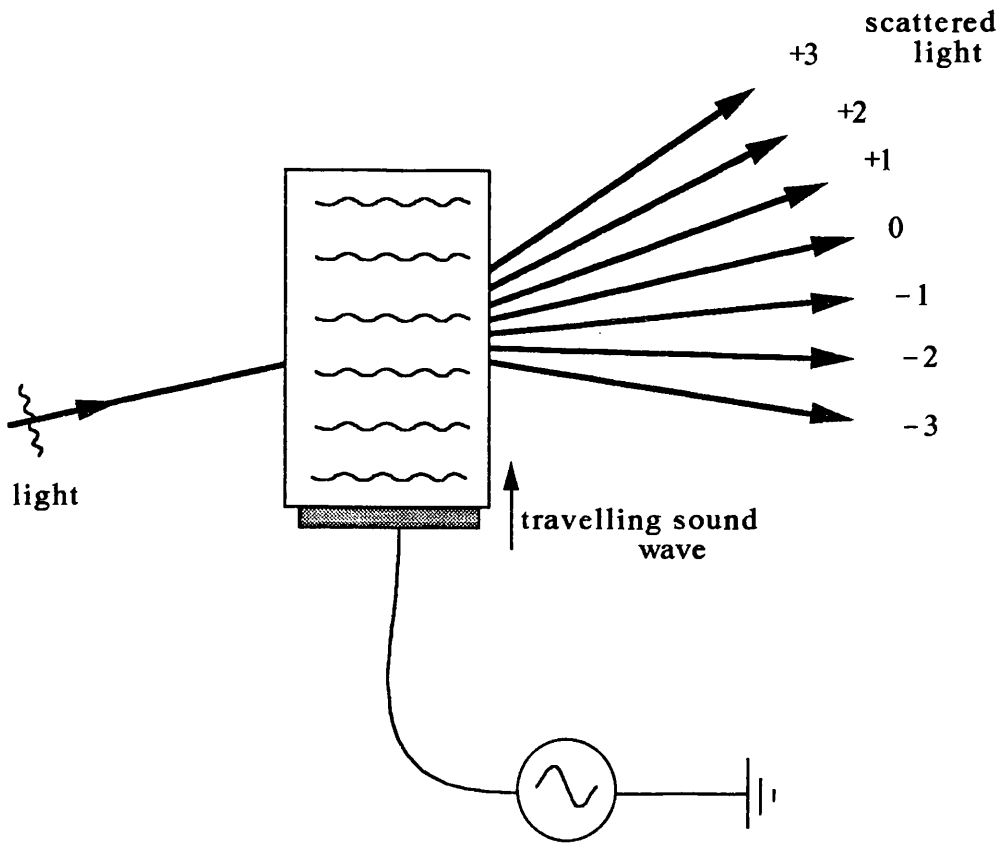


Figure 4-1 The acousto optic interaction

emerges. In the special case of light entering the cell at the so called Bragg angle, a single first order diffracted beam may appear as well as the zeroth order. The acousto-optic cell is now said to be operated in the Bragg regime; the former short interaction length cell operates in the Raman Nath regime.

The differential intensity and phase system uses an acousto-optic cell operated in the Bragg regime ie., a Bragg cell. The zeroth/ first order system uses the zeroth and first order beams as probes. In the two first order arrangement, the Bragg cell is driven by two different frequencies, generating two acoustic waves. The first order beams from each acousto-optic interaction are then used as the probes.

A full description of the acousto-optic interaction is far from simple. The most rigorous approach is to start from Maxwell's equations derive a set of wave equations and solve with the required boundary conditions. The first attempts were made by Brillouin[2], using generalized sound and light fields, but such is the complexity of the mathematics that without the advantage of modern day computers, drastic simplifications had to be made. The various theories became characterized by the model of a rectangular sound column with perfectly straight wavefronts, and perfectly collimated light. Although the theoretical results appeared to agree fairly well with experiment, this is somewhat of a surprise if diffraction of the acoustic field is considered.

The acoustic wave right next to the transducer is planar, but it has been shown experimentally by Osterhammel[3] using a schlieren picture that appreciable variations of amplitude exist across the wavefronts in the very near field. At distances greater than one Fresnel distance these amplitude fluctuations become negligible. The acoustic wavelength for a Bragg cell operating at 80 MHz is typically 50 μm, and if the transducer length is 1 cm, this corresponds to a Fresnel distance of $(1\text{ cm})^2/(50\,\mu m)$ ie., 2 metres. The incoming light is positioned typically within millimetres of the transducer, well

[2]L. Brillouin, Ann. Phys. (Paris, 17:88 (1922)

[3]K. Osterhammel, Akust. Zeit., 6:73 (1941) figure 7.

within the Fresnel distance, and therefore the acoustic wavefront will not be smooth. It can be seen that in this example at least, the acoustic waves will not be planar within the interaction region.

Nevertheless recent workers have continued with this model. The Bragg interaction is conveniently categorised as weak or strong, depending upon whether it is necessary to take into account multiple scattering within the Bragg cell. It is the strong interaction when the acoustic power is such that a large proportion of the light is diffracted out of the zeroth order. With the advent of computers, the model was extended to analyze the strong interaction of an optical beam of Gaussian[4,5] and of arbitrary[6] profiles with a single frequency planar sound beam. Hecht has developed formalism to describe the interaction of a plane optical beam with a multi-frequency planar sound beam[7]. This analysis is particularly useful for the two first order system as it gives an analytical expression for a Bragg cell driven by two discrete RF frequencies, and is described in section 4.7 .

Korpel and Poon have developed theory which is able to encompass the strong interaction of both arbitrary sound and light fields[8,9]. Their method follows analysis laid down by Van Cittert[10]. It relies on dividing the thick interaction region into many thin gratings and individually propagating an infinity of rays from one side to the other side of the cell. The strength of each emerging ray depends how it is diffracted at each grating, and the majority of the paths have a negligible contribution to the final optical field. The trick is to choose only those paths which have a significant effect, and in order to help visualize the processes, Feynman diagrams are employed. Pieper and Korpel have applied this theory to calculate the acousto-optic interaction with curved sound wavefronts[11].

---

[4]R.S. Chu and T. Tamir, J. Opt. Soc. Am., 66: 220 (1976)

[5]L.N. Magdich and V.Y. Molchanov, Opt. Spectrosc.,42(3): 299 (1977)

[6]R.S. Chu and J.A. Kong, J. Opt. Soc. Am., 70: 1 (1980)

[7]D. Hecht, IEEE Vol. SU-24 (1): 7 (1977)

[8]A. Korpel, J. Opt. Soc. Am., 69: 678 (1979)

[9]A. Korpel and T.C. Poon, J. Opt. Soc. Am., 70: 817 (1980)

[10]P.H. Van Cittert, Physica, 4: 590 (1937)

4.2 Thin acoustic field case - Raman-Nath scattering

In this section Raman-Nath scattering of light by a thin acousto-optic cell is described. Although we are not in the end interested in this sort of scattering, the results of the analysis will be used in the next section to describe Bragg scattering by a *thick* cell. The section starts by defining terms used in the rest of the chapter.

Consider a thin acousto-optic cell as shown in figure 4–2. An electrical signal at frequency Ω applied to a transducer, causes a compression wave to travel up the cell in the +x direction. This compression wave is described by the strain parameter, s, such that,

$$s(x, z, t) = \Re e\left[\underline{S}(x, z)\exp(i\Omega t)\right] \qquad ...(4-1)$$

where $\underline{S}$ is a strain phasor describing the amplitude and relative phase of strains within the cell. For example a plane wave travelling in the +x direction is described by,

$$\underline{S}(x, z) = S\exp(-iKx) \qquad ...(4-2)$$

where K is the acoustic wave propagation vector and S is the amplitude of the strain perturbations. The precise dependence of the strain on the signal voltage is governed by the transducer characteristics, but a linear proportionality is usually assumed. In this case the magnitude of S caused by a voltage drive,

$$V = V_0\cos(\Omega t)\ , \qquad ...(4-3)$$

will be proportional to V_0. For completeness, it is necessary to add an extra phase constant ϕ_s into equation (4 – 2) that takes into account any time lag between the voltage drive and the acoustic wave in the cell so that,

$$\underline{S}(x, z) = \left\{S\exp(i\phi_s)\right\}\ \exp(-iKx) \qquad ...(4-4)$$

---

[11] R. Pieper and A. Korpel, J. Opt. Soc. Am., A2:1435 (1985)

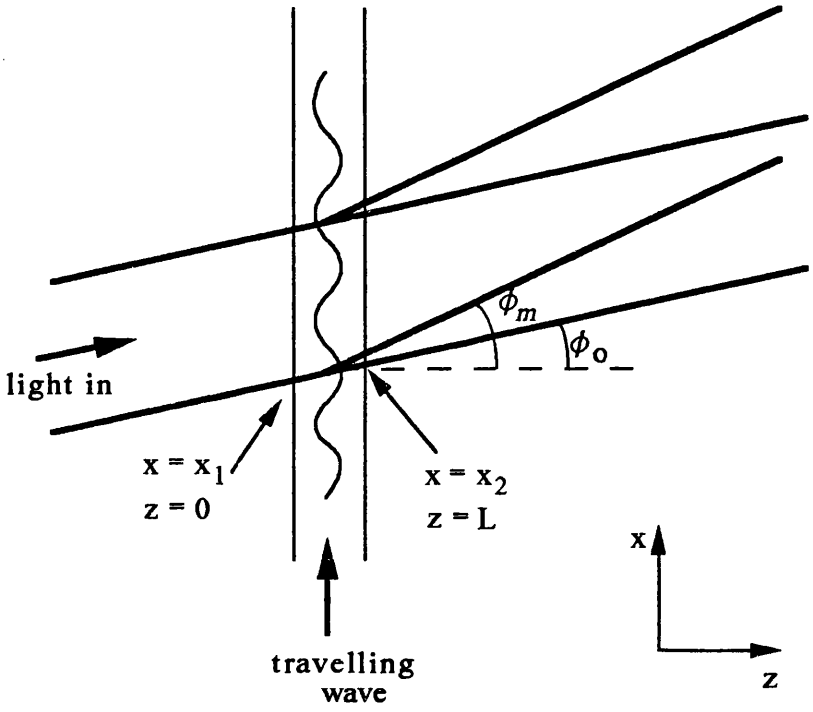


Figure 4-2 A thin acousto-optic cell

Note only one scattered beam is shown

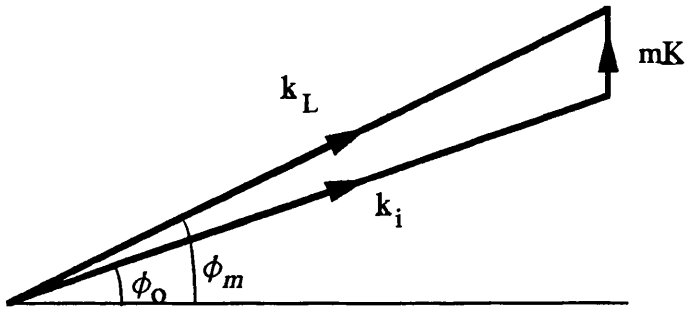


Figure 4-3 Wave vector diagram showing acoustic scattering process

If it is assumed that the sound field propagates in a medium that is both acoustically and optically isotropic and linear, the refractive index ($\underline{\Delta n}$) at all points around the cell is proportional to the strain phasor $\underline{S}$,

ie., $$\underline{\Delta n} = \frac{n_o C}{2} \underline{S} \qquad ...(4-5)$$

where n_o is the steady state refractive index of the material and C is a constant dependent on the photoelasticity properties of the material as described in reference [12]. The refractive index variation in the cell may thus be expressed as,

$$\delta n\,(x,t) = |\underline{\Delta n}|\,\cos[\,Kx - \Omega t + \phi_s\,] \qquad ...(4-6)$$

Each frequency component within the electrical drive will generate its own set of scattered beams. If the drive comprises a carrier with several close sidebands, then the scattered light from acoustic waves generated by the sidebands will not be spatially distinct. It is then easiest to include the time dependence within V_o, that is, in S and $\underline{\Delta n}$.

Providing there is a simple isotropic change in refractive index through density variation as given in equation (4 - 5), the scattered light is to a very good approximation[13] the same polarization as the incident light.

Two assumptions are necessary for Raman-Nath scattering,:

i) the sound field is thin enough to ignore diffraction effects within the cell.

ii) the sound field is weak enough to ignore optical ray bending effects,

ie., $\delta n(x_1,t) \approx \delta n(x_2,t)$ see figure 4–2.

These ensure that the collimation of the light does not change within the cell, and that spreading occurs only after it has left.

[12]A. Korpel, "Acousto-optics", in 'Applied Solid State Science', Vol. 3 (R. Wolfe, ed.), Academic Press, New York, p.71 (1972).

[13]L. Brillouin Actual. Sci. Ind., 59 (1933)

With these assumptions, the field of the light on leaving the acousto-optic cell is[14],

$$E(x, L, t) = E_0 \exp(-ikx \sin\phi_0 - ikL \cos\phi_0)$$

$$\times \sum_{m=-\infty}^{\infty} (-i)^m J_m \left\{ k_v L |\underline{\Delta n}| \operatorname{sinc}\left(\frac{K\phi_0 L}{2}\right)\right\}$$

$$\times \exp\left[-imKx + i(\omega_0 + m\Omega)t + im\left(\frac{K\phi_0 L}{2} + \phi_s\right)\right]$$

...(4 - 7)

where k and k_v are light propagation constants in the medium and vacuum; K is the acoustic propagation constant; ϕ_0 is the angle of incidence of the light at the acousto-optic cell; L is the length of the cell; E_o is the amplitude of the incident field, and ω_o is the optical frequency.

Equation (4 - 7) is the far field diffracted field due to a moving phase grating, and describes a number of diffracted beams, each adjacent order differing in frequency and propagation constant by $\pm\Omega$ and $\pm K$ respectively. The amplitude of each beam is determined by the Bessel function.

We will now study the implications of the frequency and wave vector terms of equation (4 - 7) in greater detail.

Labelling the wave vectors of the incident and transmitted light as $\underline{k}_i$ and $\underline{k}_L$, and the acoustic wave vector as $\underline{K}$, conservation of momentum dictates that $\underline{k}_L = \underline{k}_i + m\underline{K}$. This can be interpreted with reference to figure 4-3 where it can be seen that to satisfy the vector equation, the triangle must be closed. For small scattering angles, $|\underline{k}_L|, |\underline{k}_i| >> |\underline{K}|$, and the magnitude of the incident and transmitted wave vectors remains unchanged. The angle of each diffracted order is then,

[14] 'Acousto-optics', by A. Korpel, published by Marcel Dekker 1988,p. 56

$$\phi_m \approx \phi_0 + m\frac{|K|}{|k|} = \phi_0 + \frac{m\lambda}{\Lambda} \qquad ...(4-8)$$

where λ is the wavelength of light in the medium and Λ is the wavelength of sound. The small angle scattering assumption is good for most applications of acousto-optic scattering, where typically the optical wave vector may be 100 times larger than the acoustic wave vector. Refraction at the air/ material interfaces has been ignored in equation (4 - 8), although it may readily be taken into account.

According to equation (4 - 7), each diffracted beam will have an associated frequency shift in proportion to the angle through which it has been diffracted. This can be shown to be consistent with energy conservation. Writing ω_L, ω_i as the emergent and incident optical frequencies, and ε_{rms} as the time averaged phonon energy, energy conservation is described using the relation,

$$\hbar\,\omega_L = \hbar\,\omega_i + \varepsilon_{rms} \qquad ...(4-9)$$

where $\hbar$ is Planck's constant. An acoustic phonon may be considered as a harmonic oscillator which has equal time averaged kinetic and potential energies. Or equivalently the time averaged energy is equal to twice the mean kinetic energy,

ie., ε_{rms} = (momentum of phonon) x (rms velocity of phonon)

$$= \hbar\,|K|\,v_{rms} \qquad ...(4-10)$$

where v_{rms} is the same as the more familiar acoustic wave velocity. Using the relation,

$K = \dfrac{\Omega}{v_{rms}}$, the mean phonon energy is,

$$\varepsilon_{rms} = \hbar\,\Omega \qquad ...(4-11)$$

Hence, substituting equation (4 - 11) into (4 - 9), the frequency of the transmitted light is,

$$\omega_L = \omega_i + \Omega \qquad ...(4-12)$$

This is the frequency shift for a single phonon-photon collision. Further collisions will result in more frequency shifting in units of Ω. It is equally possible for the light to lose energy on a phonon collision, in which case the diffracted light will have a frequency which is reduced by Ω per collision.

We will now consider the implications of the two assumptions stated earlier in the section as necessary for Raman-Nath scattering. Both of these ensure that the all spreading of the light takes place once it has left the cell.

Looking at the first assumption, figure 4–4 shows a beam of light crossing the acousto-optic cell. Taking a pencil of width Λ within this beam, after traversing the cell it will have spread out by diffraction an amount, $\frac{\lambda L}{\Lambda}$. If each pencil is to remain well collimated within the cell and not interfere with adjacent pencil beams, this divergence must be much less than the pencil width,

ie., $$\frac{\lambda L}{\Lambda^2} << 1 . \qquad ...(4-13)$$

This condition is usually expressed in terms of the Klein-Cook parameter,

$$Q = \frac{LK^2}{k} \qquad ...(4-14)$$

so that $$Q << 1 \qquad ...(4-15)$$

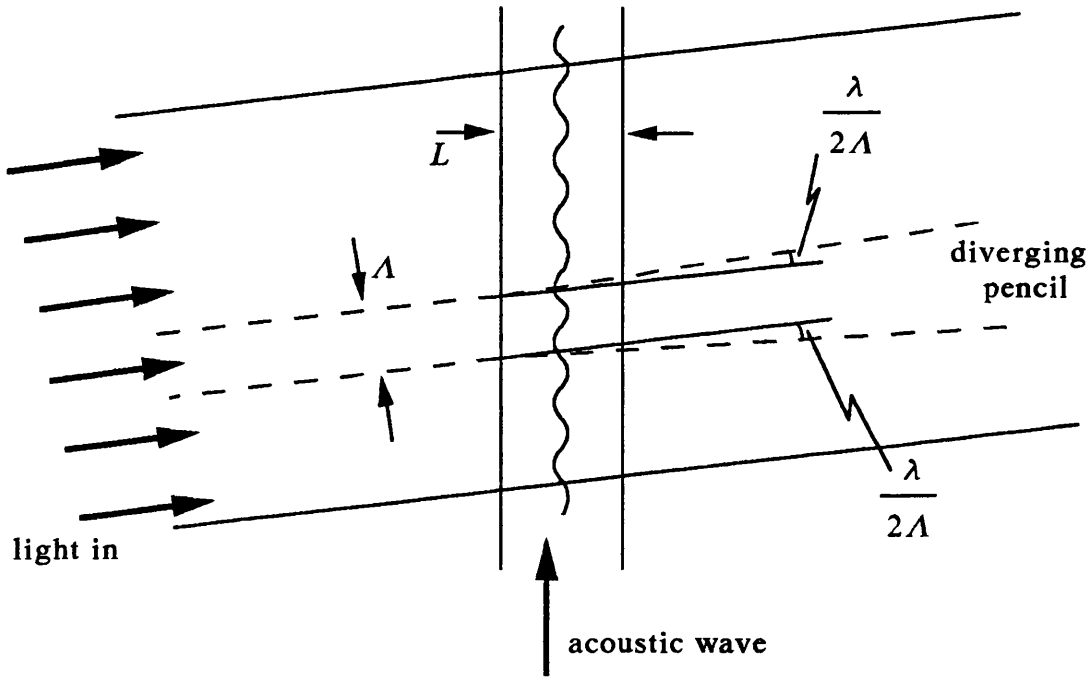


Figure 4-4 Divergence of a single pencil of light, width Λ within the propagating beam.

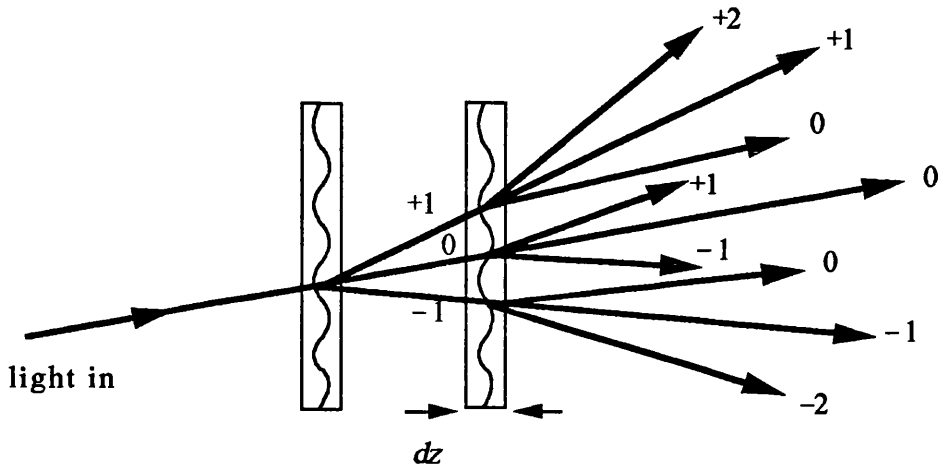


Figure 4-5 The Bragg model is formed by propagating the light succesively through many thin cells

Note the slices are shown well separated for clarity

The second condition is that the refractive index variation ($\underline{\Delta n}$) due to the acoustic wave is sufficiently small that there is no appreciable bending or focussing of the light while passing through the cell. It is shown in reference [15] that this criterion may be fulfilled if,

$$Q\upsilon << 1 \qquad \ldots(4-16)$$

where υ is the Raman-Nath parameter[16] is defined as,

$$\upsilon = k_v L |\underline{\Delta n}|$$

or using the relation, $k_v = \frac{k}{n_o}$ where k_v and k are the optical propagation constants in the vacuum and medium,

$$\upsilon = \frac{k L |\underline{\Delta n}|}{n_o} \qquad \ldots(4-17)$$

The two dimensionless quantities, Q and υ are useful parameters commonly used to define the type of acousto-optic interaction taking place. In any given experimental set up it is usual that Q will remain fixed: it is determined by the frequency of the light (k), the interaction length (L), and acoustic frequency (K). Since Q depends quadratically on the acoustic frequency, by designing acousto-optic cells with different acoustic centre frequencies, Q may vary over several orders of magnitude, ie., from $Q << 1$ to $Q >> 1$.

The amplitude of the voltage drive, V_o (equation 4 - 3) determines the depth of modulation of the refractive index ($\underline{\Delta n}$), which in turn is proportional to υ. The Raman Nath parameter thus indicates the amplitude of the voltage driving the cell. According to Klein and Cook[17] this can, under experimental conditions, range up to a value of ten. In the computer simulations presented in chapter 5, the maximum value of

---

[15]'Acousto-optics', by A. Korpel, published by Marcel Dekker 1988, pp. 59-62

[16]C.V. Raman and N.S.N. Nath, Proc. Indian Acad. Sci., A2:406 (1935)

[17]W.R. Klein and B.D. Cook, 'Unified approach to Ultrasonic light diffraction', IEEE Trans., SU-14, No. 3, p123 (1967)

υ that is used is 3.1 . For the Bragg interaction which is described in the next section, a value of υ equal to π is required to scatter 100% of the incident beam into the first order.

The parameters Q and υ together with a third parameter, α are frequently used to define the acousto-interaction. α which likewise is dimensionless, describes the angle of incidence of the incoming light. All the analysis in this chapter is for small angle scattering in which α is implicitly equal to $\pm 1/2$ (see reference 18).

4.3 Thick sound field - Bragg Scattering

This section describes Bragg scattering and defines the necessary conditions in terms of the Raman-Nath and Klein-Cook parameters. The starting point is from the results of the previous section for Raman-Nath scattering.

For a thick sound column, L increases, Q becomes greater than one and the scattering is no longer in the Raman-Nath regime. Following the analysis of Van Cittert[19], the thick sound column may be subdivided into thin slices and the light propagated from one slice to the next. The width of each slice is made sufficiently small so as to fulfil the Raman Nath conditions. Furthermore, if the slices are taken to be of infinitesimal width dz, each will act as a weak phase grating and generate only two additional orders. The scheme is illustrated in figure 4-5. The incident beam is initially split into three beams, the zeroth order beam equal in amplitude to the incident beam. At the second slice these three beams split into nine. It can be seen that light contributing to, for example the first order, has been generated by more than one sequence of interactions.

Let us consider a slice of thickness dz at position z into the cell. If dz is thin, then there is only interaction between adjacent orders, that is, any change in the optical field of

[18]same as reference 17

[19]P.H. Van Cittert, Physica, 4: 590 (1937)

the m'th order will be dependent on the (m−1)th and (m+1)th order beams *entering* the slice.

Using equation (4 – 7) it can be shown[20] that the change in optical field of the m'th order beam after passing through a slice, thickness dz at position z is,

$$dE_m = -0.25 i k C \underline{S} E_{m-1} \exp[-ikz(\cos\phi_{m-1} - \cos\phi_m)]\ dz$$
$$-0.25 i k C \underline{S}^* E_{m+1} \exp[-ikz(\cos\phi_{m+1} - \cos\phi_m)]\ dz \qquad \ldots(4-18)$$

where k is the optical propagation constant in the material, C is a constant and S is a strain phasor, as defined in the previous section. The term ϕ_m refers to the beam propagation direction (see figure 4–2). This expression shows that the changes in the m'th order beam are due to contributions from the immediate neighbouring orders, E_{m-1} and E_{m+1}.

The field of the m'th order beam emerging from the cell is calculated by integrating equation (4 – 18) from $z = 0$ to L. It can be seen however that the coupled nature of the equations makes performing such an integration far from trivial. One approximation which is particularly useful is for large L and hence $Q >> 1$. On performing the integration, if, $(\cos\phi_{m-1} - \cos\phi_m)$ and $(\cos\phi_{m+1} - \cos\phi_m)$ are non-zero, then the optical phase (given by the exponential term) rotates, resulting in overall phase cancellation.

There are at least two conditions,

$$\cos\phi_o - \cos\phi_1 = 0 \qquad \ldots(4-19a)$$
$$\cos\phi_o - \cos\phi_{-1} = 0 \qquad \ldots(4-19b)$$

in which there is phase matching between the initial light at angle ϕ_o and a neighbouring order. We will now show that these necessitate the incident light being at exactly Bragg incidence ie., $+\phi_B$ or $-\phi_B$, where ϕ_B is the Bragg angle.

---

[20] 'Acousto-optics', by A. Korpel, published by Marcel Dekker 1988, p. 64

Examining the first condition, substituting equation (4 – 19*a*) into (4 – 18),

$$dE_o = -0.25\,i\,k\,C\left\{\underline{S}\;E_{-1}\exp[-ikz(\cos\phi_{-1}-\cos\phi_o)] + \underline{S}^*E_1\;\right\}\,dz$$
$$dE_1 = -0.25\,i\,k\,C\left\{\underline{S}\;E_o + \underline{S}^*E_2\;\exp[-ikz(\cos\phi_2-\cos\phi_1)]\;\right\}\,dz$$

...(4 – 20*a,b*)

These expressions are to be integrated over the interaction region, which is of length, L. Providing this is large, then on integration any exponential terms with a non-zero argument will oscillate and give a negligible contribution to the result. Hence the E_{-1} and E_2 terms in each of the above expressions can be ignored, and we can write,

$$dE_o = -0.25\,i\,k\,C\,\underline{S}^*E_1\;dz \qquad \text{...(4 –21}a\text{)}$$
$$dE_1 = -0.25\,i\,k\,C\,\underline{S}\;E_o\;dz \qquad \text{...(4 – 21}b\text{)}$$

By the same argument, it follows that these are the only non-zero differential terms which can be obtained from equation (4 – 18) with the same condition of (4 – 19*a*). For example substituting equation (4 – 19*a*) into (4 – 18) with m=2 gives,

$$dE_2 = -0.25\,i\,k\,C\,\underline{S}\;E_1\exp[-ikz(\cos\phi_1-\cos\phi_2)]\;dz$$
$$-0.25\,i\,k\,C\,\underline{S}^*E_3\;\exp[-ikz(\cos\phi_3-\cos\phi_2)]\;dz$$

Using the stationary phase argument, it follows that on integration, both terms on the right side integrate to zero.

With the boundary condition $E_o(z=0) = E_i$, and equations (4 – 4), (4 – 5) and (4 – 17), the optical field of the transmitted light is,

$$E_o = E_i\cos\left(0.5\,v\,z/L\right) \qquad \text{...(4 – 22}a\text{)}$$

and

$$E_1 = -i\exp\left(i\,\phi_s\right)E_i\sin\left(0.5\,v\,z/L\right) \qquad \text{...(4 – 22}b\text{)}$$

ie., there are only two beams generated, a zeroth and first order.

If instead, the condition of equation (4 – 19*b*) is met, it follows that there will be a resulting zeroth and minus first order transmitted beam,

$$E_o = E_i \cos\left(0.5 v\, z/L\right) \qquad \ldots(4-23a)$$

$$E_{-1} = -i \exp\left(-i\phi_s\right) E_i \sin\left(0.5 v\, z/L\right) \qquad \ldots(4-23b)$$

The analysis required to derive the above two expressions is very similar to that used for the first condition.

This is an extremely important result as it shows that for a thick phase grating, if the angle of incidence of the light fulfills either equation (4 – 19*a*) or (4 – 19 *b*) there are only two diffracted orders produced, that is, a zeroth and plus or minus first order. For other angles of incidence, there will be no diffracted orders other than the zeroth.

Assuming small ϕ_{-1} , ϕ_0 and ϕ_1 and using equations (4 – 8), the angles satisfying (4 – 19*a,b*), are called the Bragg angles ($\pm\phi_B$), where,

$$\phi_B = \frac{K}{2k} \qquad \ldots(4-24)$$

It has also been found that appreciable power transfer between two orders is possible not only for Bragg incidence, but also for incidence at multiple Bragg angles. The mechanism responsible for this is not as yet fully understood and more information on this subject may be found in reference [21].

In defining the necessary criteria for Bragg diffraction, consider figure 4–5. Unlike Raman-Nath, Bragg diffraction relies on phase cancellation of all but two beams. A pencil of light within the incident beam, after being diffracted, will interfere with adjacent pencils *only* if they have spread out sufficiently to overlap. This is the opposite requirement of that stated in the previous section for Raman-Nath scattering, and leads to

---

21 'Acousto-optics', by A. Korpel, published by Marcel Dekker 1988, p. 144

the condition,

$$Q >> 1 \qquad \ldots(4-25)$$

A more comprehensive criterion includes the modulation level[22] :

$$\frac{Q}{v} = \rho \qquad \ldots(4-26)$$

where the fraction of the total power in the higher diffracted orders is less than or approximately equal to $\frac{1}{\rho^2}$. For example for 90% efficiency of operation, $\rho = \sqrt{10}$.

4.4 Multifrequency Bragg interaction

It has been seen that an acousto optic device driven by the electrical drive, $V = V_0 \cos(\Omega t)$ [equation (4 – 3)], and operated in the Bragg regime, will divide light incident at the Bragg angle into a zeroth and first order beam. The electric field amplitude of each beam (assuming incidence at $+\phi_B$) is given by equations (4 – 22),

ie.,
$$E_o = E_i \cos(0.5 v\, z/L) \qquad \ldots(4-22a)$$
$$E_1 = -i \exp(i\phi_s)\, E_i \sin(0.5 v\, z/L) \qquad \ldots(4-22b)$$

where the parameter v is proportional to V_0. It was further shown that the first order will be frequency shifted by an amount Ω

The question arises, what would be the optical field if the drive comprised many different frequencies. We will look first at the case of a drive modulated by a frequency much less than Ω. In fourier space this is represented by a carrier Ω with several close sidebands. Each frequency component generates its own independent acoustic wave which diffracts the light in proportion to its frequency. If the frequency difference

[22]M.G. Moharam, T.K. Gaylord and R. Magnusson, 'Criteria for Bragg regime diffraction by phase gratings', Opt. Commun., Vol. 32, No. 1, 1980, pp. 14-18.

between the various sidebands is sufficiently small, for a given diffraction order, the divergence angles between the diffracted beams will be less than the natural spreading *within* each beam due to diffraction. So each set of diffracted beams will coalesce into one, and it will be as if only one acoustic wave is present.

Another way to view the situation is that if the modulation is sufficiently slow, the light will pass through the cell before there has been any noticeable change in the intensity of the acoustic wave. The slow modulation may thus be included within the V_0 term, ie., replacing it with $V_0(t)$, and using $v(t)$ in equations (4 - 22*a*,*b*).

For higher modulation frequencies it is necessary to explicitly include each set of acoustic waves. Even though it is usual to assume that there is no coupling between the various acoustic waves within the acousto-optic cell[23], optical coupling does occur *between* the various diffracted beams. Each generated order depletes the source beam from which all other beams come resulting in no straight dependence between the amplitude of one electrical carrier frequency and the intensity of its corresponding diffracted beams.

This coupling also involves the generation of intermodulation beams. Referring to the 'thin slice' model of figure 4-5 where for example we look at light which interacts with acoustic wave Ω_1 in the first slice and Ω_2 in the second slice, it can be seen that light diffracted into the +1 order at the first slice and −1 order at the second slice will have a total frequency shift of $\Omega_1 - \Omega_2$.

Figure 4-6 shows the intermodulation products for Raman-Nath scattering when only two acoustic frequencies are present. In the Bragg regime, phase cancellation results in very little energy being diffracted out of the zeroth and (±) first orders.

[23]In the present discussion non-linear acoustic effects are ignored, although in high power devices, particularly deflectors, they can be of great importance.

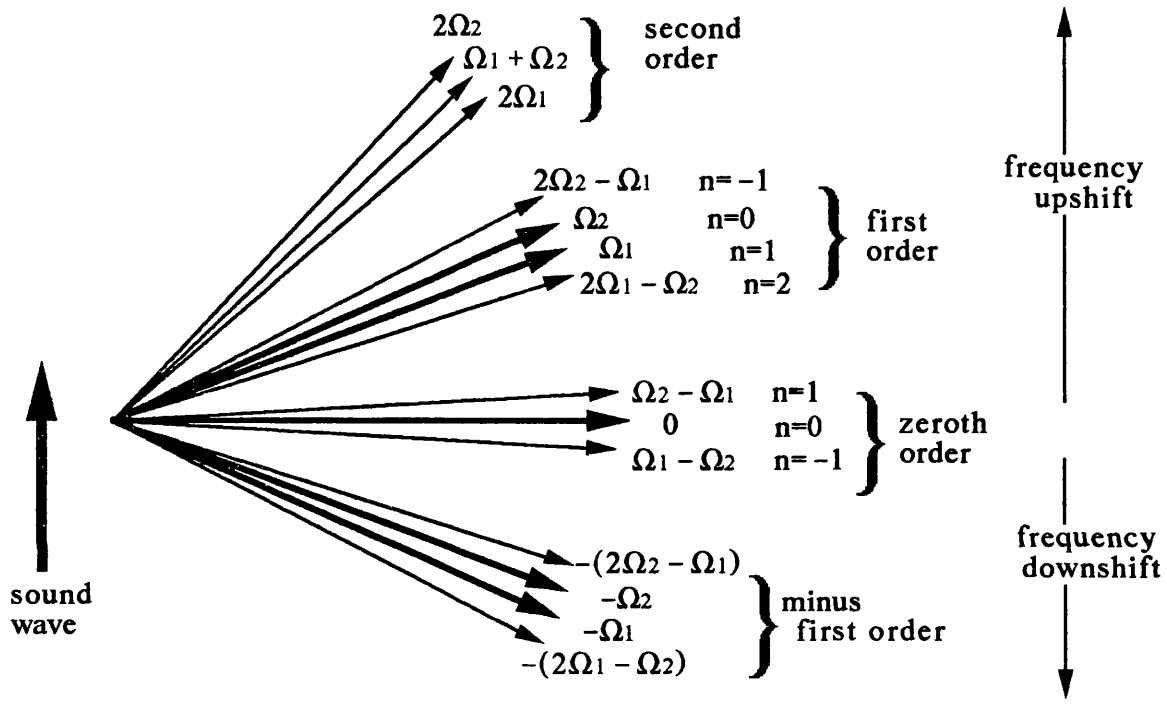


Figure 4-6 Generation of intermodulation beams in an acousto-optic cell driven by two carrier frequencies, Ω_1 and Ω_2.

4.5 Feynman diagram method

In section 4.2 the thin slice model was introduced, and the important relation given by equation (4 – 18) was solved for a single frequency acoustic plane wave. Such a simple analytical solution is very much the exception rather than the rule, and it can be expected that some numerical method will be required to deal with situations of several acoustic frequencies, non-planar acoustic or light fields, or Q of the order of unity. Korpel and Poon[24] have developed a diagrammatic method which uses Feynman diagrams to clarify the more complicated acousto-optic interaction problems. This section explains this method using a few illustrative examples.

Returning to equation (4 – 18) and we will make the substitutions,

$$a = 0.25\, kC \qquad \ldots(4-27)$$

$$\underline{S}^{+}_{m-1} = \underline{S}\exp\left[-ikz(\cos\phi_{m-1} - \cos\phi_m)\right] \qquad \ldots(4-28a)$$

$$\underline{S}^{-}_{m+1} = \underline{S}^{*}\exp\left[-ikz(\cos\phi_{m+1} - \cos\phi_m)\right] \qquad \ldots(4-28b)$$

so that equation (4 – 18) becomes,

$$dE_m = \left\{ -i\,a\,E_{m-1}\,\underline{S}^{+}_{m-1} - i\,a\,E_{m+1}\,\underline{S}^{-}_{m+1} \right\} dz \qquad \ldots(4-29)$$

The $\underline{S}^{+}_{m-1}$ and $\underline{S}^{-}_{m+1}$ terms determine the coupling between adjacent orders. $\underline{S}^{+}_{m-1}$ describes the fraction of the $(m-1)$'th order beam entering the thin slice that is diffracted into the m'th order beam leaving. Similarly, the $\underline{S}^{-}_{m+1}$ describes coupling from the $(m+1)$'th into the m'th order.

Integrating equation (4 – 29) with the boundary conditions,

$$E_m(z) = 0 \text{ for } z \le 0,\ m \neq 0;\quad E_0(z) = E_i \text{ for } z \le 0,$$

[24] A. Korpel and T.C. Poon, J. Opt. Soc. Am., 70: 817 (1980)

the following iterative relation is obtained,

$$E_o(z) = E_i - ia\int_0^z \underline{S}_{-1}^{+} E_{-1}\, dz - ia\int_0^z \underline{S}_{+1}^{-} E_{+1}\, dz \qquad \text{for m = 0}$$

...(4 – 30*a*)

$$E_m(z) = -ia\int_0^z \underline{S}_{m-1}^{+} E_{m-1}\, dz - ia\int_0^z \underline{S}_{m+1}^{-} E_{m+1}\, dz \qquad \text{for m} \neq 0$$

...(4 – 30*b*)

Consider now the field of the first order beam emerging from the cell. Taking equation (4 – 30*a*) with $m = 1$,

$$E_1(L) = -ia\int_0^L \underline{S}_0^{+}(z)\, E_0(z)\, dz - ia\int_0^L \underline{S}_2^{-}(z)\, E_2(z)\, dz\ ,$$

and substituting equations (4 – 30*a,b*) in,

$$E_1(L) = -ia\int_0^L \underline{S}_0^{+}(z)\left\{E_i - ia\int_0^z \underline{S}_{-1}^{+}(z_1)\, E_{-1}(z_1)\, dz_1 - ia\int_0^z \underline{S}_{+1}^{-}(z_1)\, E_{+1}(z_1)\, dz_1\right\} dz$$

$$- ia\int_0^L \underline{S}_2^{-}(z)\left\{-ia\int_0^z \underline{S}_1^{+}(z_1)\, E_1(z_1)\, dz_1 - ia\int_0^z \underline{S}_3^{-}(z_1)\, E_3(z_1)\, dz_1\right\} dz$$

...(4 – 31)

It can be seen that the field $E_1(L)$ comprises a dendritic series of integrals. Any one particular branch terminates when it arrives at E_i, but there will never come a point when all the branches will end.

Figure 4-7 illustrates two arbitrary paths by which the incident light, E_i is diffracted into the first order beam which is transmitted out of the acousto-optic cell. Each path represents a string of interlinked integrals. Tracing along path 1, light is initially diffracted into the first order. It may now either be diffracted back into the zeroth order, into the second order, or alternatively, if it has reached the end of the cell, may contribute to the transmitted first order beam. Path 1 chooses the second option, and the light is diffracted into the second order. Again there are another three choices to be made, and this time it is diffracted back into the first order. The light is now at the end of the acousto-optic cell and emerges as part of the transmitted first order beam.

The contribution of the light following this sequence of diffraction interactions is expressed as a series of integrals:

$$^{path\ 1}E_1\ (L) = (-ia)^3 \int_0^L \underline{S}_2^-(z_3) \left\{ \int_0^{z_3} \underline{S}_1^+(z_2) \left\{ \int_0^{z_2} \underline{S}_0^+(z_1)\, dz_1 \right\} dz_2 \right\} dz_3\ E_i \qquad ...(4-32)$$

A second path also illustrated on figure 4-7 may equally well be followed and a integral expression written down. The distinguishing feature of any path is the *number* of diffraction interactions and the *order* in which they occur. There are indeed a whole infinity of eligible paths and the task is to reduce the problem to evaluating only those with a significant contribution to the final field. Paths for example containing coupling coefficients equal to zero, may be ignored. Also it can be expected that the modulus of the coupling coefficients are less than one, in which case paths containing more integrals will tend to contribute less. Hence by judicious choice of paths, a good estimate may be made of a particular diffracted order with the minimum of computational effort.

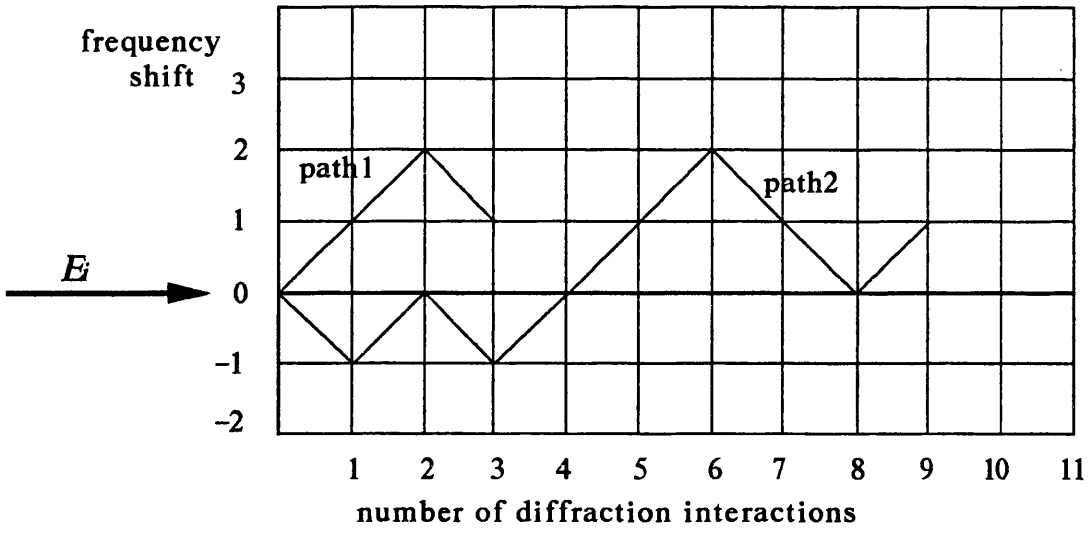


Figure 4-7 Two paths for generating a first order diffracted beam

The exact form of the $\underline{S}^{+}_{m}$ and $\underline{S}^{-}_{m}$ coupling terms will determine whether the cell is operating in the Bragg, Raman-Nath, or some intermediate regime. With the assumption of small angle scattering and using equations (4 - 8) and (4 - 24), the coupling coefficients of equations (4 - 28*a,b*) can be re-written as,

$$\underline{S}^{+}_{m-1} = \underline{S}\exp\left[-i\,Kz(\phi_o + \phi_B(2m-1))\right] \qquad ...(4-33a)$$

$$\underline{S}^{-}_{m+1} = \underline{S}^{*}\exp\left[i\,Kz(\phi_o + \phi_B(2m+1))\right] \qquad ...(4-33b)$$

For the single frequency case and with the incoming light incident at the Bragg angle ($\phi_o = -\phi_B$), this may be further simplified in terms of the Klein-Cook parameter Q, (using also equations (4 - 14) and (4 - 24),

$$\underline{S}^{+}_{m-1} = \underline{S}\exp\left[-iQ(m-1)z/L\right] \qquad ...(4-34a)$$

$$\underline{S}^{-}_{m+1} = \underline{S}^{*}\exp\left[iQmz/L\right] \qquad ...(4-34b)$$

For ideal Bragg diffraction, as was discussed earlier, Q tends to infinity. It is then clear that the only non-zero coupling factors are those that connect levels 0 and +1, ie.,

$$\underline{S}_o^{+} = \underline{S} \qquad ...(4-35a)$$

and

$$\underline{S}_1^{-} = \underline{S}^{*} \qquad ...(4-35b)$$

Therefore if the incident angle of the light is ϕ_B the only non-zero paths for Bragg scattering are those which alternate between the zeroth and first orders. So to calculate the field of the first order beam, the contributions along the following paths must be calculated:

path 1 = E_i to (1st order)

path 2 = E_i to (1st order) to (zeroth order) to (1st order)

path 3 = E_i to (1st order) to (zeroth order) to (1st order) to (zeroth order) to (1st order)

etc.

Evaluating[25] the total contributions along each path gives identical expressions as those shown in equations (4 - 22), (4 - 23). The transmitted zeroth order beam can be calculated in a similar way.

4.6 Solving the two frequency Bragg interaction

The Feynman diagram method has been shown to considerably simplify evaluation of the diffracted optical fields. This technique will now be applied to a case of specific interest, to estimate the intensity of the diffracted orders for a Bragg cell in which the electrical drive comprises two different frequencies :

$$V(t) = V_1(t)\cos\Omega_1 t + V_2(t)\cos\Omega_2 t \qquad ...(4-36a)$$

with

$$V_1(t) = \sin\omega_a t \qquad ...(4-36b)$$

$$V_2(t) = \cos\omega_a t \qquad ...(4-36c)$$

It is assumed that the light is incident at the mean Bragg angle of both acoustic waves ; that the frequency difference, $\Delta\Omega = \Omega_2 - \Omega_1$ is sufficiently large that two distinct first order beams are generated, and that the acoustic fields are planar. With these assumptions, the coupling constants for the up and down transitions are,

$$\underline{S}_o^+ = {}^{\Omega_j}\underline{S} \qquad ...(4-37a)$$

and

$$\underline{S}_1^- = {}^{\Omega_j}\underline{S}^* \qquad ...(4-37b)$$

where j=1, 2 depending on the frequency involved. It may not be possible for a single incoming beam to exactly satisfy the Bragg condition for both planar acoustic waves. This may result in the existence of other non-zero coupling terms which will be small

[25] A. Korpel and T.C. Poon, J. Opt. Soc. Am., 70: 817 (1980)

providing that the frequency difference $\Delta\Omega$ is much smaller than both Ω_2 and Ω_1. In modern Bragg cell devices, stepped transducers can be fabricated which simulate a phased array. The propagation angle of each acoustic wave can then be well matched to the incident light.

It was pointed out in section 4.1 that the light crosses the acoustic field *within* one Fresnel distance from the transducer, in which case the acoustic wave fronts will not be plane. Nevertheless we will retain the above representations of the coupling terms.

Consider the generation of the intermodulation first order beam at frequency $(2\Omega_2 - \Omega_1)$. Figure 4–8 shows the shortest interaction path, and using equations (3 - 30*a,b*), the contribution along this path is,

$$E(L; 2\Omega_2 - \Omega_1) = (-ia)^3 \int_0^L {}^{\Omega_2}\underline{S}\ dz_3 \int_0^{z_3} {}^{\Omega_1}\underline{S}^*\ dz_2 \int_0^{z_2} {}^{\Omega_2}\underline{S} E_i\ dz_1$$

$$= (-ia)^3 \frac{L^3 E_i\ {}^{\Omega_1}\underline{S}^* ({}^{\Omega_2}\underline{S})^2}{6}$$

...(4 - 38)

The strain phasors ${}^{\Omega_1}\underline{S}$, ${}^{\Omega_2}\underline{S}$ can be re-expressed (using equations 4 - 4, 4 - 5, 4 - 17 and 4 - 27) in terms of the Raman-Nath parameter for each of the two acoustic waves, v_1, v_2:

$${}^{\Omega_j}\underline{S} = \frac{v_j}{2aL} \exp(i\phi_j) \quad ; \qquad j = 1,\ 2 \qquad \ldots(4 - 39)$$

where ϕ_j is the phase lag between the drive amplitudes V_1, V_2 and the resulting acoustic field. The actual value of ϕ_j has no effect on the resultant interference intensity and so will not be included in any of the following equations.

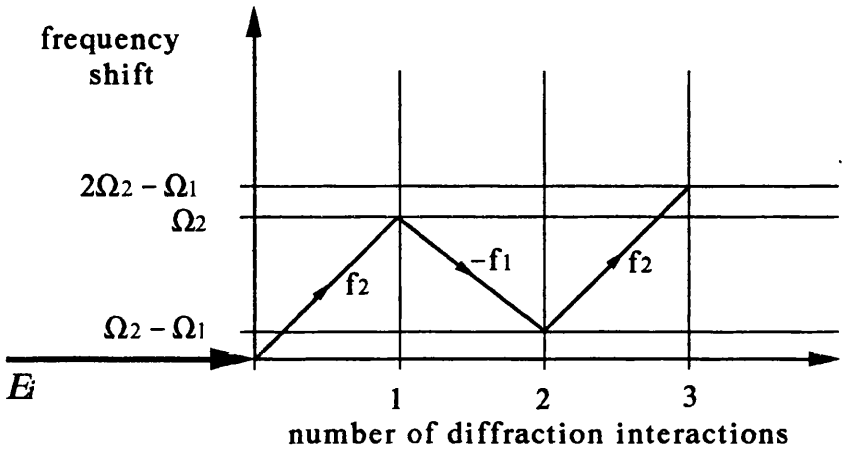


Figure 4-8 Shortest path for generation of $2\Omega_2 - \Omega_1$ intermodulation beam

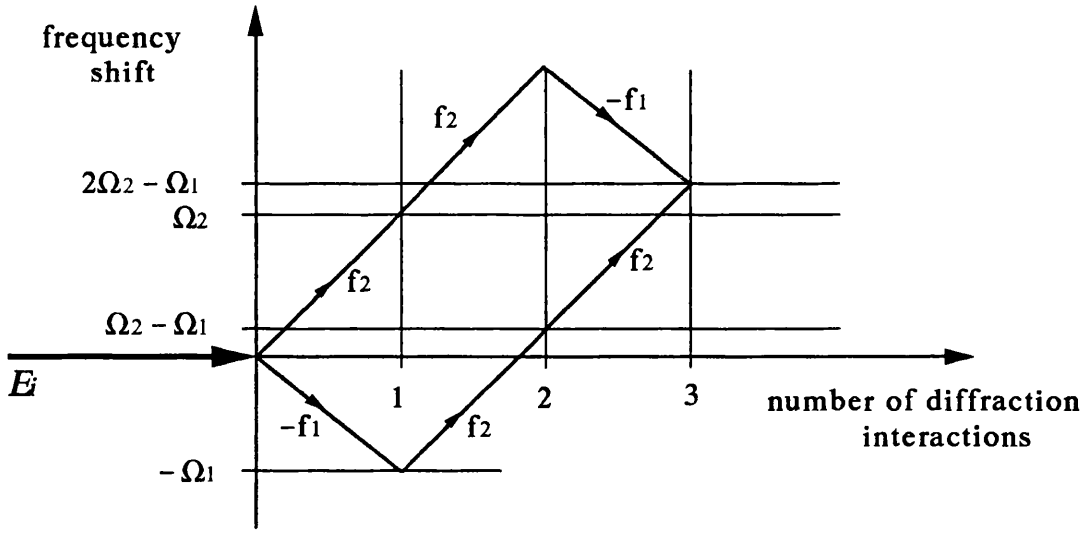


Figure 4-9 Two extra paths present in Raman-Nath scattering for generation of $2\Omega_2 - \Omega_1$ intermodulation beam

Hence the contribution at frequency $(2\Omega_2 - \Omega_1)$ is,

$$E(L; 2\Omega_2 - \Omega_1) = \left(\frac{i}{6}\right)\left(\frac{v_2}{2}\right)^2\left(\frac{v_1}{2}\right)E_i \qquad ...(4-40)$$

In order to accurately calculate the intensity of this beam, many more paths must be examined. Since the only non-zero coupling coefficients are between the zeroth and first orders, any path which gives a significant contribution must alternately move between levels 0 and 1 only. Denoting (f_1) and $(-f_1)$ as up and down transitions caused by an interaction with acoustic wave Ω_1, and similarly (f_2) and $(-f_2)$ for the second acoustic frequency, then a series of allowed transitions contributing to the $(2\Omega_2 - \Omega_1)$ beam are shown in table 4–I.

A path is unique providing the number and sequence of transitions are different to any other path. The transitions within each one of these series may therefore be re-arranged into any order with the proviso that an up transition must always be followed by a down transition. The number of distinct paths which can be made from each series of transitions has been written in column a) of the table.

The field of the light in the $(2\Omega_2 - \Omega_1)$ beam may now be evaluated as follows :

$$\begin{aligned} {}^{total}E_1(L; 2\Omega_2 - \Omega_1) = {} & 1\text{x}(\text{path } 3) \\ & + 2\text{x}(\text{path } 5a) + 3\text{x}(\text{path } 5b) \\ & + 6\text{x}(\text{path } 7a) + 12\text{x}(\text{path } 7b) + 3\text{x}(\text{path } 7c) \\ & + 10\text{x}(\text{path } 9a) + 40\text{x}(\text{path } 9b) + 30\text{x}(\text{path } 9c) + 4\text{x}(\text{path } 9d) \\ & + \text{etc.} \end{aligned} \qquad ...(4-41)$$

path	transitions	No. of permutations a) Bragg	b)Raman-Nath
3a	$(f_2)(-f_1)(f_2)$	${}^1C_1.{}^2C_0 = 1$	${}^3D_{1002} = 3$
5a	$(f_2)(-f_1)(f_2)\ (-f_2)(f_2)$	${}^2C_1.{}^3C_0 = 2$	${}^5D_{1013} = 20$
5b	$(f_2)(-f_1)(f_2)\ (-f_1)(f_1)$	${}^2C_2.{}^3C_1 = 3$	${}^5D_{2102} = 30$
7a	$(f_2)(-f_1)(f_2)\ (-f_1)(f_1)\ (-f_1)(f_1)$	${}^3C_3.{}^4C_2 = 6$	${}^7D_{3202} = 210$
7b	$(f_2)(-f_1)(f_2)\ (-f_1)(f_1)\ (-f_2)(f_2)$	${}^3C_2.{}^4C_1 = 12$	${}^7D_{2113} = 420$
7c	$(f_2)(-f_1)(f_2)\ (-f_2)(f_2)\ (-f_2)(f_2)$	${}^3C_1.{}^4C_0 = 3$	${}^7D_{1024} = 105$
9a	$(f_2)(-f_1)(f_2)\ (-f_1)(f_1)\ (-f_1)(f_1)\ (-f_1)(f_1)$	${}^4C_4.{}^5C_3 = 10$	${}^9D_{4302} = 1260$
9b	$(f_2)(-f_1)(f_2)\ (-f_1)(f_1)\ (-f_1)(f_1)\ (-f_2)(f_2)$	${}^4C_3.{}^5C_2 = 40$	${}^9D_{3213} = \ldots$
9c	$(f_2)(-f_1)(f_2)\ (-f_1)(f_1)\ (-f_2)(f_2)\ (-f_2)(f_2)$	${}^4C_2.{}^5C_1 = 30$	${}^9D_{2124} = \ldots$
9d	$(f_2)(-f_1)(f_2)\ (-f_2)(f_2)\ (-f_2)(f_2)\ (-f_2)(f_2)$	${}^4C_1.{}^5C_0 = 4$	${}^9D_{1035} = \ldots$
	etc.		

TABLE 4-I Interaction paths contributing to the $(2\Omega_2 - \Omega_1)$ beam.

By changing the order of the interactions within each series, different paths may be described. The number of permutations are shown in columns *a)* and *b)* for an acousto-optic cell operated in the Bragg and Raman-Nath regimes.

The combination formulae are as follows :

$$ {}^nC_r = \frac{n!}{r!\ (n-r)!} \quad \text{and} \quad {}^nD_{jklm} = \frac{n!}{j!\ k!\ l!\ m!} $$

In the Bragg regime, only paths which have an up transition followed by a down transition will contribute to the transmitted light. Hence the number of permutations is the product of two ${}^nC_r$ functions, the first describing the number of down transitions and the second, the up transitions. In the Raman-Nath regime, it is not necessary to have an up transition immediately followed by a down transition, so there are many more allowed transitions. The number of permutations is then given by ${}^nD_{jklm}$, where n is the total number of transitions, and j, k, l, m are the number of $(-f_1), (f_1), (-f_2)$ and $(+f_2)$ transitions respectively.

It is of interest to note that for Raman-Nath scattering, since there are more than two non-zero coupling coefficients, many more paths exist for a given selection of transitions. There are now two additional short paths, which are illustrated in figure 4-9. Column *b* of table I indicates the number of paths for each selection of transitions, and it can be seen that there is a considerable increase above the number for the Bragg interaction. This suggests that the intermodulation products in the Raman-Nath regime are considerably worse than under Bragg, and has indeed been confirmed in computer simulations and experiments performed by Hecht[26]

4.7 An analytic solution for two frequency Bragg diffraction

The analysis of the previous section suggests that a series solution may be formulated to describe the diffracted beam intensities. This has been done by Hecht[27] for the two frequency Bragg case, and his solution will be presented in this section.

Since this solution is already in existence, it may at first sight seem that the analysis of the earlier sections is superfluous. Hecht's solution is however for a bounded plane acoustic wave interacting with a planar light beam, and it is possible to use the acousto-optic analysis of this chapter to model more realistic representations of sound and light fields.

Even though the referenced paper shows the solution, it is not at all clear how it was obtained. A different method is used to that outlined in this chapter, and I have been unable to comprehend the way in which the end result is obtained. The solution contains a combinational term C_{Mns} which is similar to those used in table 4-I, but few clues are given as to how it actually enters his solution. Also not all the mathematical symbols are properly defined.

[26]D.L. Hecht, IEEE Trans., SU-24: 7 (1977)

[27]D.L. Hecht, IEEE Trans., SU-24: 7 (1977)

From the analysis of the previous section, the meaning of the combinational term C_{Mns} is far clearer. Hecht uses in his paper variables V_1 and V_2 to describe the intensity of the acoustic wave. Comparison of his result with analysis in this chapter revealed that these variables are exactly equivalent to the Raman-Nath parameters, υ_1 and υ_2. The Feynman diagram approach has thus helped considerably to clarify his solution.

Hecht's formulation is given below with all the various terms now correctly defined. Certain symbolism is changed from that used in the referenced paper so as to match the terms which have been used in earlier sections. These changes are only minor and are listed below.

The assumptions for the model are as follows:

i) all light enters or leaves the Bragg cell at the Bragg angle ;

ii) all interactions are between planar sound and light waves ;

iii) the incident light consists of a single beam with electric field E_i ;

iv) $Q/\upsilon_j >> 1$ for $j = 1, 2$;

v) the material of the Bragg cell is isotropic with no optical or acoustic non-linearities ; optical polarization is ignored.

The field of the first and zeroth orders is :

<u>zeroth order:</u>

$$E_n^o(z\,;\,\upsilon_1,\,\upsilon_2) = i \sum_{\substack{r=0 \\ \text{for } r \text{ even only}}}^{\infty} A_{nr}\;\; E_i \left(\frac{z}{L}\right)^r \qquad \ldots(4-42)$$

first order:

$$E_n^1(z;\, v_1,\, v_2) = i \sum_{\substack{r=0 \\ \text{for } r \text{ even only}}}^{\infty} \frac{-1}{r+1}\left[\frac{v_1}{2}A_{(n-1)r} + \frac{v_2}{2}A_{nr}\right] E_i \left(\frac{z}{L}\right)^{r+1} \qquad \ldots(4-43)$$

where,

$$A_{nr} = \sum_{s=0}^{\left[\frac{(M-|n|)}{2}\right]} C_{Mns}\frac{(-1)^M}{r!}\left\{\frac{v_1}{2}\text{x}\frac{v_2}{2}\right\}^{2s+|n|}\left\{\left(\frac{v_1}{2}\right)^2 + \left(\frac{v_2}{2}\right)^2\right\}^{M-|n|-2s} \qquad \ldots(4-44a)$$

$$M = \frac{r}{2} \qquad \ldots(4-44b)$$

$$C_{Mns} = \frac{M!}{(M-|n|-2s)!\,(s+|n|)!\;s!} \qquad \ldots(4-44c)$$

and,

i) $[x]$ denotes the greatest integer not greater than x; if $x$is negative, then $[x]=0$;

ii) summations are performed for r even only ;

iii) $A_{no} = 0$ for $|n| > 0$;

iv) $0^x = 0$ for $x \neq 0$

$= 1$ for $x = 0$;

v) The variable n refers to the zeroth order or first order beam component, as shown in figure 4–6 ;

vi) When comparing the above equations (4 – 42) to (4 – 44) with the referenced paper, note that, V_1 , V_2 have been replaced by v_1 , v_2 ; $a_{nr} = A_{nr}\, L^r$, and there is an extra factor of i in the equations (4 – 42), (4 – 43) so that the series solution is consistent with the preceding analysis in this chapter.

The field of the transmitted light is obtained at $z = L$.

4.8 Application to two first order system

In the two first order system (figure 3–2), the Bragg cell is driven with an electrical signal,

$$V(t) = V_1(t)\cos \Omega_1 t + V_2(t)\cos \Omega_2 t \qquad ...(4-36a)$$

where V_1 and V_2 are modulations in phase quadrature with an associated frequency much less than Ω_1 and Ω_2.

In chapter 3 (section 3.2) it was described how light from the laser (μ_1) enters the Bragg cell where it is divided into two first orders, κ_1, κ_2. After reflection from the object, these two beams re-enter the Bragg cell, where they are diffracted into the cluster of first order beams, μ_1, μ_2 and μ_3. The g_1 and g_2 modulation functions were introduced in order to describe the modulation of the light after interaction with one or other of the two acoustic waves.

If all the beams μ_1, μ_2, μ_3, κ_1 and κ_2 are assumed to be at Bragg incidence, and the Bragg cell is operated well within the Bragg regime ie.,$Q >> 1$, then

$$g_1(t) = \left| \frac{E_1^1(z; v_1, v_2)}{E_i} \right| \qquad ...(4-45a)$$

$$g_2(t) = \left| \frac{E_0^1(z; v_1, v_2)}{E_i} \right| \qquad ...(4-45b)$$

where v_1 and v_2 are proportional to the modulation voltages, V_1 and V_2 and the splitting/ recombination variables,

$$A_1 = A_2 = a_1 = a_2 = 1 \qquad ...(4-46)$$

If the beams are not at exact Bragg incidence, then the diffraction efficiency is reduced. The solution presented in the previous section is unable to include this situation. The effects may however be included within the splitting/ recombination variables, reducing them to a value less than one.

4.9. Summary

A mathematical model of the acousto-optic interaction has been described. This may be applied to solving acousto-optic interactions in the Raman-Nath, Bragg or some intermediate regime. In particular, a series solution of the two frequency Bragg interaction may be obtained, and it has been discussed how this solution may be applied to the study of the two first order system.

CHAPTER 5

DIFFERENTIAL PHASE AND INTENSITY SENSITIVITY OF THE 2 FIRST ORDER SYSTEM

5. Differential phase and intensity sensitivity of the 2 first order system

Theory presented in chapter 3 set-out modulation criteria which are necessary for an accurate differential intensity response. If the modulations are incorrect, the differential response includes some absolute intensity information, thereby reducing the differential intensity sensitivity. The differential intensity responses were analysed for detectors in three different positions (see figure 3 - 2). Although differential intensity may be measured from any of the detectors, D3 required special selection of the modulations. Consequently, it was suggested that only detectors D1 and D2 should be used.

The aim of this chapter is to enable specifications to be defined for a practical implementation of the differential phase and intensity technique. To this end, it is crucial to know the degradation in system performance for small deviations from the modulation criteria of chapter 3. In the first section, the effects are investigated of unequal splitting/ recombination coefficients (A_1, A_2, a_1, a_2) and the optical modulation functions (g_1, g_2) not in exact phase quadrature.

The optical modulation functions may be controlled only via the electrical drive modulations. In order to evaluate the reduction in sensitivity for electrical modulations of unequal amplitude or not in exact phase quadrature, the analysis of chapter 3 must be combined with the acousto-optic analysis of chapter 4. The second section lays out the necessary formalism.

The third section presents results of computer simulations modelling the acousto-optic interaction within the Bragg cell. These demonstrate the modulated waveforms of the light at different points in the system and the dependence of the differential intensity resolution on the electrical modulations. A value is calculated for the interference intensity and this is used to ascertain the differential phase sensitivity for measurements

made in this system. The model is used to investigate the effects of intermodulation beams on the differential intensity and phase responses.

In the last section, the use of an extra detector, monitoring the total light of both probe beams immediately after the first pass through the Bragg cell is shown to be a feasible method of ensuring true differential intensity measurement from detector D1.

5.1 Unequal splitting/ recombination coefficients and quadrature phase error in the modulation functions

It was shown in section 3.3.2 that true differential intensity may be obtained from detectors D1 and D2 if,

i) the modulations (g_1, g_2) are identical but for a time delay corresponding to a quadrature phase shift at frequency ω_a (any arbitrary form of modulation may be used) ;

ii) the splitting/ recombination coefficients are equal, ie., $A_1 = A_2$ and $a_1 = a_2$.

This section will continue on from this example, and consider the cases where the modulations are still identical but not in exact phase quadrature, and the splitting/ recombination coefficients are no longer exactly equal.

Two parameters β, γ are introduced to describe quadrature phase error and splitting/ recombination mismatch.

Redefining the quadrature relation (equation 3 – 38*a*) in terms of the phase error term β,

$$c_n = d_n \exp i\left(\frac{\pi}{2} + \beta\right) n$$

or $$c_n \approx (i^n + in\beta)\, d_n \qquad \text{for } \beta << 1 \qquad ...(5-1)$$

The splitting and recombination coefficients (equations 3 – 38*b*, 3 – 38*c*) will be re-written in terms of the intensity mismatch term γ, which is assumed to be much less than unity:

$$A_1^2 = (1 + \gamma) A_2^2 \qquad ...(5-2)$$

$$a_1^2 = (1 + \gamma) a_2^2 \qquad ...(5-3)$$

The response of detectors D1 and D2 is obtained after substituting (5 – 1) to (5 – 3) into (3 –24) and (3 – 25) :

detector 1:

$$S_1(\omega) = \frac{1}{4}\rho_1 P_L \zeta A_1^2 \sum_{m,n=-\infty}^{\infty} c_m c_n \{r_1^2 + r_2^2 (1 - \gamma - i\beta(m+n))\, i^{-(m+n)}\} \times \delta(\omega - (m+n)\omega_a) \qquad ...(5-4)$$

Hence the $2\omega_a$ component is,

$$S_1(2\omega_a) = \frac{1}{2}\rho_1 P_L \zeta A_1^2 . |z_1| . \mathrm{F}\{\cos(2\omega_a t - \phi_1)\} \qquad ...(5-5a)$$

where $$z_1 = \{r_1^2 - r_2^2 (1 - \gamma - 2i\beta)\} \sum_{\substack{m,n=-\infty \\ \text{for } m+n=+2}}^{\infty} c_m c_n \qquad ...(5-5b)$$

and $$\cos\phi_1 = \frac{\Re(z_1)}{|z_1|} \qquad ...(5-5c)$$

detector 2:

$$S_2(\omega) = \frac{1}{16}\rho_2 R_L \zeta a_1^2 A_1^2 \ \delta(\omega - (m + n + q + r)\omega_a)$$

$$\times \sum_{m,n,q,r=-\infty}^{\infty} c_m c_n c_q c_r \{r_1^2 + r_2^2 (1 - 2\gamma - i\beta (m + n + q + r))\ i^{-(m+n+q+r)}\}$$

...(5 - 6)

and the $2\omega_a$ component is,

$$S_2(2\omega_a) = \frac{1}{8}\rho_2 R_L \zeta a_1^2 A_1^2 . |z_2| . \mathrm{F}\{\cos\ (2\omega_a t - \phi_2)\}$$

...(5 - 7*a*)

where

$$z_2 = \{r_1^2 - r_2^2 (1 - 2\gamma - 2i\beta)\} \sum_{\substack{m,n,q,r=-\infty \\ for\ m+n+q+r=+2}}^{\infty} c_m c_n c_q c_r$$

...(5 - 7*b*)

and

$$\cos\phi_2 = \frac{\Re(z_2)}{|z_2|}$$

...(5 - 7*c*)

Equations (5 - 5*a,b,c*) and (5 - 7*a,b,c*) were obtained using equations (3 - 23), (3 - 33*a,b,c,d*) and (3 - 34).

In order to compare the accuracy of measurements, it is useful to define the fractional error σ as,

$$\sigma \overset{\text{df}}{=} \frac{|\text{Actual response as measured} - \text{true differential response}|}{|\text{true differential response}|}$$

...(5 - 8)

Using this definition, the fractional error of measurements from each detector is,

detector 1:

$$\sigma_1 = \frac{r_2^2\sqrt{\gamma^2 + 4\beta^2}}{r_1^2 - r_2^2} = \sqrt{\gamma^2 + 4\beta^2} \cdot \frac{R}{\Delta R} \qquad ...(5-9)$$

detector 2:

$$\sigma_2 = \frac{2r_2^2\sqrt{\gamma^2 + \beta^2}}{r_1^2 - r_2^2} = 2\sqrt{\gamma^2 + \beta^2} \cdot \frac{R}{\Delta R} \qquad ...(5-10)$$

If the greatest tolerated differential intensity error is $\sigma = {}^1/_2$, then the differential intensity resolution (${}^{\Delta R}/_R$) from each detector is,

detector 1 : $\psi_1 = 2\sqrt{\gamma^2 + 4\beta^2}$ ...(5-11)

detector 2 : $\psi_2 = 2\sqrt{4\gamma^2 + 4\beta^2}$ ...(5-12)

It can be seen that a non-zero γ causes a drop in resolution which is worse by a factor of two for measurements from detector D2. This is because after one pass through the Bragg cell, the difference in intensity of the two beams is proportional to γ, whereas after two passes, the intensity difference of light from each of the probed areas is proportional to 2γ.

A non-zero value of β however results in an identical reduction in resolution at both detectors. For a given small phase error all that is important is the frequency at which the measurement is made. It makes no difference how many times the light passes through the Bragg cell.

It follows from equations (5 - 11) and (5 - 12) that a phase error of $\beta = 2$ mrad with $\gamma = 0$ results in a differential intensity resolution of 8×10^{-3} from either detector. Alternatively, an intensity mismatch error of $\gamma = 1 \times 10^{-3}$ (with $\beta = 0$) corresponds to a differential intensity resolution of 2×10^{-3} for detector D1 and 4×10^{-3} for detector D2.

5.2 Phase and amplitude variations in the electrical modulations

The quadrature phase error term β introduced in the last section refers to the relative phases of the optical modulation functions. Experimentally, these may only be controlled via the electrical drive to the Bragg cell. However varying the phases of the electrical modulations will in addition modify the modulation functions. In order to properly account for the effects of phase as well as amplitude variations in the electrical modulations, it is necessary to use the acousto-optic theory introduced in chapter 4.

In the two first order system, the Bragg cell is driven with an electrical signal,

$$V(t) = V_1(t) \cos \Omega_1 t + V_2(t) \cos \Omega_2 t \qquad \ldots(4-36a)$$

where Ω_1 and Ω_2 are RF frequencies (10's to 100's of MHz) with associated amplitudes V_1, V_2, which are modulated in phase quadrature. The frequency of modulation is much less than Ω_1 and Ω_2.

The optical modulation functions are given according to equations (4 - 45*a,b*),

ie.,
$$g_1(t) = \left| \frac{E_1^1(z; v_1, v_2)}{E_i} \right| \qquad \ldots(4-45a)$$

$$g_2(t) = \left| \frac{E_0^1(z; v_1, v_2)}{E_i} \right| \qquad \ldots(4-45b)$$

where v_1 and v_2 are the Raman-Nath parameters and are proportional to V_1 , V_2. $E_0^{\,1}$ and $E_1^{\,1}$ are given in equations (4 – 42) and (4 – 43).

It can be seen in the above equations (4 –45*a,b*) that each modulation function is dependent on the amplitude of both RF drive frequencies. Hence as outlined above, altering the phase of either V_1 or V_2 will affect both g_1 and g_2 .

The analysis in chapter 3 manipulated the modulation functions in Fourier space, introducing the c_n and d_n coefficients. Equations (4 – 45*a,b*) express g_1 and g_2 in real space, and to accommodate these numerically derived functions with the minimum of computation, it is useful to modify the theory presented in section 3.3, using discrete instead of continuous Fourier transforms. The following analysis will derive expressions for the differential intensity resolution for measurements made at frequency $2\omega_a$ from detectors D1 and D2.

The frequency response from detector D1 is given by equation (3 – 16). For periodic modulations, this can be re-expressed in terms of the sine and cosine Fourier integrals,

$$v_{nS} = \frac{2}{\tau}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} g_n^2(t)\sin 2\omega_a t \;\mathrm{dt} \quad ; \quad (n=1,2) \qquad \ldots(5-13a)$$

$$v_{nC} = \frac{2}{\tau}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} g_n^2(t)\cos 2\omega_a t \;\mathrm{dt} \quad ; \quad (n=1,2) \qquad \ldots(5-13b)$$

where $$\tau = \frac{2\pi}{\omega_a} \qquad \ldots(5-13c)$$

so that the differential intensity response is,

$$S_1(2\omega_a) = (v_{1C}\cos 2\omega_a t + v_{1S}\sin 2\omega_a t)\, c_1 A_1^2 r_1^2 + (v_{2C}\cos 2\omega_a t + v_{2S}\sin 2\omega_a t)\, c_1 A_2^2 r_2^2 \quad . \qquad ...(5-14)$$

c_1 is a constant term given by $c_1 = \rho_1 P_L \zeta$, ie., it absorbs effects of laser power variations (P_L), loss of power at the beam splitter (ρ_1), and the detection efficiency of the photodiode (ζ).

The previously defined expressions for the differential (ΔR) and absolute reflectivities (R),

$$\Delta R = r_1^2 - r_2^2 \qquad ...(3-36)$$

and

$$R = \frac{1}{2}\left(r_1^2 + r_2^2\right) \qquad ...(3-45a)$$

can now be used to re-write equation (5 – 14) in terms of a sum of the differential and absolute reflectivity components :

$$S_1(2\omega_a) = \Delta R\,(v_{1C}\cos 2\omega_a t + v_{1S}\sin 2\omega_a t)\, c_1 A_1^2 + R\left\{\left[v_{1C}A_1^2 + v_{2C}A_2^2\right]\cos 2\omega_a t + \left[v_{1S}A_1^2 + v_{2S}A_2^2\right]\sin 2\omega_a t\right\} c_1$$

$$...(5-15)$$

Hence the fractional error as defined in equation (5 – 8) from detector D1 is,

$$\sigma_1 = \frac{\sqrt{\left[v_{1C}A_1^2 + v_{2C}A_2^2\right]^2 + \left[v_{1S}A_1^2 + v_{2S}A_2^2\right]^2}}{A_1^2\sqrt{(v_{1C})^2 + (v_{1S})^2}} \cdot \frac{R}{\Delta R} \quad . \qquad ...(5-16)$$

If the smallest allowable fractional error is $^1/_2$, then the differential intensity resolution of

detector D1 is,

$$\psi_1 = 2\frac{\sqrt{\left[v_{1C}A_1^2 + v_{2C}A_2^2\right]^2 + \left[v_{1S}A_1^2 + v_{2S}A_2^2\right]^2}}{A_1^2\sqrt{(v_{1C})^2 + (v_{1S})^2}} \qquad ...(5-17)$$

Turning attention now to the response of detector D2, first define the integrals,

$$v_{nS} = \frac{2}{\tau}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} g_n^4(t)\sin 2\omega_a t \; dt \qquad ...(5-18a)$$

$$v_{nC} = \frac{2}{\tau}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} g_n^4(t)\cos 2\omega_a t \; dt \qquad ...(5-18b)$$

where $$\tau = \frac{2\pi}{\omega_a} \qquad ...(5-18c)$$

The response from detector D2 at $2\omega_a$ is then,

$$S_2(2\omega_a) = \Delta R\,(v_{1C}\cos 2\omega_a t + v_{1S}\sin 2\omega_a t)\,c_2 a_1^2 A_1^2 + R\left\{\left[v_{1C}a_1^2A_1^2 + v_{2C}a_2^2A_2^2\right]\cos 2\omega_a t + \left[v_{1S}a_1^2A_1^2 + v_{2S}a_2^2A_2^2\right]\sin 2\omega_a t\right\}c_2 \qquad ...(5-19)$$

where $c_2 = \rho_2 P_L \zeta$ is a constant.

The fractional error from detector D2 is,

$$\sigma_2 = \frac{\sqrt{\left[v_{1C}a_1^2A_1^2 + v_{2C}a_2^2A_2^2\right]^2 + \left[v_{1S}a_1^2A_1^2 + v_{2S}a_2^2A_2^2\right]^2}}{a_1^2A_1^2\sqrt{(v_{1C})^2 + (v_{1S})^2}} \cdot \frac{R}{\Delta R} \qquad ...(5-20)$$

Finally the differential intensity resolution of detector D2 is,

$$\psi_2 = 2\frac{\sqrt{\left[v_{1C}a_1^2A_1^2 + v_{2C}a_2^2A_2^2\right]^2 + \left[v_{1S}a_1^2A_1^2 + v_{2S}a_2^2A_2^2\right]^2}}{a_1^2A_1^2\sqrt{(v_{1C})^2 + (v_{1S})^2}} \quad ...(5-21)$$

5.3 Computer simulations of the acousto-optic interaction

This section will present the results of several computer simulations modelling the acousto-optic interaction, and is divided into four sub-sections. The first section examines the differential intensity response and the effects of imperfect modulation. In the second section the differential phase sensitivity is calculated, and the last two sections will look at the effects of intermodulation orders on the differential intensity and phase responses. Two types of electrical modulations are considered, triangular modulation of each RF frequency (figure 5–1*a*) and sine/ cosine modulation (figure 5–2*a*). The g_1 and g_2 functions are evaluated using equations (4 – 45*a,b*).

5.3.1 differential intensity

The intensities of the two probe beams κ_1 and κ_2 are proportional to $A_1^2 g_1^2$ and $A_2^2 g_2^2$ (see equations (3 – 8*a,b*)). Hence plots of $g_1^2(t)$ and $g_2^2(t)$ versus time will be proportional to the intensities of the beams, and these are are plotted as traces i) and ii) of figures 5–1*b* and 5–2*b* for triangular and sine/ cosine electrical modulations. If the zeroth order beam (ζ_1 on figure 3–2) is left unblocked, then its intensity is proportional to $g_0^2 = \left|E_0^{\circ}(L; v_1 v_2)\right|^2$, where E_0° is in given equation (4 – 42). This is plotted in trace iii) of figures 5–1 *b* and 5–2*b*.

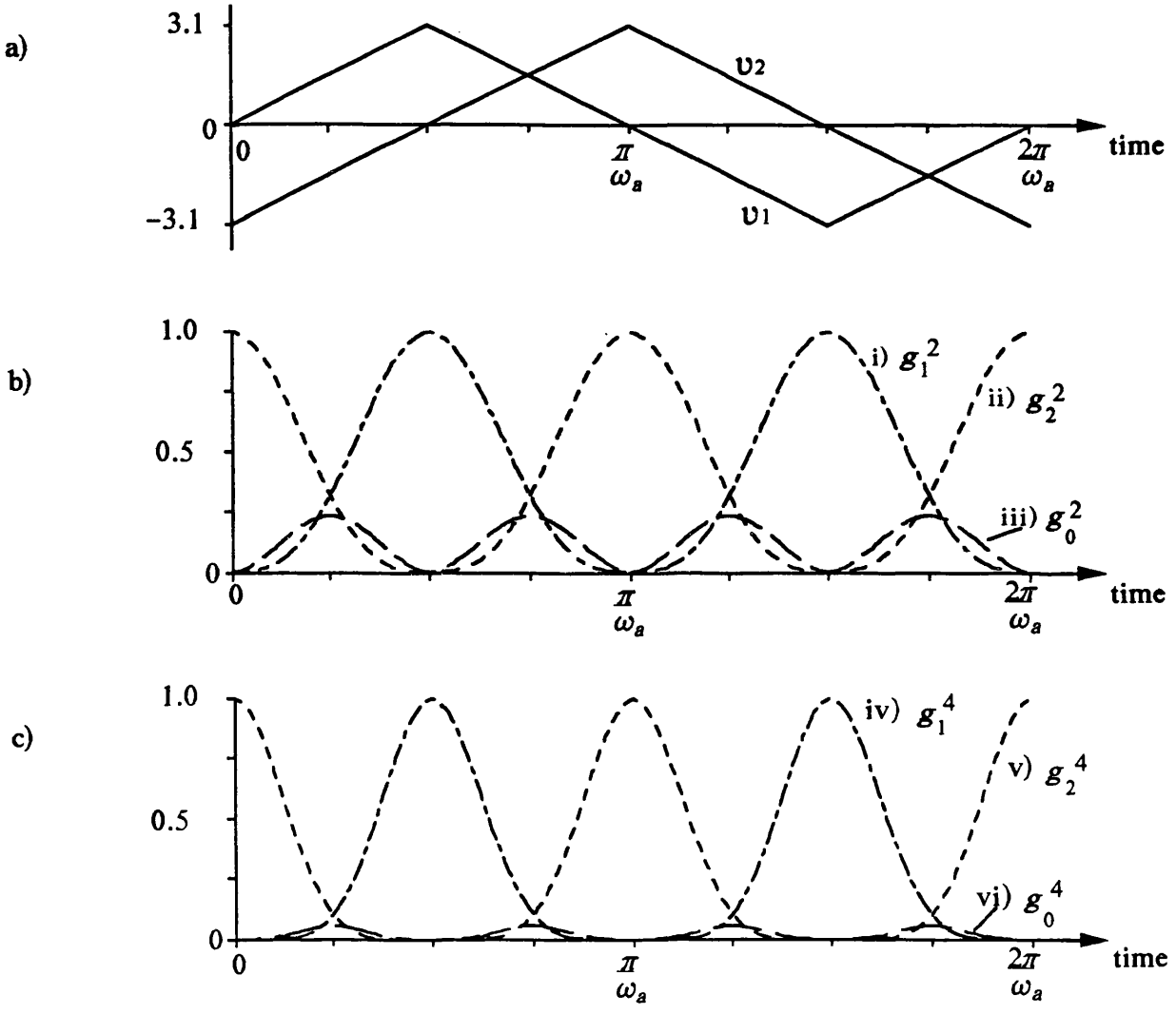


Figure 5–1 Computed values of modulation functions for triangular electrical modulations as shown in 1a) where v is the Raman-Nath parameter. 1b) shows i) g_1^2, ii) g_2^2 and iii) g_o^2; 1c) shows iv) g_1^4, v) g_2^4 and vi) g_o^4.

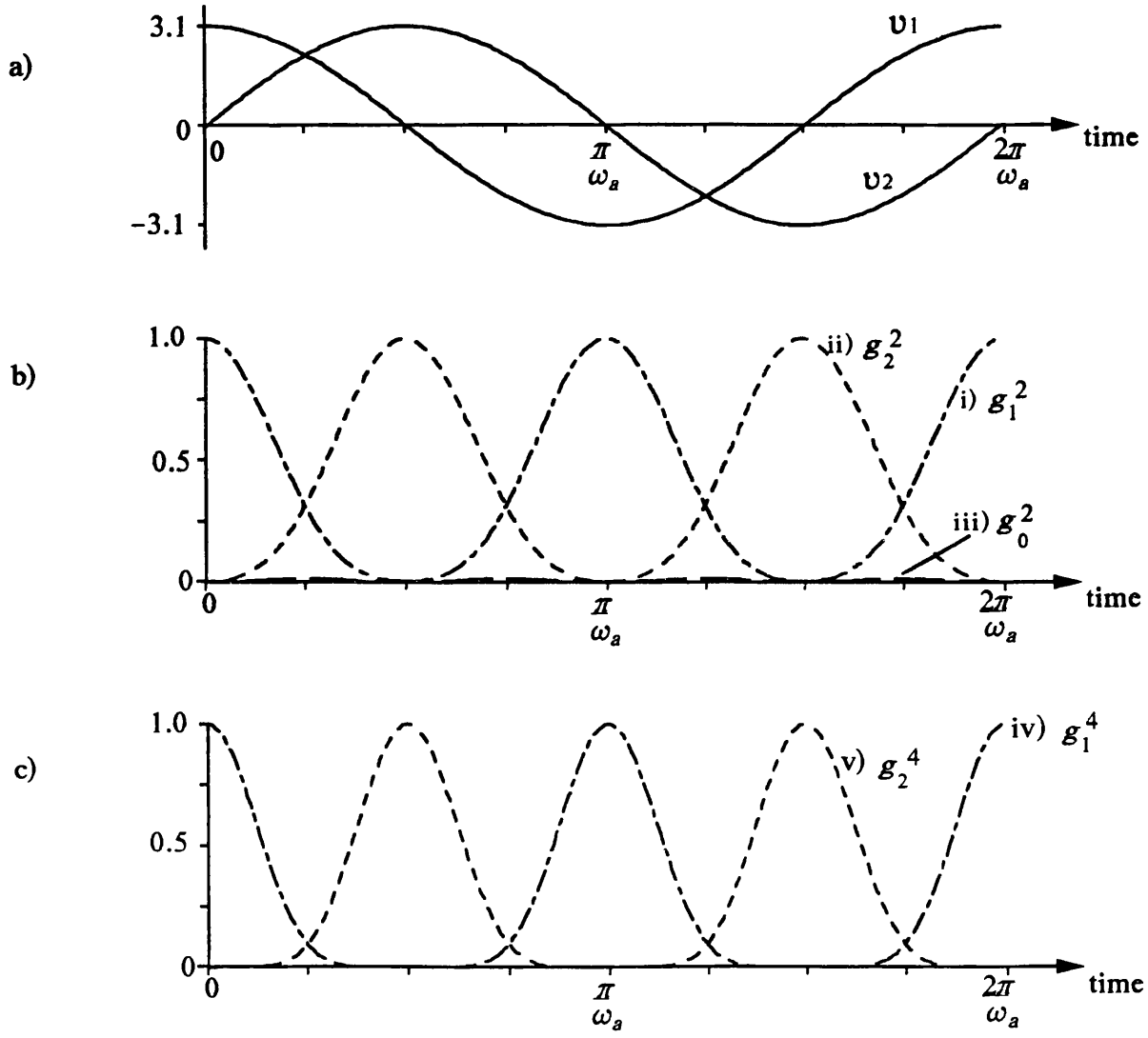


Figure 5–2 Computed values of modulation functions for sine/ cosine electrical modulations as shown in 2a) where v is the Raman-Nath parameter. 2b) shows i) g_1^2 , ii) g_2^2 and iii) g_0^2 ; 2c) shows iv) g_1^4 and v) g_2^4 . g_0^4 is not resolved on this scale.

The vertical axis of figures 5–1*a,b* is v, the Raman-Nath parameter and is proportional to the electrical modulations, V_1 and V_2. It can be seen that with the electrical modulations in phase quadrature, the resulting optical modulation of each beam is in anti-phase. The intensity modulation is at frequency $2\,\omega_a$, twice the electrical modulation frequency. Detector D1 collects all the reflected light of κ_1 and κ_2, and if the reflectivities of both probed areas is the same, the addition of curves i) and ii) gives no resultant $2\omega_a$ frequency component. The differential response is zero corresponding to no variation in object reflectivity. If however each beam experiences different reflectivities, then unequal intensities of each beam are detected resulting in a $2\omega_a$ frequency component which is proportional to the reflectivity difference.

The intensity of the light incident at the Bragg cell is constant, so it follows that the total transmitted light should likewise not fluctuate. It is found that for triangular modulations, the zeroth order intermodulation orders $\Omega_2-\Omega_1$, $\Omega_1-\Omega_2$ (see figure 4–6) account for all but 0.02% of the remaining power. For sine/ cosine modulations this value is 0.25%.

It can be seen from 5–1 *b* and 5–2 *b* that whether it is triangular or sine/ cosine electrical modulation, it makes no perceivable difference to the optical modulation of the two first order beams. For both types of modulations, a similar amount of light is lost into unused beams. The only significant difference is the way in which this optical power is distributed between the zeroth and other intermodulation orders.

After a second passage through the Bragg cell, the intensity of the light from κ_1 and κ_2 which has been diffracted into μ_1 will be proportional to ${g_1}^4$ and ${g_2}^4$. If the zeroth order beam ζ_1 had been left unblocked and reflected from the object, its contribution to the intensity of μ_1 will be ${g_0}^4 = \left|E_0^{\,o}(L;v_1\,v_2)\right|^4$. ${g_1}^4$, ${g_2}^4$ and ${g_0}^4$ are plotted in figures 5–1 *c* and 5–2 *c* as traces iv, v and vi respectively.

In all the results presented in this section, the triangular and sine/ cosine modulations vary

between $v = \pm 3.1$. This value was obtained by plotting the above mentioned curves for different ranges of v and choosing the value which gave undistorted waveforms with a maximum depth of modulation. It is not obvious from the series solutions of the g_1 and g_2 functions exactly what value should be chosen. Taking the much simpler case of a Bragg cell driven by a single RF drive, only a zeroth and first order beam are generated, and the respective intensities are obtained using equations (4 - 22*a,b*). For triangular electrical modulations as shown in figure 5-1 *a* which vary between $\pm\pi$, it follows that the intensities of the zeroth and first order beams are modulated by $\cos^2 2\omega_a t$ and $\sin^2 2\omega_a t$. This suggests that for the two frequency Bragg cell simulations presented here, v should vary in the range of $\pm\pi$ instead of ± 3.1. This is only a 1% correction and will not significantly affect the results.

It is interesting to compare the above results of figure 5-1*b* to the situation of a single frequency RF drive which is triangularly modulated. The intensities of the zeroth and first orders is described by sine squared and cosine squared curves which have an intercept point of 0.5 . In the two first order system, after a single Bragg interaction, this intercept point falls to 0.30 . This is because both acoustic waves are trying to diffract light into their respective first orders. There is only a finite amount of light available, so the light diffracted by each is correspondingly reduced, and the peaks of figure 5-1*b* are squeezed. After a second passage through the Bragg cell, there is further narrowing of the peaks, and the intercept point falls to 8.8×10^{-2} ie., $(0.30)^2$.

Using the analysis of section 5.2, the differential intensity resolution has been calculated for various types of 'imperfect' electrical modulations. Figure 5-3 shows the variation in resolution as the relative modulation amplitudes are varied. (The modulations remain in phase quadrature throughout.) V_1 is held constant at 3.1, and V_2 is varied about this value. It can be seen that for a given change in V_2 , the resolution is higher for a measurement from detector D1 (single pass) than D2. Also at both detectors, the resolution is better by a factor of two for sine/ cosine than triangular modulation.

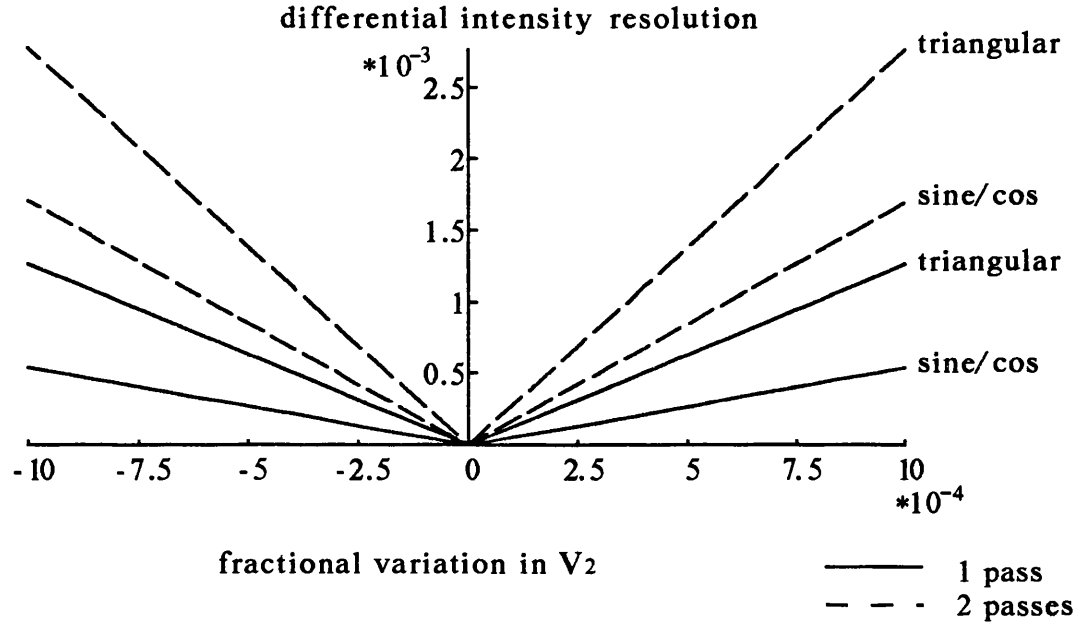


Figure 5-3 Differential intensity resolution for small differences in modulation amplitudes.

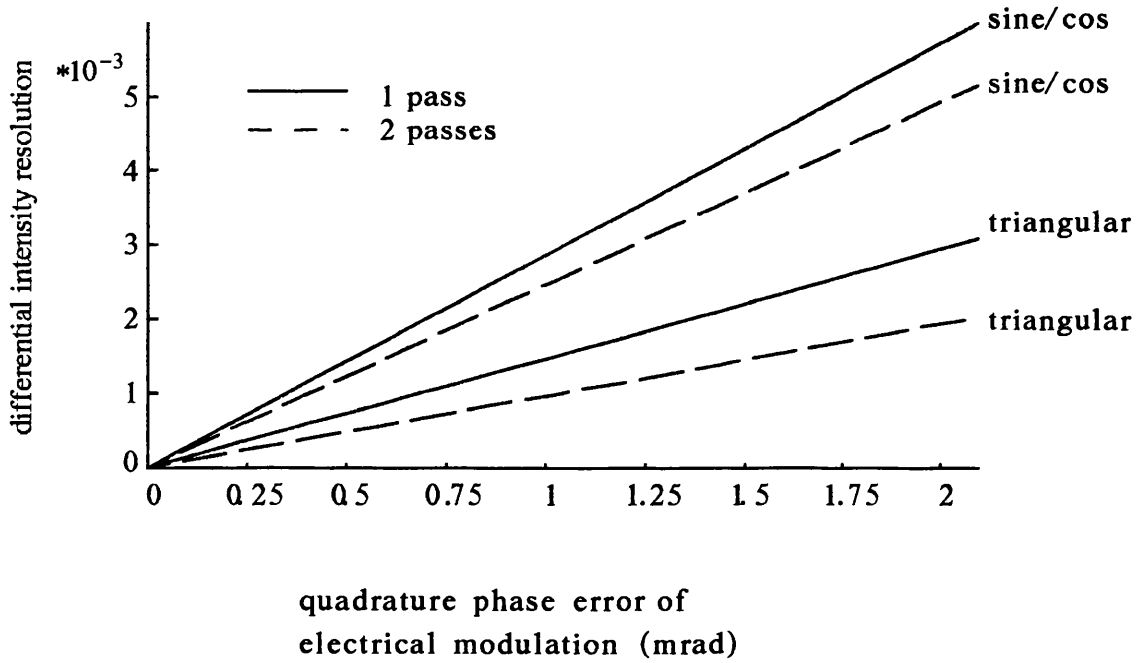


Figure 5-4 Differential intensity resolution for small deviations from exact quadrature modulation.

The demands on the electrical modulation are nonetheless stringent : a differential intensity resolution of 5×10^{-4} (or 13×10^{-4} with triangular modulation) measured from detector D1 requires variations in the voltage modulation of less than 0.1%.

Figure 5–4 shows the effect of a quadrature phase error, with the amplitudes of modulation maintained at 3.1 . It can be seen that in this case the best resolution is from detector D2, using triangular modulation.

From the two sets of results presented in figures 5–3 and 5–4, it is not obvious which is the optimum combination of modulation and detector. The exact choice will depend upon the practical difficulties of constraining phase and modulation amplitude noise. If detector D1 is chosen, it does appear that triangular modulation is perhaps the better choice.

In section 5.1 it was shown that for a 2 mrad error in the *optical* modulation functions the resulting differential intensity resolution is 8×10^{-3} , and that a given phase error will degrade the resolution in the same way at both detectors. From figure 5–4, a phase error of 2mrad in the *electrical* modulation results in a resolution ranging between 2×10^{-3} and 6×10^{-3} which now depends on the chosen position of the detector and the modulation type. It is interesting that the former far simpler approach which does not model the true situation generates a value of the correct order of magnitude. By taking into account the variations in amplitude of both beams, the estimate of the resolution improves to a value less than 8×10^{-3} which is further reduced after a second pass through the Bragg cell.

5.3.2 differential phase

In section 3.5, the effect of the modulations on the differential phase signal level was discussed, and a table was drawn up showing the drop in signal when sine/ cosine *optical* modulations are used. We will now look at the more relevant case concerning the effect of various *electrical* modulations on the differential phase response.

The interference intensity after one and two passes through the Bragg cell is proportional[1] to $g_1 g_2$ and $(g_1 g_2)^2$, and these are shown in figure 5–5. It was found that, within the resolution of the graph, there is no difference between using sine/ cosine or triangular modulations. A comparison is made in table 5–I of the interference for sine/ cosine optical modulation and sine/cosine or triangular electrical modulations. This shows that for sine/ cosine or triangular *electrical* modulations, the signal level at detectors D2/ D3 is 9dB worse in power than for sine/ cosine *optical* modulation.

no modulation		sine/ cosine *optical* modulation		sine/ cosine optical, or triangular *electrical* modulation	
D 1	D 2/D3	D 1	D 2/D3	D 1	D 2/D3
$\Delta\omega$	$2\Delta\omega$	$\Delta\omega \pm 2\omega a$	$2\Delta\omega$	$\Delta\omega \pm 2\omega a$	$2\Delta\omega$
$1/2$	$1/4$	$1/4$	$1/8$	0.15	4.4×10^{-2}
0dB	6dB	6 dB	12 dB	10.5 dB	21 dB

Table 5–I Comparative interference from detectors D1, D2/ D3. The top row is the relative interference intensities, and the second row indicates the detected electrical power.

[1] see equations (3 – 46) and (3 – 55)

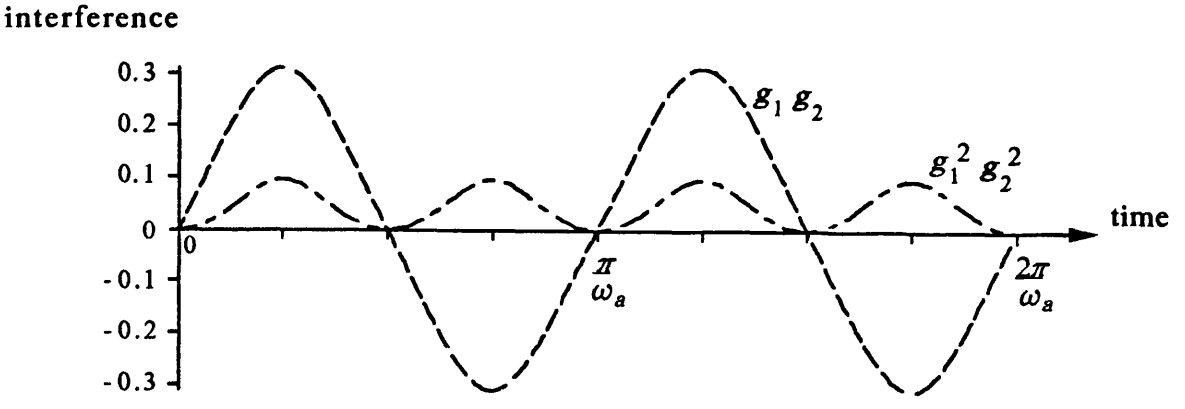


Figure 5-5 Computed values of $(g_1 g_2)$ and $(g_1 g_2)^2$. Plots are identical for triangular and sine/ cosine electrical modulations.

We will now calculate the differential phase sensitivity. From the discussion of chapter 3 it became apparent that detector D1 was not well suited for interference type measurements, so only the sensitivities of detectors D2 and D3 will be examined.

The differential phase response is given by equation (3 – 46) and can be expressed in the following form:

$$I_{2(\omega_1 - \omega_2)} = E_c \cos\ (2\Delta\omega t + 2\delta\theta' + \phi'') + i_n(t) \qquad ...(5-22a)$$

where

$$E_c = 2r_1 r_2 a_1 a_2 A_1 A_2 \rho_x \zeta P_L \left| g_1^2(t)\ g_2^2(t) \right|_{dc} \qquad ...(5-22b)$$

$$\Delta\omega = \omega_1 - \omega_2 \qquad ...(5-22c)$$

$$\delta\theta' = \delta\theta + \frac{\delta\alpha}{2} \quad . \qquad ...(5-22d)$$

and $i_n(t)$ is the noise current. Appendix C shows that the signal to noise ratio for coherent detection of the phase is then,

$$S/N = \frac{E_c^2\ (2\delta\theta')^2}{\frac{1}{2}\ \overline{i_n^2(t)}} \qquad ...(5-23)$$

where $\overline{i_n^2(t)}$ is the time averaged noise current.

In a common optical interferometer, the noise due to microphonics can be reduced much below electrical noise which comprises shot and Johnson (thermal) noise. The noise current is then,

$$\overline{i_n^2(t)} = 2e\,\overline{I_D}BF + \frac{4k_B T_R BF}{R_L} \qquad ...(5-24)$$

where $\overline{I_D}$ is the time average of the total current flowing in the photodiode; R_L is the photodiode load resistor ; B is the overall system measurement bandwidth ; F is the

noise figure of the photodiode preamplifier ; T_R is the noise temperature of the photodiode load resistor, and k_B is the Boltzmann constant.

Shot noise is determined by the number of events n taking place at the photodiode ie., photon absorptions. The mean number of events is proportional to the time averaged current, $\overline{I_D}$, and so it is this rather than the RMS current level which is used in the shot noise expression. Optimum sensitivity is achieved if the thermal noise is much less than the shot noise. This is achieved either by using a high value of the load resistance R_L or by increasing the laser power. For example, a load resistance, R_L of $200\,\Omega$ and an optical power of 1mW will ensure the shot noise term will dominate, and the signal to noise ratio becomes,

$$S/N = \frac{E_c^2 (2\delta\theta')^2}{e\,\overline{I_D}\,B\,F} \qquad ...(5-25)$$

We need to determine the mean current at each photodiode. The time average of any heterodyne interference components is zero, and so for the purpose of determining the time averaged current at each detector, such frequency terms can be negleceted. Hence using equations (3 - 7) to (3 - 14) the current in photodiodes D2 and D3 is,

$$I_{D_2} = \rho_2 P_L \zeta \left[r_1^2 a_1^2 A_1^2 g_1^4 (t) + r_2^2 a_2^2 A_2^2 g_2^4 (t) \right] \qquad ...(5-26)$$

$$I_{D_3} = \rho_3 P_L \zeta \left[r_1^2 a_1^2 A_1^2 g_1^4 (t) + r_2^2 a_2^2 A_2^2 g_2^4 (t) \right] + \rho_3 P_L \zeta \left[r_1^2 b_1^2 A_1^2 + r_2^2 b_2^2 A_2^2 \right] g_1^2 (t)\, g_2^2 (t) \qquad ...(5-27)$$

Using equations (4 - 45*a,b*) and the electrical modulations described in figures 5-1 *a* and 5-1 *b*, the following values have been calculated :

$$\overline{g_1^4 (t) + g_2^4 (t)} = 0.589 \qquad \overline{g_1^2 (t) \times g_2^2 (t)} \equiv \left| g_1^2 (t)\, g_2^2 (t) \right|_{dc} = 4.424 \times 10^{-2}$$

It was found that there was little difference in these values for either type of electrical

modulations. For example $|g_1^2(t)\ g_2^2(t)|_{dc}$ is 0.03% larger for triangular than sine/cosine modulation.

The following assumptions will be made :

i) $a_1 \approx a_2 \approx A_1 \approx A_2 \approx b_1 \approx b_2 \approx 1$

ii) $r_1 \approx r_2 \approx 1$

iii) 50:50 beam splitters are used, and the beam splitter dividing the light between detectors D2 and D3 is not used (only one of these detectors is used at any one time). Hence $\rho_2 = \rho_3 = 1/16$.

iv) there are no losses of light in the Bragg cell or lenses.

Equations (5 – 22*b*), (5 – 26) and (5 – 27) then evaluate as follows :

$$E_c = 2\rho_x\ \zeta P_L\ |g_1^2(t)\ g_2^2(t)|_{dc} \qquad \ldots (5-28a)$$

$$\overline{I_{Dx}} = \rho_x\ \zeta P_L \left\{ \overline{g_1^4(t) + g_2^4(t)} + 2\delta_{Dx}\ \overline{g_1^2(t) \times g_2^2(t)} \right\} \qquad \ldots (5-28b)$$

where $\overline{I_{Dx}} = \overline{I_{D_2}}$, $\delta_{Dx} = 0$ for detector D2

and $\overline{I_{Dx}} = \overline{I_{D_3}}$, $\delta_{Dx} = 1$ for detector D3.

Substituting into equation (5 – 25), the signal to noise ratio at each detector D2, D3 is,

$$(S/N)_{Dx} = \frac{16\rho_x\ \zeta P_L \left(|g_1^2(t)\ g_2^2(t)|_{dc} \right)^2\ (\delta\theta')^2}{eBF\ \{ \overline{g_1^4(t) + g_2^4(t)} + 2\delta_{Dx}\ \overline{g_1^2(t) \times g_2^2(t)} \}} \qquad \ldots (5-29)$$

The minimum detectable $\delta\theta'$ is now calculated by asserting that for this case the signal to noise ratio should equal unity. Hence the smallest detectable phase feature is,

$$(\delta\theta')_{Dx} = \sqrt{\frac{eBF\ \{ \overline{g_1^4(t) + g_2^4(t)} + 2\delta_{Dx}\ \overline{g_1^2(t) \times g_2^2(t)} \}}{16\rho_x\ \zeta P_L \left(|g_1^2(t)\ g_2^2(t)|_{dc} \right)^2}} \qquad \ldots (5-30)$$

The factor ζ is given by equation (3 - 12*b*) :

$$\zeta = \frac{\eta e}{h\nu} \quad . \qquad \ldots (3 - 12b)$$

Using a value of 0.8 for the quantum efficiency and a laser wavelength of 632.8nm, ζ is equal to 0.4083 . Taking a measurement bandwidth (B) of 1 Hz, a laser power (P_L) of 100μW, and a noise figure (F) of 2 for the photodiode pre-amplifier, the phase resolution from detector D2 is 1.6×10^{-6} rad (or 1.6×10^{-3} Å in terms of topography), and 1.7×10^{-6} rad from D3 . It is evident that the effect of detecting the side beams μ_2 and μ_3 reduces the sensitivity by less that 10%.

If the modulation is switched off, the optimum phase sensitivity is achieved when the electrical drives are adjusted so that $g_1(t) = g_2(t) = \sqrt{0.5}$. We then have, $\overline{g_1^4(t) + g_2^4(t)} = 0.5$ and $\overline{g_1^2(t) \times g_2^2(t)} = 0.25$, and the resulting phase sensitivity is $(\delta\theta')_{D2} = 2.5 \times 10^{-7}$ rad, or 2.5×10^{-4} Å in terms of topography. This phase resolution can be compared to a value of 5×10^{-5}Å which is claimed for the differential phase contrast interferometer[2] described in section 2.2.2, part I, (using similar parameters). It appears that the present system is a factor of five less sensitive than the cited instrument, but this is probably due to the way in which effects of elements such as beam splitters have been accounted. For example, in these calculations, the beam splitters have caused a four-fold drop in sensitivity.

Comparing the performance of the two first order system with and without modulation highlights the conflict in the conditions required for optimum phase or intensity measurements. Differential phase is an interferometric measurement and requires both probed areas to be illuminated simultaneously whereas the differential intensity response relies on sequential examination of each area. In the above examples the phase sensitivity is reduced by a factor of five (ie $^{16}/_{2.5}$), and chapter 8 discusses an 'indirect interference' method which enables this problem to be considerably overcome.

---

[2]M. Vaez Iravani, and C.W. See, "Linear and differential techniques in the scanning optical microscope", Proc. SPIE conference on 'Scanning microscopy technologies and applications', Los Angeles, CA, 1988, Vol. 897, p.43

5.3.3 The effect of intermodulation products on the differential intensity response

We have seen that out of the three beams, μ_1 , μ_2 and μ_3 , it is only μ_1 which holds the information required for differential phase and intensity measurement. Moreover, detection of μ_2 and μ_3 can considerably reduce the accuracy of the differential intensity response. This one example highlights the importance of taking into account all extraneous light which is collected by the detector.

Figure 3–2 shows the primary modulation products ie., light which has interacted with only one of the acoustic waves per pass of the Bragg cell. There are in addition intermodulation beams which account for a small fraction of the total optical power. We will look first at the effect of these on the performance of detector D1. After the first passage through the Bragg cell, there are two intermodulation beams either side of κ_1 and κ_2 at frequencies $(2\Omega_1 - \Omega_2)$ and $(2\Omega_2 - \Omega_1)$. These are illustrated in figures 4–6 and 5–6*a* where thick lines are used to indicate the main beams and the thin lines for the intermodulation beams. The angular separation of these from κ_1 and κ_2 is given by,

$$\alpha = \beta . \frac{\Omega_2 - \Omega_1}{\Omega_2} \qquad ...(5-31)$$

where β is the angular divergence of κ_1 and κ_2 . Typically $\alpha < 0.1\,\beta$, and if this value is less than the natural divergence *within* each beam due to diffraction, it is physically not possible to separate the intermodulation orders from the main probe beams. There will in addition be further intermodulation beams, but they will hold even less optical power and so will be ignored in the present discussion.

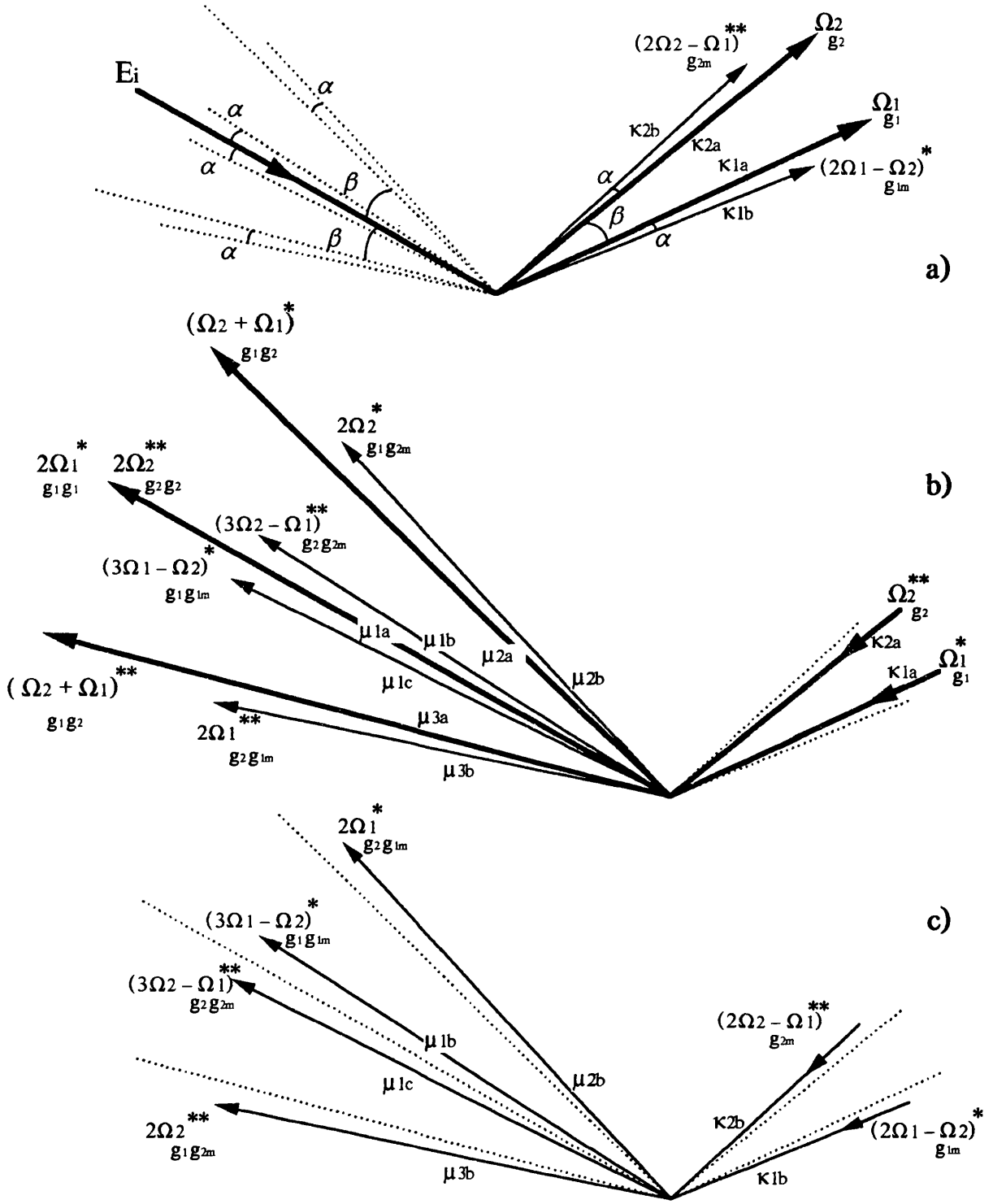


Figure 5-6 S c h e m a t i c showing generation of diffracted orders at Bragg cell. a) First pass through Bragg cell ; b) Second pass - interaction of κ_{1a} and κ_{2a} ; c) Second pass - interaction of κ_{1b} and κ_{2b}.
Thick lines indicate strongly diffracted beams, thin lines are intermodulation orders. For each emerging beam, the optical field amplitude (g coefficient), optical frequency shift (Ω) and source beam (* or **) is indicated.

The intensities of the intermodulation beams are,

$$(g_{1m})^2 = |E^1_{-1}(L;\, \upsilon_1,\, \upsilon_2)|^2 \qquad ...(5-32)$$

$$(g_{2m})^2 = |E^1_{2}(L;\, \upsilon_1,\, \upsilon_2)|^2 \qquad ...(5-33)$$

where E^1_n $(n = -1, 2)$ is given in equation (4 - 43). These functions are plotted in figures 5–7*a,b* for triangular and sine/ cosine modulations. The labelling ' i, ii, i_m and ii_m' refer to the same beams on all the figures 5–1 *b*, 5–6*a* and 5–7*a,b*.

It can be seen from figure 5–7 *a,b* that both the $(2\Omega_1 - \Omega_2)$ and $(2\Omega_2 - \Omega_1)$ intermodulations have $2\omega_a$ frequency components which are approximately 1×10^{-2} and 2×10^{-3} in amplitude for sine/ cosine and triangular modulations respectively. The $2\omega_a$ frequency component of the intermodulations are in anti-phase, so detecting both at detector D1 will give a differential response. This is important as it means that the intermodulation beams do not contribute any detrimental absolute intensity signal to the response of detector D1. The differential response will be of opposite sign to that from κ_1 and κ_2 and will result in at the most, a 1% error in a differential intensity measurement. If these intermodulation orders are collected by the detector D1, then any potential error will be minimized by ensuring that *all* of both beams are detected.

After a second passage through the Bragg cell, the situation becomes somewhat more complicated. κ_1 and κ_2 are divided into three main beams and four intermodulation beams as illustrated in figure 5–6 *b*. Light which has originated from κ_1 is denoted by one asterisk, and from κ_2 by two asterisk. The intermodulation beams which were generated on the first passage through the Bragg cell are divided into four as shown in figure 5–6 *c* (second order intermodulation products have been ignored). For each beam on figures 5–6*a,b,c*, the respective amplitudes are indicated. The intensity of a detected beam is equal to the square of this value.

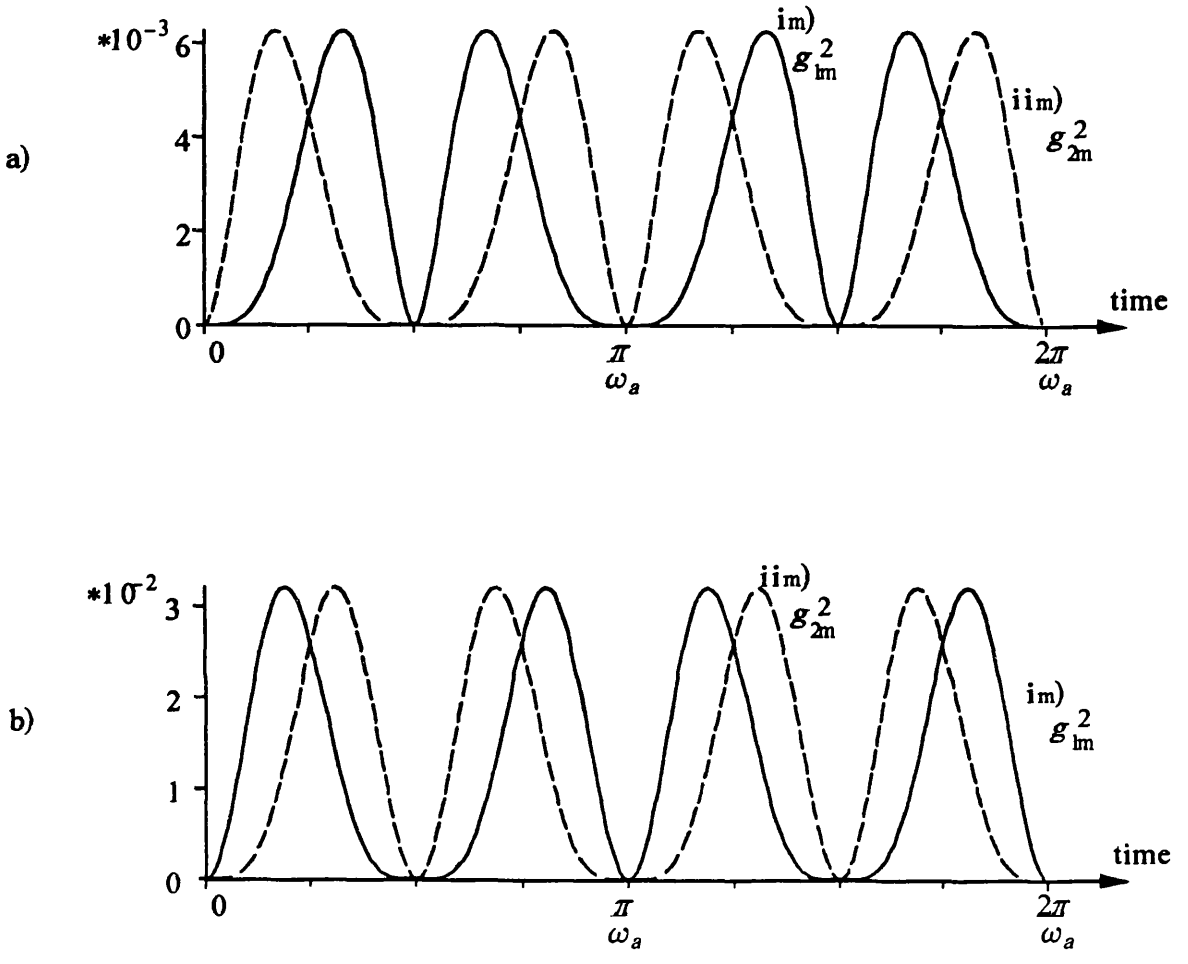


Figure 5–7 Plots of intermodulation terms for a) triangular and b) sine/ cosine electrical modulations. Trace im) shows g^2_{1m} and trace iim) shows g^2_{2m}.

Detector D2 responds to the light of four intermodulation and one central beam, all contained within μ_1 , so that the intensity response is,

$$I \propto \left[g_1^2 r_1^2 + g_2^2 r_2^2\right] + \left[g_{1m}^2 g_1^2 r_1^2 + g_2^2 g_{2m}^2 r_2^2\right] + \left[g_{1m}^2 g_1^2 (r_1')^2 + g_2^2 g_{2m}^2 (r_2')^2\right] \quad ...(5-34)$$

The first bracket is due to primary modulations The second bracket is from the intermodulation orders μ_{1b} , μ_{1c} shown in figure 5-6*b* , and will contribute some differential intensity response at frequency $2\omega_a$. Similarly the third bracket is due to the intermodulation beams μ_{1b} , μ_{1c} shown in figure 5-6*c* and will result in a differential intensity signal approximately equal to that from the second bracketed terms. (The differential response of the last bracket is from a slightly different area to the first two bracketed terms, hence the primed labels on r_1 and r_2). The magnitude of the intermodulation terms is however about three orders of magnitude less than the 'regular' modulation of the first bracket, and so it is clear that the intermodulation beams will not significantly affect the differential response from detector D2.

The fact that the intensities of the intermodulation beams (figure 5-7) are less for triangular than sine/cosine electrical modulation is an interesting observation in itself. Since both types of electrical modulations were of the same amplitude, it therefore follows that the mean acoustic power in the Bragg cell is greater for the sine/cosine modulation than the triangular electrical modulation. This is the reason for worse intermodulation when using sine/cosine electrical modulation.

Minimization of the intermodulation beam intensities is achieved by ensuring that there is significant power only ever in one acoustic wave, at any moment in time. For example

square modulation of each RF carrier would be 'ideal' since the Bragg cell will then operate as if it is being driven by a single RF frequency, and it would thus be possible to diffract all the optical power into the first order. This type of modulation however will result in no interference (hence no differential phase response) and so is obviously not any use in the present system. Nevertheless it is particularly relevant for a further implementation of the differential intensity and phase technique discussed in chapter 8 under further work.

5.3.4 The effect of intermodulation products on the differential phase response

The intermodulation beams will affect the differential phase response. Referring to figures 5-6*b,c*, the interference response at frequency $2\Delta\omega$ from light within beams μ_{1a}, μ_{3b}, μ_{2b}, μ_{1c} and μ_{1b} is (in order of appearance),

$$I_{2\Delta\omega} \propto \left| g_1^2 g_2^2 \right|_{dc} \cos\left(2\Delta\omega t + 2\,\underset{\kappa_{1a}\,\kappa_{2a}}{\delta\theta_1} \right)$$

$$+ \left| g_1 g_2 g_{1m} g_{2m} \right|_{dc} \left\{ \cos\left(2\Delta\omega t + 2\,\underset{\kappa_{2b}\,\kappa_{2a}}{\delta\theta_2} \right) + \cos\left(2\Delta\omega t + 2\,\underset{\kappa_{1b}\,\kappa_{1a}}{\delta\theta_3} \right) \right.$$

$$\left. + \cos\left(2\Delta\omega t + 2\,\underset{\kappa_{2b}\,\kappa_{1a}}{\delta\theta_4} \right) + \cos\left(2\Delta\omega t + 2\,\underset{\kappa_{2a}\,\kappa_{1b}}{\delta\theta_5} \right) \right\}$$

...(5 - 35)

where the labels under $\delta\theta_1$, $\delta\theta_2$ etc. indicate which set of probing beams were responsible for the respective phase term.

The first term is the expected interference signal holding the phase difference from κ_{1a} and κ_{2a}. The other four terms are from intermodulation beams. It may noticed that

$\delta\theta_2$ and $\delta\theta_3$ depend on light exclusively from within either κ_1 or κ_2, and so will be much smaller than any of the other differential phase terms.

Figure 5-8*a,b* show $(g_1\, g_2 g_{1m} g_{2m})$ for both triangular and sine/ cosine electrical modulations. From the graphs, the dc levels are approximately $4\text{x}10^{-4}$ and $2\text{x}10^{-3}$. It is therefore apparent that the triangular modulation is some 5 times less susceptible to the effects of the intermodulation beams.

The effect of the additional phase terms in equation (5 – 35) is to slightly modify the measured phase. The value of $|g_1^2\,(t)\;g_2^2\,(t)|_{dc}$ is $4.4\text{x}10^{-2}$ (as given in section 5.3.2), and therefore for triangular electrical modulations, the dc component of each intermodulation beam is a factor of 100 less than that from the main beam μ_{1a}. The net effect will be to change the measured phase value by at the most 2% (only the μ_{1b} and μ_{1c} beams are important). In this case, detecting the side beams, μ_2 and μ_3 will have no significant effect on the differential phase measurement.

Interference at frequency $2\Delta\omega$ may also arise from mixing of light after the second pass through the Bragg cell as follows :

interfering beams	$2\Delta\omega$ component
$\mu_{3a} * \mu_{1c}$	$\left\| g_1^2 g_{1m}\, g_2 \right\|_{dc} \cos\left(2\Delta\omega t + 2\,\delta\theta_{1\,\kappa_{1a}\,\kappa_{2a}} \right)$
$\mu_{2a} * \mu_{1b}$	$\left\| g_1\; g_{2m}\, g_2^2 \right\|_{dc} \cos\left(2\Delta\omega t + 2\,\delta\theta_{1\,\kappa_{1a}\,\kappa_{2a}} \right)$
$\mu_{3a} * \mu_{1b}$ †	$\left\| g_1^2 g_{1m}\, g_2 \right\|_{dc} \cos\left(2\Delta\omega t + 2\,\delta\theta_{1\,\kappa_{1b}\,\kappa_{2a}} \right)$
$\mu_{2a} * \mu_{1c}$ †	$\left\| g_1\; g_{2m}\, g_2^2 \right\|_{dc} \cos\left(2\Delta\omega t + 2\,\delta\theta_{1\,\kappa_{1a}\,\kappa_{2b}} \right)$
$\mu_{1a} * \mu_{3b}$ †	$\left\| g_{2m}\, g_1^3 \right\|_{dc} \cos\left(2\Delta\omega t + 2\,\delta\theta_{4\,\kappa_{2b}\,\kappa_{1a}} \right)$
$\mu_{1a} * \mu_{2b}$ †	$\left\| g_{1m}\, g_2^3 \right\|_{dc} \cos\left(2\Delta\omega t + 2\,\delta\theta_{5\,\kappa_{2a}\,\kappa_{1b}} \right)$

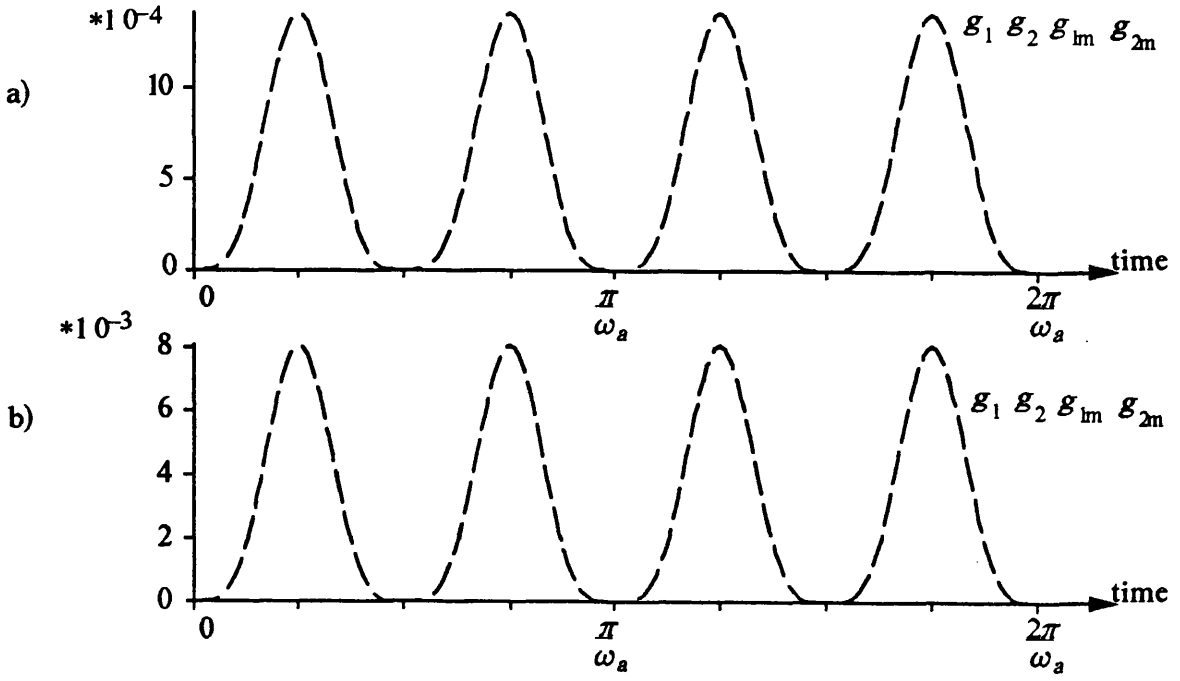


Figure 5-8 Plots of ($g_1\ g_2\ g_{1m}\ g_{2m}$) intermodulation term for a) triangular and b) sine/ cosine electrical modulations.

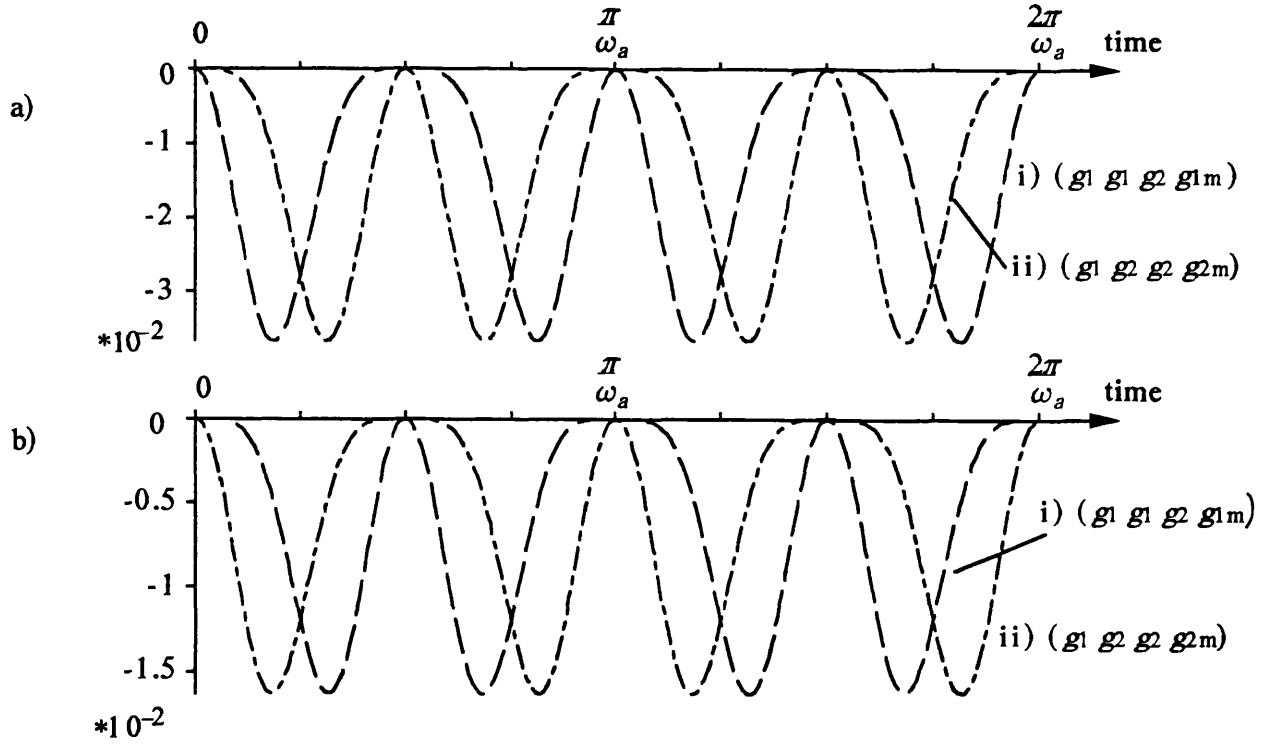


Figure 5-9 Plots of intermodulation terms for a) triangular and b) sine/ cosine electrical modulations. Trace i) shows $g^2_1 g_2 g_{1m}$ and trace ii) shows $g_1 g^2_2\ g_{2m}$.

All the beam components are on figure 5-6*b* except those denoted by † which are on figure 5-6*c*. This type of interference never involves more than one component from each cluster of beams within μ_1, μ_2 and μ_3, and so will only affect measurement from detector D3.

Figure 5-9*a,b* show $(g_1^{\,2}\, g_2\, g_{1\,m})$ and $(g_1\, g_2^{\,2}\, g_{\lambda m})$ for both types of electrical modulations. This time triangular modulation is a factor of two worse than sine/ cosine modulation, and the dc values are approximately, 4×10^{-3} and 8×10^{-3}. The functions, $(g_1^{\,3}\, g_{2\,m})$ and $(g_2^{\,3}\, g_{1\,m})$ have not been plotted, but their dc values are at least an order of magnitude less than that of $(g_1^{\,2}\, g_2\, g_{1\,m})$ and $(g_1\, g_2^{\,2}\, g_{\lambda m})$.

Interference *between* rather than *within* the beams can have more serious ramifications. Phase cancellation effects as discussed in chapter 3 (section 3.4.2) will mean that interference intensity is less than the values given above by perhaps a factor of 10. Also the interference intensity will depend upon the angular spreading of each of the constituent interfering beams, further reducing the effect. However what is of most concern is that the measured phase will now be influenced by movement of the detector. This may result in phase jitter not previously expected in this common path system, when optical components are moved. In order to avoid potential problems, it is always important to avoid mixing light from within μ_1, μ_2 and μ_3. This does not prohibit detector D3 from being used, but merely means that care must be exercised in its use.

It is assumed in all these discussions that light is always at Bragg incidence so that the series solution of the modulation functions given in chapter 4 are valid. The effect of being off Bragg incidence will be to reduce the amplitude of diffracted light and so too, any consequential interference. This may have an adverse effect on system performance depending upon how the relative intensities of the various interference terms discussed above are affected.

5.4 The use of feedback to improve the differential intensity measurement accuracy

It has become apparent that in order to achieve a good differential intensity resolution, the modulations of the beams must be extremely tightly controlled. In an experimental set-up there are three important ways in which the modulations of the beams are controlled. Rotation of the Bragg cell will reduce the amplitude of the incident and emerging light which is not at the exact Bragg angle. This effect is described by the A_1, A_2, a_1 and a_2 coefficients, the value of which will depend upon the design of the Bragg cell. These may in principle be evaluated using the acousto-optic theory of chapter 4, taking into account the finite sizes of the Bragg cell and transducer. The remaining two controls are variation of the phase and amplitude of the electrical modulations.

In the discussions of earlier sections, consideration has been given to the situation when two of the parameters are correct and one is slightly wrong. On an experimental rig, some method is required to test whether each of the conditions are met and accordingly make adjustments. Herein lies a problem because there is no straightforward means by which all three conditions can be individually tested. Indeed, the question arises, is it really necessary to have each condition individually fulfilled? Could not perhaps an error in the splitting coefficients be compensated by an amplitude error in the electrical drive modulations.

To answer these practical problems, it is necessary to turn back to the analysis of chapter 3. We will consider the response of an extra detector, D_{mon} as shown in figure 5–10, which monitors the total intensity of the two probe beams before they are incident on the sample. Just one condition will be imposed, that the $2\,\omega_a$ frequency component from this detector is zero. Referring to equation (3 – 24), which is the frequency response of detector D1 and putting $r_1 = r_2 = 1$, the following condition is being enforced :

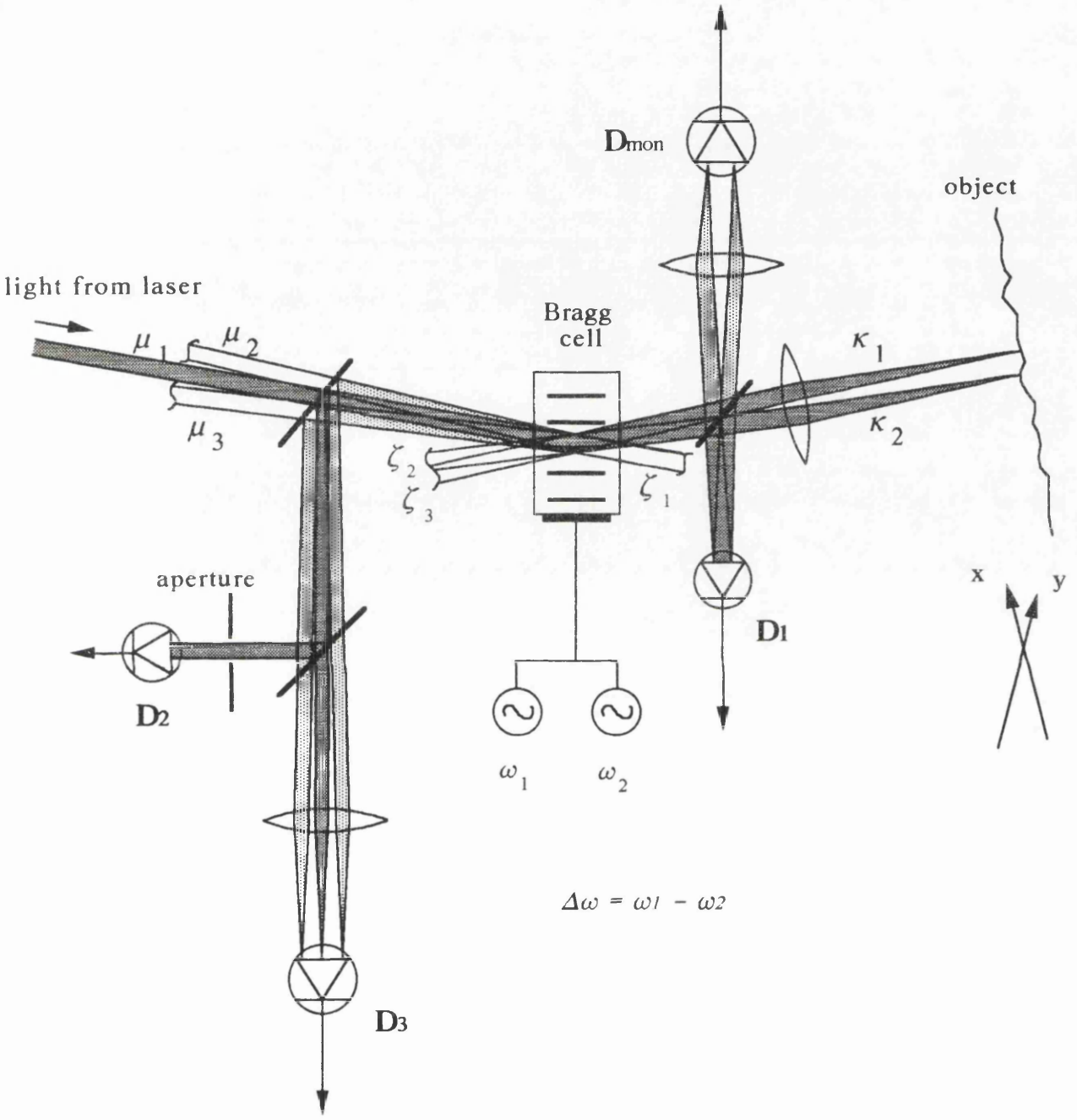


Figure 5–10 The two first order system with monitoring detector

$$\sum_{\substack{m,n=-\infty \\ for\ m+n=\pm 2}}^{\infty} \left(A_1^2 c_m c_n + A_2^2 d_m d_n \right) \delta(\omega - (m+n)\omega_a) = 0 \qquad \ldots(5-36)$$

Substituting this back into equation (3 – 24), the $2\omega_a$ frequency response of detector D1 is,

$$S_1(2\omega_a) = \frac{1}{2}\rho_1 R_L \zeta \left(r_1^2 - r_2^2 \right) A_1^2 \sum_{\substack{m,n=-\infty \\ for\ m+n=\pm 2}}^{\infty} c_m c_n \delta(\omega - (m+n)\omega_a) \qquad \ldots(5-37)$$

Using expressions (3 – 34) and (3 – 36), this becomes,

$$S_1(2\omega_a) = \rho_1 R_L \zeta A_1^2 \Delta R \, |e_8| \,.\, \mathrm{F}\{\cos(2\omega_a t - \varepsilon_8)\} \qquad \ldots(5-38a)$$

where
$$e_8 = \sum_{\substack{m,n=-\infty \\ for\ m+n=+2}}^{\infty} c_m c_n \qquad \ldots(5-38b)$$

and
$$\cos\varepsilon_8 = \frac{\Re(e_8)}{|e_8|} \qquad \ldots(5-38c)$$

The current response of detector D1 at frequency $2\omega_a$ is, (ignoring the phase term),

$$I_{2\omega_a}(t) = \rho_1 R_L \zeta A_1^2 \Delta R \, |e_8| \,.\, \cos(2\omega_a t) \qquad \ldots(5-39)$$

Hence the cancelling of all $2\omega_a$ components at D_{mon} is sufficient to ensure a response at detector D1 proportional only to ΔR ie., true differential intensity.

The validity of the differential intensity result from D1 may thus be constantly monitored and suitable adjustments made. In contrast, the situation for the second detector, D2 is more tricky. When the two beams pass through the Bragg cell for a second time, each probe interacts a further time with each acoustic wave. Consequently even if the $2\omega_a$ frequency component were nulled at detector D_{mon}, it is not sufficient to guarantee true differential intensity measurement from detector D2. This may be seen by examining (3 - 25), which is the frequency response of detector D2. It is apparent that the required condition for a true differential intensity response from D2 at frequency $2\omega_a$ is,

$$\sum_{\substack{m,n,q,r=-\infty \\ for\ m+n=\pm 2}}^{\infty} \left(a_1^2 A_1^2 c_m c_n c_q c_r + a_2^2 A_2^2 d_m d_n d_q d_r \right) \delta(\omega - (m+n+q+r)\omega_a) = 0 \qquad ...(5-40)$$

As is the case for detector D1, it is not necessary to have exactly identical modulations of each probe beam, providing the above condition is met. For detector D2 however, its fulfilment may be verified only by setting up the system with some 'reference' sample of zero reflectivity variation and making adjustments until the $2\omega_a$ frequency component at detector D2 is nulled. This has the disadvantage of prohibiting constant monitoring of the differential intensity measurement accuracy. Furthermore, although both methods are influenced by the evenness in sensitivity of the detectors over their light sensitive areas, the set up accuracy of D2 is also limited by the quality of the test sample.

Lastly we will investigate whether this 'nulling' method is of any potential use for measurements made from detector D3. Following the procedure of detector D2, using a reference sample and adjusting the system so that there is no $2\omega_a$ component at D3, the following condition is ensured :

$$\sum_{\substack{m,n,q,r=-\infty \\ for\ m+n=\pm 2}}^{\infty} \left\{ a_1^2 A_1^2 c_m c_n c_q c_r + a_2^2 A_2^2 d_m d_n d_q d_r \right\} .\delta(\omega-(m+n+q+r)\ \omega_a)$$

$$+ \{ b_1^2 A_1^2 + b_2^2 A_2^2 \} \sum_{\substack{m,n,q,r=-\infty \\ for\ m+n=\pm 2}}^{\infty} c_m c_n d_q d_r\ \delta(\omega-(m+n+q+r)\ \omega_a)$$

$$= 0 \qquad \ldots(5-41)$$

Substitution of this back into (3 – 26) reveals that it is an insufficient condition to guarantee true differential intensity measurement. So this 'nulling' method is not applicable for measurements made from detector D3.

5.4.1 Discussion

We have seen that the use of an extra detector D_{mon} allows the differential intensity measurement accuracy from detector D1 to be constantly monitored. It is no longer necessary to be individually concerned about the phase and amplitudes of the modulations and the splitting/ recombination coefficients. Rather it is sufficient just to null the $2\omega_a$ frequency component as measured from D_{mon}. The use of a reference sample for setting up detector D2 is much less desirable, and the method is not applicable for detector D3. Detector D1 is therefore the preferred detector for differential intensity metrology.

Even though it is not strictly necessary to modulate the two beams in an identical manner, intuitively it is expected that the system would be most stable when the modulations are as similar as possible. Fine adjustments can be continuously made using D_{mon} to provide a feedback signal. The feedback may operate on the amplitude or phase of the electrical modulations, or the orientation of the Bragg cell (probably the least favoured option).

The accuracy that this feedback method can achieve will be limited by noise. One source of noise arises from mechanical vibration of the Bragg cell relative to the incident light (microphonics). The variable γ introduced in section 5.1 describes the splitting/ recombination mismatch and is determined by the relative angle of the propagating acoustic and light fields, and the acoustic field distribution within the cell. In chapter 4 it was seen that a necessary condition for the Bragg interaction is that the Klein-Cook parameter Q (equation (4 - 14)) is much greater than unity ie., a long interaction length. This is necessary to suppress the multitude of beams characteristic of Raman-Nath scattering. Increasing the Q factor has the consequence of making the Bragg interaction more angularly sensitive as shown by equations (4 - 34*a*,*b*), and may even result in the incident light being diffracted by only one of the acoustic frequencies. This can be overcome by using a phased array type device where the two acoustic waves are propagating in the Bragg cell at slightly different angles so that the incident light is at the Bragg angle for both of them.

This does not answer the question as to the effect on the amplitudes of the beams if there is vibration of the Bragg cell relative to the light. If the Bragg cell interaction is so sensitive to variation in angle, perhaps it would be better to detune the Bragg cell. Reducing the interaction length and increasing the angular spectrum (or divergence) of the acoustic waves (which as mentioned in the previous chapter occurs naturally in a real Bragg cell since it is operated within the Fresnel distance), will have the desired effect. Clearly any reduction in Q must not result in any significant amount of power being diffracted into extra 'Raman-Nath' type beams.

A reduction in Q effectively increases the bandwidth of the cell. Some compromise must be found so that vibrations have negligible effect on the value of γ while the acousto-optic interaction remains well enough within the Bragg regime. A further consideration of probably lesser importance is that the acoustic field must be constant across the width of the incident optical beam.

It should be noted that the bandwidth to which reference is made has nothing to do with the bandwidth of the transducer which is on the side of the cell, and generates the sound waves. Here we are concerned only with the bandwidth which is due to phase cancellation of light at different angles of incidence, and this distinction may not be immediately obvious in the specifications of an 'off the shelf' device which is used in the system.

Having chosen a Bragg cell in which microphonics induced variations in γ is acceptably small, the next concern is noise within the Bragg cell drive electronics and compensation circuitry of the feedback loop. The results of figures 5-3 and 5-4 in section 5.3.1 describe the expected differential intensity resolution for a particular amount of phase or amplitude noise in the electrical modulations. These figures should be used in the design of the Bragg cell drive and feedback electronics.

5.5 Summary

In this chapter we have looked at the factors which limit the differential intensity and phase sensitivities. The most troublesome aspect of the system is the modulation necessary for the differential intensity response. Exactly how precise this modulation must be in order to attain a given sensitivity is a most important question and considerable effort has been afforded to its solution. Both graphical and analytical results have been presented showing the quantitative dependence of the differential intensity resolution on the splitting ratio of light at the Bragg cell and the electrical drive modulations.

Having designed electronics which can modulate the Bragg cell drive with sufficient accuracy, some means must be found of verification. For this purpose, the use of an extra detector has been discussed which will provide continuous monitoring of the modulations. It is envisaged that this will form part of a feedback loop and ensure that any necessary adjustments are continuously made to the modulations. Since this feedback method may only be used for measurement from detector D1, this is the clear choice for differential intensity metrology.

Calculations of the differential phase sensitivity are, by comparison, far simpler, and it was shown that the phase resolution is 1×10^{-6} rad in a 1 Hz bandwidth. This is for detectors D2 and D3, there being less than 10% difference between the two detectors, and assumes shot limited noise.

The effects of intermodulation beams on both differential responses has been studied. In general, sine/ cosine electrical modulations generated more intense intermodulation beams than triangular modulations. However for both responses, the resulting errors in each differential response is typically less than 2%, and so not a serious problem.

CHAPTER 6

EXPERIMENTAL WORK

6 Experimental work

The earlier chapters have described the principle of the differential intensity and phase measurement technique, and in particular, theory outlining the necessary modulation criteria. In conjunction with the development of this theory, three different optical systems have been built and measurements taken.

The first setup used the zeroth and first order beams as probes, and the second and third were implementations of the two first order system. Each used the same Bragg cell (Isomet type 1205C-2 $PbMoO_4$; with 13mrad divergence of zeroth first orders @ 80 MHz) and laser (Melles Griot 05LHP151, λ = 632.8Å, 5mW output power, linear polarization, 0.8mm beam diameter with 1mrad divergence). In the zeroth/ first order system, an off the shelf Bragg cell driver was used (Isomet type 232A-1), and for the latter two setups, the required Bragg cell drive signals were synthesized using separate frequency sources, mixers and amplifiers. For all three arrangements, triangular modulations were used ; the frequency was 10 kHz for the zeroth/ first order system and around 100kHz for the latter two arrangements.

The chapter is divided into 5 sections. Results from each experimental setup will be presented in the first three sections. The final section will present the most recent implementation of the system and describe practical alignment techniques. The results shown in this chapter were all taken before completion of the theory outlined in earlier chapters, and as such do not make full benefit of their conclusions.

6.1 Results from the zeroth/ first order system

The system arrangement is illustrated in figure 3–1. A single 3cm focal length lens was used, resulting in a focussed beam diameter[1] on the object surface of 28μm and a separation of 390μm. The differential intensity measurements were made using a detector which monitored the light before it passed a second time through the Bragg cell, in a similar manner to detector D1 in the two first order system (figure 3–2). This was done in order to improve the measurement accuracy.

Two complementary samples were examined, one with almost pure topography and the second with primarily reflectivity detail. Figure 6–1 shows differential phase and intensity line traces across a silicon wafer whose surface has been active plasma etched to produce a series of parallel tracks as shown in figure 6–1*a*. The nominal step heights are 180, 300, 400, 550 and 700 Å, which were confirmed to within ±10% using a mechanical stylus probe.

The differential phase result[2] (figure 6–1 *b*) shows not only peaks corresponding to the edges of the etched tracks, but also an overall phase change across the wafer which is due to wafer warp. This background variation has been subtracted to produce the trace 6–1*c* which can be seen to comprise a series of up and down peaks. The true differentiating nature of the system is immediately apparent since the direction of the peaks depend upon whether the step has a negative or positive gradient. The height of each peak corresponds to the extra optical path length imposed when the beams are either side of the step. Since the light is reflected from the surface, a physical step height of for example half an optical wavelength will result in a measured phase change of 2π radians. This assumes there is no compositional difference between the two probed areas, which is a valid assumption for this particular sample. The differential phase measurements are shown in table 6–I and compare well with those made by a mechanical stylus probe.

---

[1]All measurements of beam diameter in this chapter refer to the distance at which the intensity falls to half its maximum value.

[2]It should be noted in this chapter, that all graphs showing differential phase have a relative rather than an absolute scale.

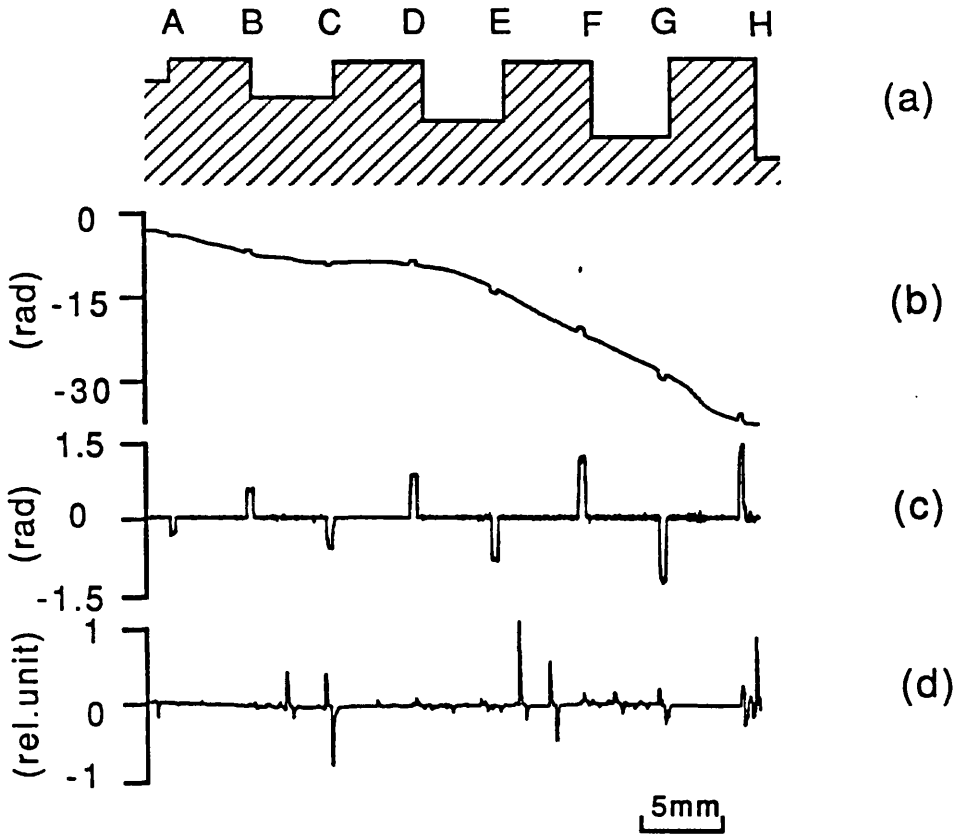


Figure 6-1 Differential optical phase and intensity line traces measured by zeroth/ first order system across an etched silicon wafer. (a) etched silicon surface, (b) differential phase, (c) differential phase (background variation subtracted), and (d) differential intensity.

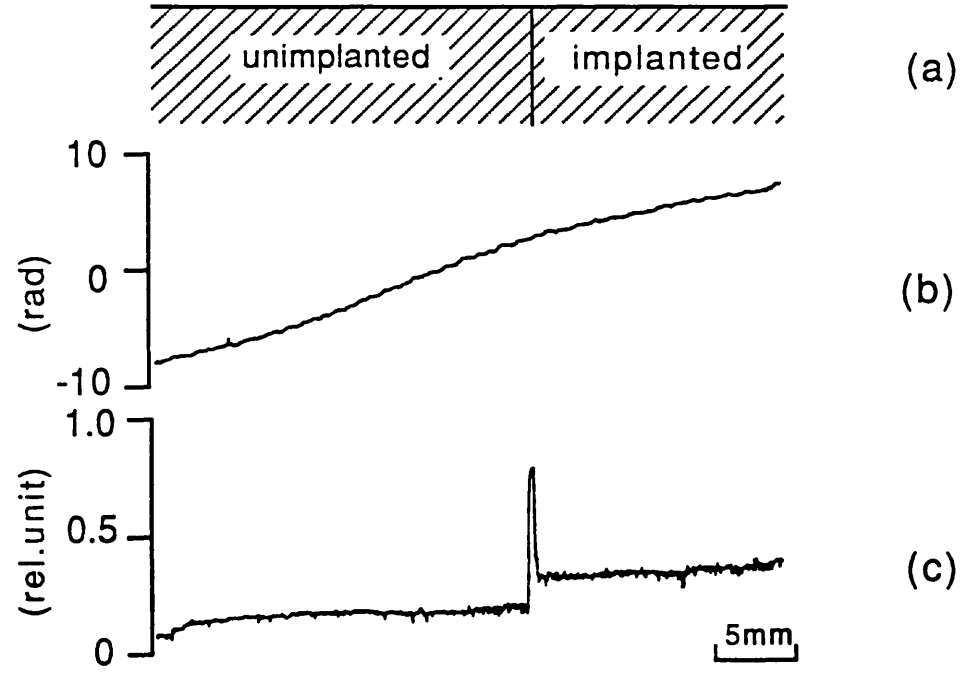


Figure 6-2 Differential optical phase and intensity line traces measured by zeroth/ first order system across a partly implanted silicon wafer. (a) silicon sample, (b) differential phase, and (c) differential intensity.

Step	A	B	C	D	E	F	G	H
Stylus probe	180	310	315	430	440	530	540	700
Zeroth/ first order	160	316	328	449	457	594	586	722
Two first order	194	293	267	437	468	589	567	—

Table 6-I Comparison of step heights measured with stylus probe, zeroth/ first order and two first order differential phase systems. All measurements are in Angstroms.

The width of each peak is equal to the probe beam separation and will reduce as the two probe beams are brought closer together, that is, the system would become a better differentiator. It can be noticed that each peak has a flat top. This is because the separation of the two beams is much greater than the dimensions of the feature responsible for the phase contrast, ie., the step. For measurement of step height, it is important that the probe beams are well clear of the interface.

The differential intensity result is shown in figure 6-1*d*. Differential contrast on this sample is caused by scatterers such as step discontinuities, scratches, surface roughness from the etching process and surface contamination (eg. dust). The characteristic differential contrast from a point object is a pair of bipolar peaks with a separation equal to the distance between the two focussed probe beams. The width of each peak is determined by the focussed beam diameter.

Figure 6-2 shows differential phase and intensity line traces across a silicon wafer, half of which is implanted with As^+ at a doping level of 10^{13} ions/cm^2 (see figure 6-2*a*). Looking first at the differential phase result in figure 6-2*b*, it can be seen that any contrast caused by a change in the phase angle of the reflection coefficient is far smaller

than that due to surface topography. The overall slope of the line trace is indicative of warp of the wafer surface. The differential intensity result is shown in figure 6-2 *c*. Here the change in reflection coefficient is clearly visible. Also however, the dc levels on either side of the interface are different and non-zero. This is attributable to unequal modulation in each of the probe beams.

At the time that these results were taken, there were severe difficulties encountered in achieving symmetrical modulations in each of the probe beams. It was found that the incident light was divided not only into a zeroth and first order, but that 2nd and 3rd order beams of lower intensity were produced. There are a number of possible explanations for these extra beams:

i) the interaction length of the Bragg cell is too short ;

ii) reflections of the acoustic wave in the Bragg cell ;

iii) multiple reflections of the optical beam within the Bragg cell ;

iv) a distorted Bragg cell drive signal containing RF frequency harmonics.

Further experimental observations discounted ii) and iii). It was found that on swapping the Bragg cell drive for an alternate unit, the same effect was manifest and this seemed to preclude point iv). It was therefore concluded that i) was the most likely explanation.

In more recent experiments, the same Bragg cell has been used, but instead of the standard Bragg cell drive unit, the necessary drive signals have been generated using separate synthesizers and amplifiers. Moreover it has been found that the above problems are no longer prevalent, suggesting that many of the previously encountered difficulties may be attributed to the Bragg cell drive unit. This experience shows that if a Bragg cell is required for accurate intensity modulation, it is *always* important to be aware of the frequency content of the Bragg cell driving voltage when under load.

It should therefore be possible to improve the differential intensity measurement accuracy above that demonstrated in the results of figure 6–2*c*. There are none the less other difficulties associated with this system. For metrology applications, it is important to have the two probe beams separated on the object surface. If the separation is too great, then additional effects are included in the measurement which are not of immediate interest for example warp of the wafer surface. The effect of wafer warp may be subtracted from the measured values, but this is inconvenient and it would be far more desirable if the ratio R_{ss} of probe beam separation to spot size may be reduced. Currently this is equal to 13.5, and a value of 3 or 4 would be better suited for this type of metrology application. Of course the measurement of wafer warp is a measurement in its own rights, but at present we are concerned with features of smaller lateral dimensions.

In order to reduce R_{ss}, it is necessary to reduce the Bragg cell drive frequency. It is apparent from equation (4 – 14) that if the interaction is to remain true Bragg, then there is an inverse square relation between the required length of the Bragg cell and Bragg cell drive frequency. In short, a reduction in R_{ss} of 4 necessitates a 16 fold increase in the required length of the Bragg cell.

Such a method of reducing R_{ss} is far from practical especially for imaging applications where a value less than unity is required. In order to surmount this problem the concept of the two first order system was developed. This latter technique offers much greater versatility and all subsequent experimental work has been concerned with this method.

6.2 Results from the two first order system – 1

The two first order arrangement is shown in figure 3-2. The differential phase measurements were made using detector D3 and the differential intensity, using detector D1. The beam separation and focussed beam diameter on the sample was 260μm and 240μm, resulting in a R_{ss} value of 1.1 .

The system was used to examine the two samples described in the previous section. It can be seen that the differential intensity results from the implanted silicon wafer (figure 6-3*a*) are an improvement over those from the zeroth/ first order system. The dc levels on either side of the interface are now equal, and this can be attributed solely to the modulation of each beam being more identical. The differential phase results (figure 6-3*a*) show again that any differential phase contrast caused by the ion implantation is smaller than that from surface topography.

The differential intensity peak is now triangular in shape compared to the previous square peak from the zeroth/ first order system. As the first probe beam moves across the interface, the differential intensity response will gradually increase. However no sooner has one probe crossed the interface than the second also starts to cross, and the differential intensity response falls off. Hence there is no longer a flat top to the peak. The slope of the sides will be dependent on the focussed beam diameters.

Figure 6-4 shows results from the plasma etched silicon sample. The differential phase response (figure 6-4 *a*) again has an underlying variation due to warp of the wafer, and this has been subtracted to reveal trace 6-1 *b*. As was seen in the results from the implanted silicon, the peaks are now triangular in shape.

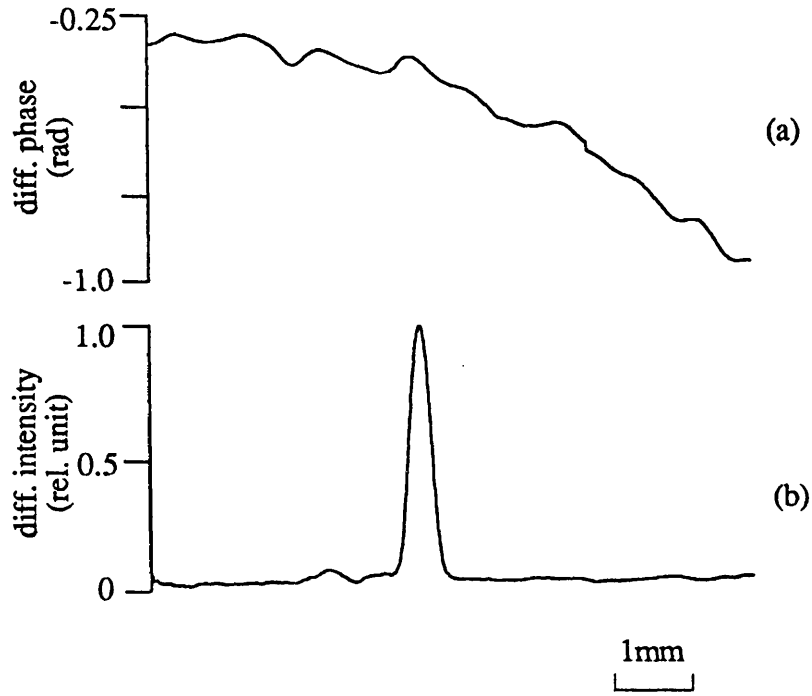


Figure 6-3 Differential optical phase and intensity line traces measured by two first order system across a partly implanted silicon wafer. (a) differential phase, and (b) differential intensity.

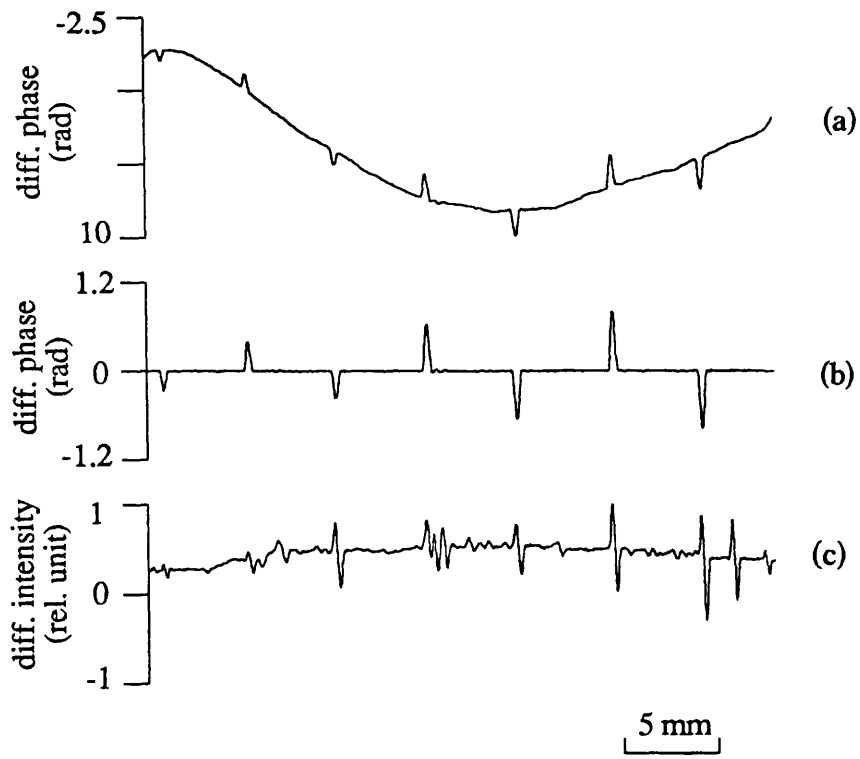


Figure 6-4 Differential optical phase and intensity line traces measured by two first order system across an etched silicon wafer. (a) differential phase, (b) differential phase (background variation subtracted) , and (c) differential intensity.

The heights of the peaks have been measured and are shown in table 6-I. Since the value of R_{ss} is not less than unity, it would be expected that the measurements from the zeroth/ first order and two first order systems should agree. Where there are significant differences between the two sets of measurements, this is most likely due to a lack of uniformity of the specimen.

The differential intensity result is shown in figure 6-3*c*. It can be seen that the dc level is still non-zero and fluctuates about a mean value. There are two prime reasons for this . First, warp of the wafer surface caused the light to move across detector D1. This may be overcome using a larger area detector. The second reason is that deficiencies in the Bragg cell drive electronics meant that that the modulations of each beam did not perfectly satisfy the necessary criteria as defined in chapter 3.

6.3 Results from the two first order system - 2

For this last set of results, the electronics were redesigned allowing better control over the modulation of each RF frequency. In order to facilitate the use of a shorter focal length objective lens, two further lenses were used to project the point of divergence in the Bragg cell further down the system as described in section 3.1 . The calculated focussed beam diameter was 4.7μm, and the probe beam separation was 6.3μm, resulting in a value for R_{ss} of 1.3 . The beam separation is increased above this value for a few results presented at the end of the section. Both the differential phase and intensity measurements were taken from detector D3.

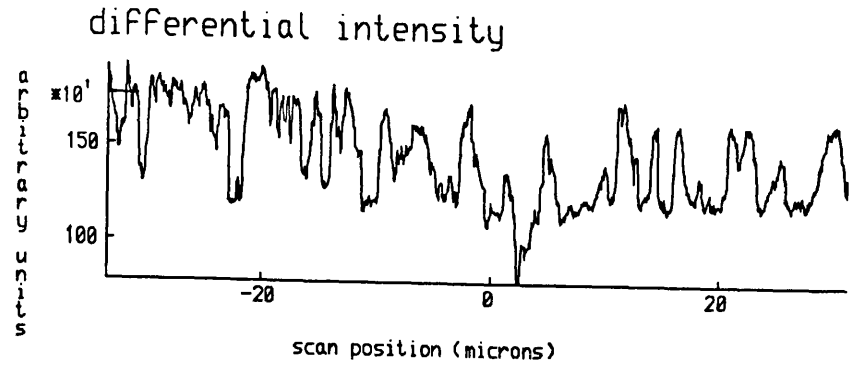
Results from the etched silicon sample, the ion implanted silicon sample, and a third sample consisting of a silicon wafer with aluminium deposited on one half will now be discussed in the next three sections.

6.3.1 Results from the two first order system – 2 : etched silicon sample

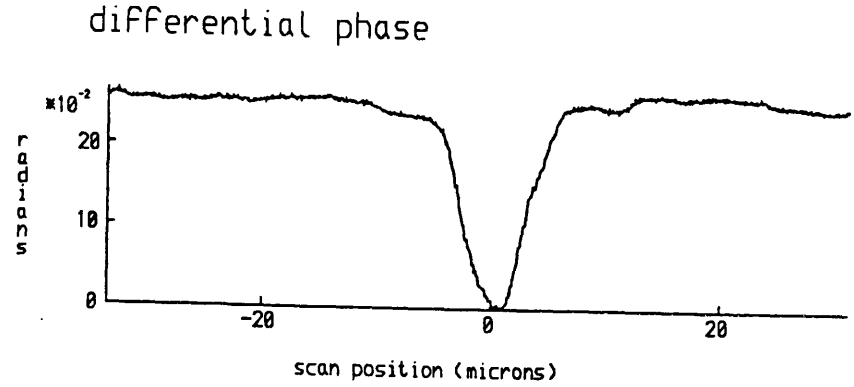
Figure 6–5 shows line trace measurements over the 180Å step of the etched silicon sample. With the the reduction in both the the focussed beam size and separation, there is a considerable improvement in lateral resolution. The differential phase result is shown in figure 6–5 *b*. The height of the peak corresponds to a step height of 129Å which is less than the expected value. This is most likely caused by the focussed beam diameter being somewhat larger than the diffraction limited value of 4.7μm, so that the spots partially overlapped.

The differential intensity measurement is shown in figure 6–5 *a*. At the step it can be seen that there are a pair of bipolar peaks as each probe beam individually crosses the step. Each is aligned with the sloping sides of the differential phase response. With the reduction in the focussed beams diameters used for these measurements, there is considerably finer differential intensity detail. It is also apparent that there is still some absolute intensity information contained in the response. Despite improvements in the electronics since the last series of measurements, there is still room for improvement. It was also shown in chapter 3 that the differential intensity response of detector D3 is likely to include a small absolute intensity component, which is still present even when all the modulation conditions are satisfied. These measurements were however performed using detector D3 before such conclusions were known.

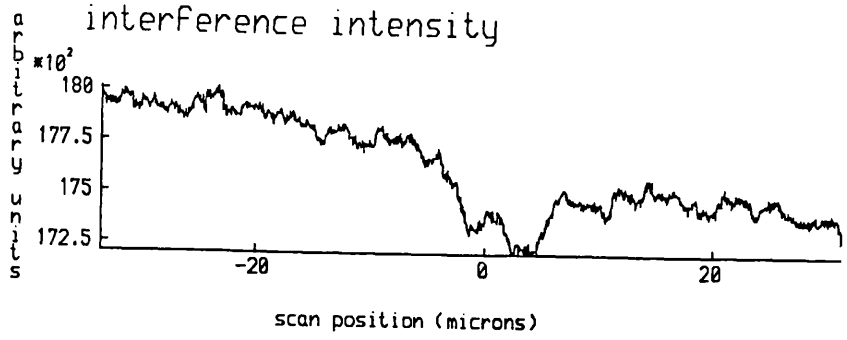
A third measurement which is shown in figure 6–5 *c* is the intensity of the $2\Delta\omega$ interference signal. This is proportional to the product of the reflectivities experienced by each beam. It can be seen that whilst the first probe is crossing the edge of the step, there is a big drop in the interference intensity. The intensity increases somewhat when the probes are either side of the step edge (this corresponds to the minimum point on the differential phase response). There is a further reduction in interference intensity as the second probe crosses the step. Finally the interference intensity increases when both probes are on the same side of the step.



(a)



(b)



(c)

Figure 6-5 Differential optical phase and intensity line traces measured by two first order system across step A of etched silicon wafer. (a) differential intensity, (b) differential phase, and (c) interference intensity.

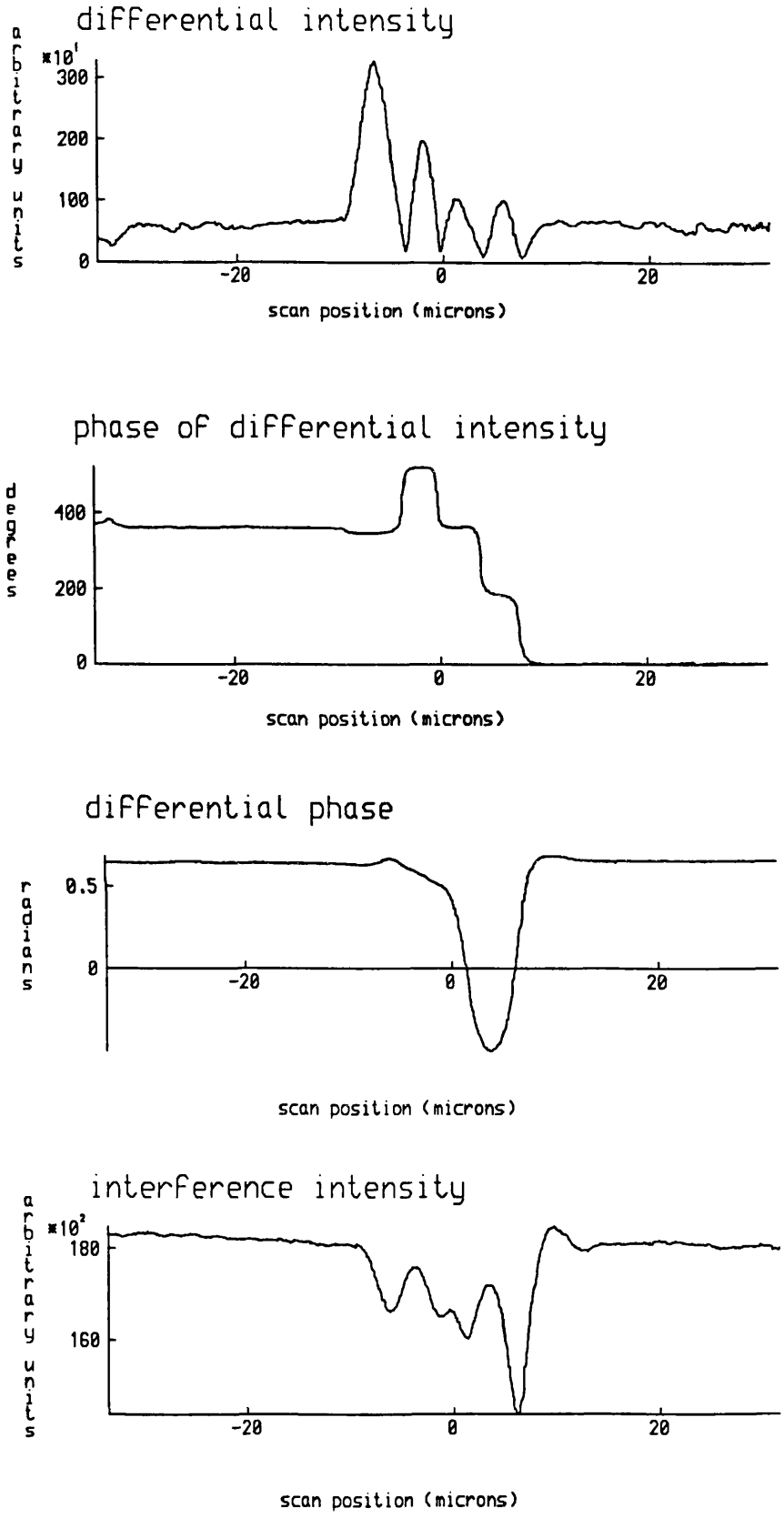


Figure 6-6 Differential optical phase and intensity line traces measured by two first order system across step G of etched silicon wafer. (a) differential intensity, (b) phase of differential intensity, (c) differential phase, and (d) interference intensity.

Line trace measurements from a deeper track (G) are shown in figure 6–6. The differential intensity is shown in figure 6–6 *a* and the phase of the differential intensity, in figure 6–6 *b*. The differential intensity measurement is made using a lock-in amplifier to monitor the $2\,\omega_a$ frequency component of detector D3. The lock-in amplifier will not give a negative output, and instead it is necessary to monitor the phase output in order to determine the sign of the differential response. Hence it can be seen that each peak in figure 6–6 *a* is alternately positive and negative, ie., two pairs of bipolar peaks.

So the differential intensity response suggests that the object comprises two strong scattering objects, and these are likely to be two consecutive steps. The differential phase results (figure 6–6 *c*) do indeed provide further evidence along these lines since the response now has a leading sloped edge. To draw further conclusions is difficult without a more detailed use of imaging theory or using higher resolution optics.

The interference intensity is shown in figure 6–6 *d* and this shows a series of troughs as each beam moves across each of the step edges.

6.3.2 Results from the two first order system – 2 : ion implanted silicon sample

Results from the ion implanted silicon sample are shown in figure 6–7. It is apparent from the differential intensity response (figure 6–7 *a*) that there is some sort of structure at the edge of the implanted region. It is likely that this is a consequence of the fabrication process.

The differential phase result (figure 6–7 *b*) is interesting because unlike the previous results from this specimen, the improved lateral resolution has allowed the effects of the interface to be resolved. A pair of bipolar peaks can be clearly discerned : as far as the

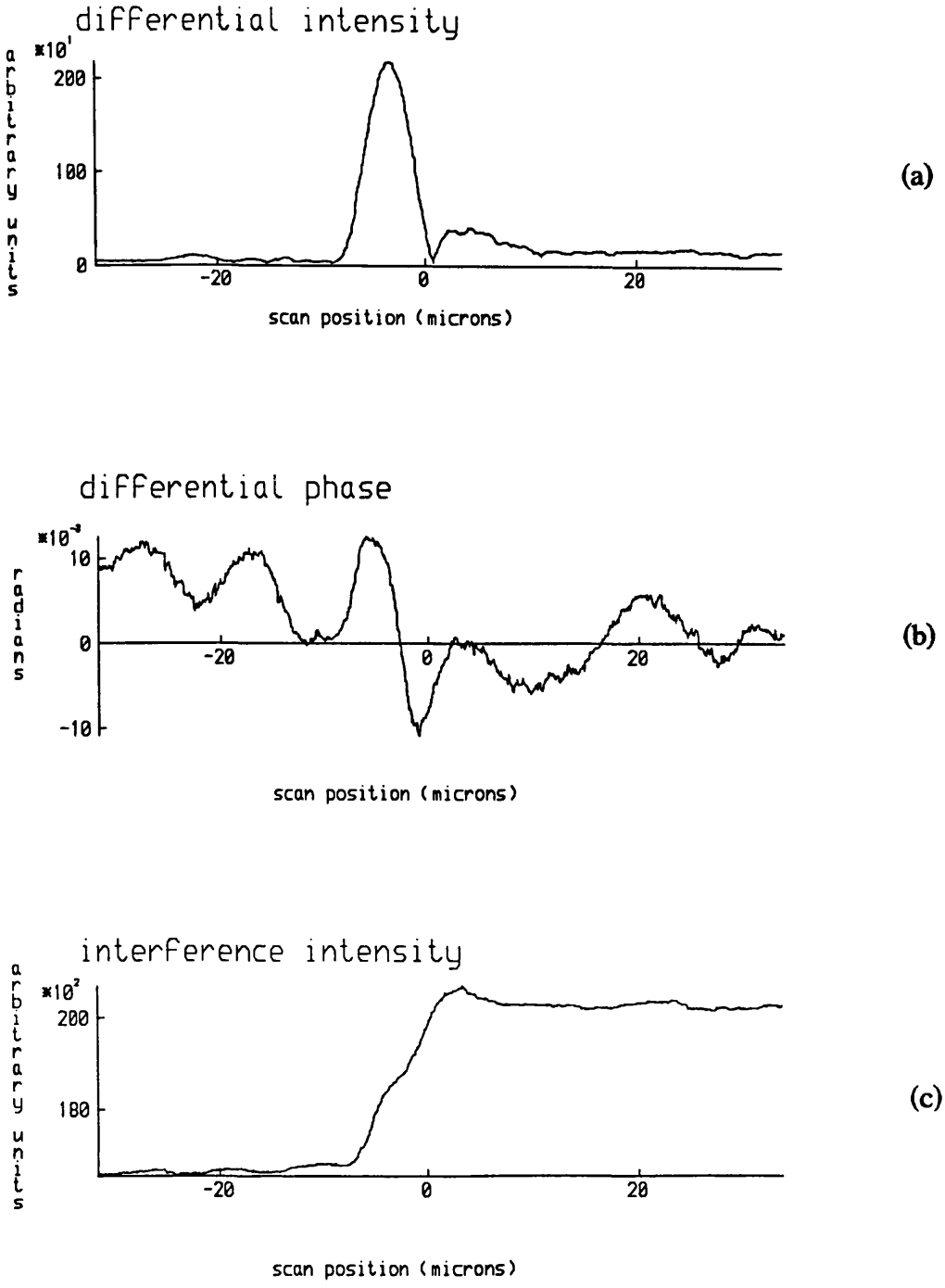


Figure 6-7 Differential optical phase and intensity line traces measured by two first order system across a partly implanted silicon wafer. (a) differential intensity, (b) differential phase, and (c) interference intensity.

differential phase response is concerned, the interface acts as like a point scatterer, perturbing the phase of the light as it scans across. The amplitude of the bipolar peaks is ±10 mrad.

Figure 6–7*c* shows the interference intensity. It can be seen that as each probe beam moves across the step, there is a rapid change in the response. Also when one beam is either side of the interface, the response partially levels out.

6.3.3 Results from the two first order system – 2 : aluminium/ silicon step

This set of results is from a silicon wafer which has had aluminium deposited on one half. The depth of the layer is nominally 500Å. As a consequence of there being both changes in height and conductivity, it is expected that there will be strong contrast in phase *and* intensity as the probe beams are scanned across the step.

Due to the method of fabrication, the aluminium/ silicon interface is not abrupt but rather graduated as shown schematically in figure 6–8. In order to gain some understanding of the phase and intensity responses to be expected from such a structure, it is useful to divide the step into three regions, labelled 1, 2 and 3.

In region 1 the thickness of the aluminium layer is greater than the skin depth, so there will be little differential intensity contrast. Variation in topography will lead to differential phase contrast.
In region 2 the thickness of the aluminium is approximately the same as that of the skin depth (50Å). This results in rapid variation in both surface conductivity and reflectivity and hence both differential phase and intensity contrast.

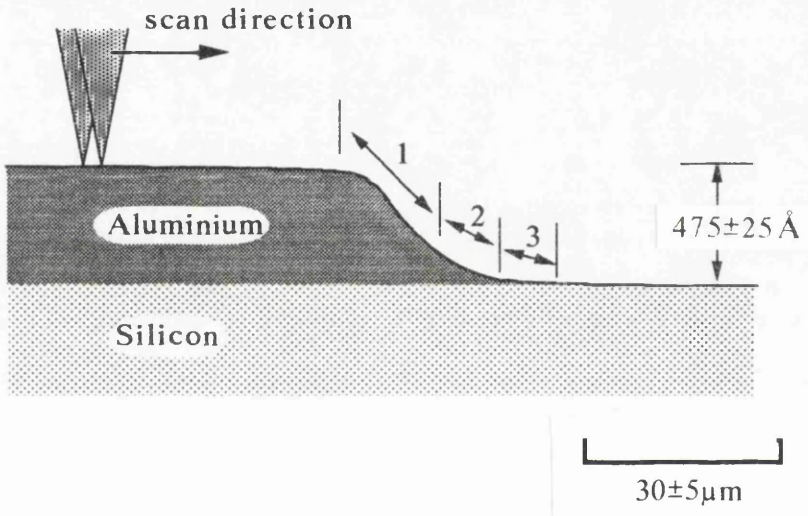


Figure 6-8 Schematic of the aluminium / silicon step sample (measurements were made using a stylus probe instrument)

	focussed beam diameter	beam separation	R_{ss}
Figure 6–9	4.7μm	6.3μm	1.3
Figure 6–10	4.7μm	22μm	4.7
Figure 6–11	4.7μm	38μm	8.1

Table 6–II Details of the focussed beam diameter and beam separation used to collect the results shown in figures 6–9 to 6–11. R_{ss} is the ratio of the values in the first two columns.

label	
a	side of first probe enters region 1
b	centre of first probe enters region 1
c	centre of second probe enters region 1
d	side of first probe enters region 2
e	side of second probe leaves region 2
f	side of second probe leaves region 3

Table 6–III Summary of labels as used on figures 6–9, 6–10 and 6–11.

In region 3 there is little change in reflectivity, but there will be some differential phase contrast as the thickness of the remaining layer of aluminium reduces to zero.

Line scan measurements are shown in figures 6–9 to 6–11. By changing the difference frequency ($\Delta\omega$) of the Bragg cell drive, the separation of the probes has been varied as shown in table 6–II. A value of R_{ss} above 8 was achieved using a difference frequency of 48MHz (the centre frequency of the Bragg cell was 80MHz). Arbitrary scales have been used for all the differential intensity and interference intensity responses. Increasing the difference frequency leads to an inevitable variation in the diffraction efficiency of the Bragg cell and hence total optical power incident on the object. This was not monitored and so comparison may only be made of the relative magnitude of differential and interference intensities between each set of results.

The graphs of figures 6–9 to 6–11 have each had a series of corresponding points marked on them, labelled from *a* - *f*. The significance of each is summarised in table 6–III.

Turning attention first to the results of figure 6–9, the dimensions of the probing beams are illustrated in figure 6–9 *e*. The first feature of significance is apparent from the differential intensity response where around –20μm, there are a pair of bipolar peaks. (The phase plot in figure 6–9 *b* indicates that the second of the bipolar peaks is inverted). These are generated as each of the probe beams passes into region 1, at the top of the slope.

Label *a* on figure 6–9 *a,c* indicates the position at which the edge of the first probe beam starts to cross into region 1. As the centre of this probe passes into region 1 (label *b*) and starts to move down the slope, the differential phase begins increasing (*b* on figure 6–9 *c*). There is a flattening in the differential phase response at *c* as the centre of the second probe passes into region 1.

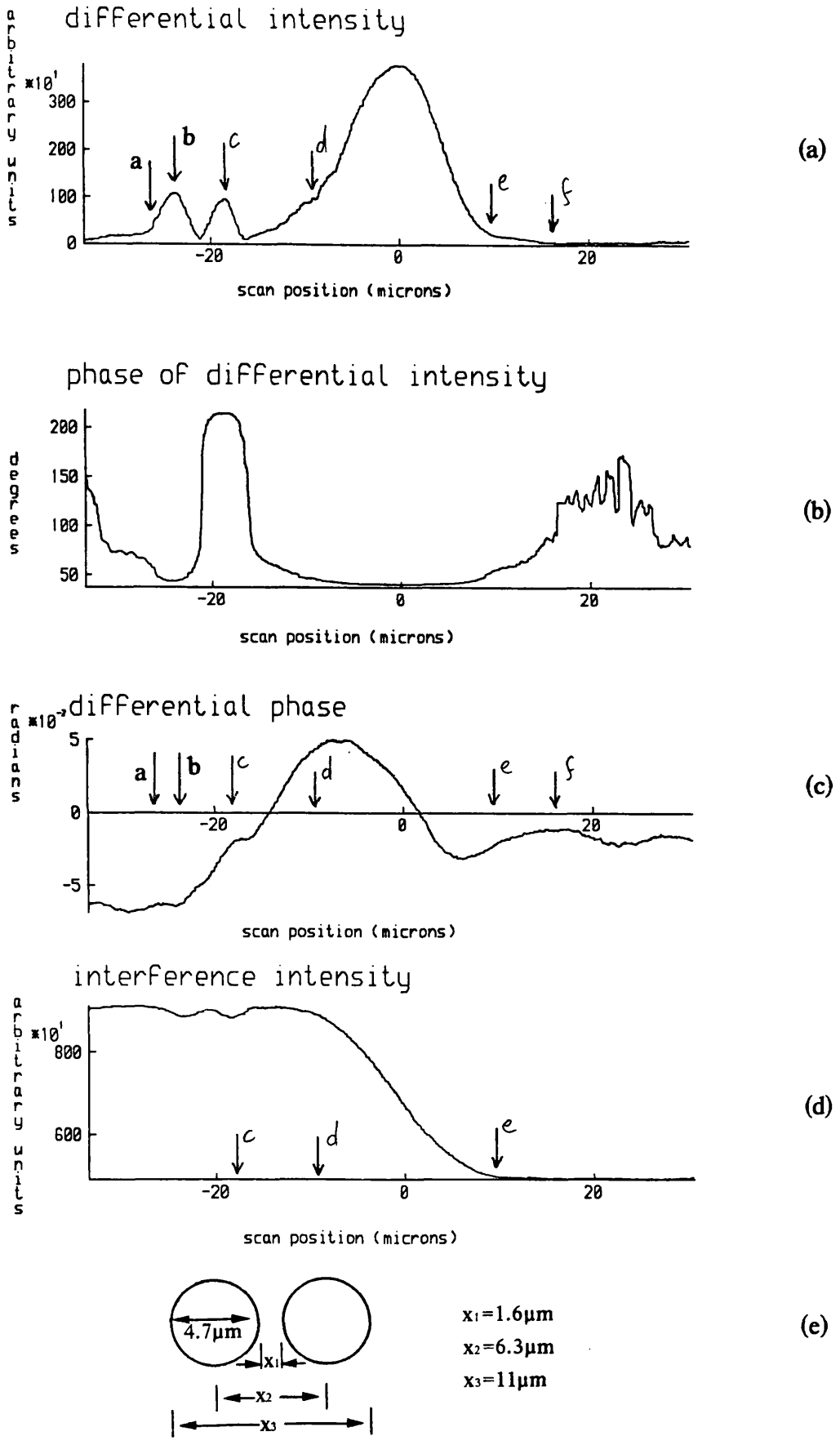


Figure 6-9 Differential optical phase and intensity line traces measured by two first order system across the aluminium/ silicon step sample - R_{ss} = 1.3 . (a) differential intensity, (b) phase of differential intensity, (c) differential phase, (d) interference intensity, and (e) dimensions of probing beams.

The transition into region 2 may best be seen from the interference intensity response (figure 6-9 *d*). As already mentioned, it is only over this region that the intensity of the light will be significantly affected. Therefore the edge of the first probe enters region 2 at label *d* and the edge of the second probe beam leaves this region at label *e*. The corresponding positions of labels *d* and *e* are marked on figures 6-9 *a* and *c*.

It can be seen that between points *d* and *e* on figure 6-9 *a* the response is very similar to that of the implanted silicon (figure 6-7 *a*). For both results, the same beam diameters and separations were used. It is apparent however that the width of the peak in figure 6-9*a* is over twice that of figure 6-7 *a*. The reason for this is that the intensity drops continuously over region 2 rather than abruptly as was case for the former results.

The differential phase contrast between *d* and *e* is due to both changes in topography and the electrical properties of the material.

A further label *f* has been used to indicate where the second probe beam exits the region 3.

We will now relate the distances between the labelled points *a* - *f* to the dimensions of regions 1, 2 and 3. The following symbols will be defined:

- w_1 — width of region 1 on sample
- w_2 — width of region 2 on sample
- w_3 — width of region 3 on sample
- w_{13} — extent of step ie., $w_{13} = w_1 + w_2 + w_3$
- x_1 — minimum distance between two probe beams
- x_2 — distance between mid-point of foci of probe beams
- x_3 — distance between opposite edges of the probe beams

x_1, x_2 and x_3 are indicated on figures 6-9 *e*, 6-10 *d* and 6-11 *d*.

Measured values from the line scans relate to physical distances on the sample according to the relations,

$$a\,f = w_{13} + x_3 \qquad \ldots(6-1)$$

$$d\,e = w_2 + x_3 \qquad \ldots(6-2)$$

$$e\,f = w_3 \qquad \ldots(6-3)$$

The following values are from figure 6–9: $a\,f = 42\mu m$, $d\,e = 19\mu m$ and $e\,f = 6\mu m$. Using these together with $x_3 = 11\mu m$, the dimensions of regions 1, 2 and 3 are,

$$w_{13} = 31\mu m, \quad w_2 = 8\mu m \quad \text{and} \quad w_3 = 6\mu m.$$

In arriving at these values several assumptions have been made. These include the model of the step in figure 6–8 which divides the step into three distinct regions. Also it is likely that the focussed beam diameter is greater than the diffraction limited value of $4.7\mu m$. Nevertheless we will see that these values for w_{13}, w_2 and w_3 are consistent with the remaining results.

In the second set of results (figure 6–10), the separation of the probing beams is increased by a factor of three so that x_3 is now equal to $27\mu m$ (see figure 6–10*d*). Labels *d* and *e* may be readily located on the interference intensity plot (figure 6–10*c*), and the distance *de* measured from this graph is $35\mu m$. It can be seen from equation (6 – 2) that the new values of $de = 35\mu m$ and $x_3 = 27\mu m$ are consistent with the previously calculated value of w_2.

The interference intensity is no longer a steady slope as in figure 6–9*d*. This is because the minimum separation of the two beams ($x_1 = 17\mu m$) is now much larger than the distance across region 2. So the first probe beam crosses this region and enters region 3, whilst the second probe beam remains in region 1. Only then does the second probe enter

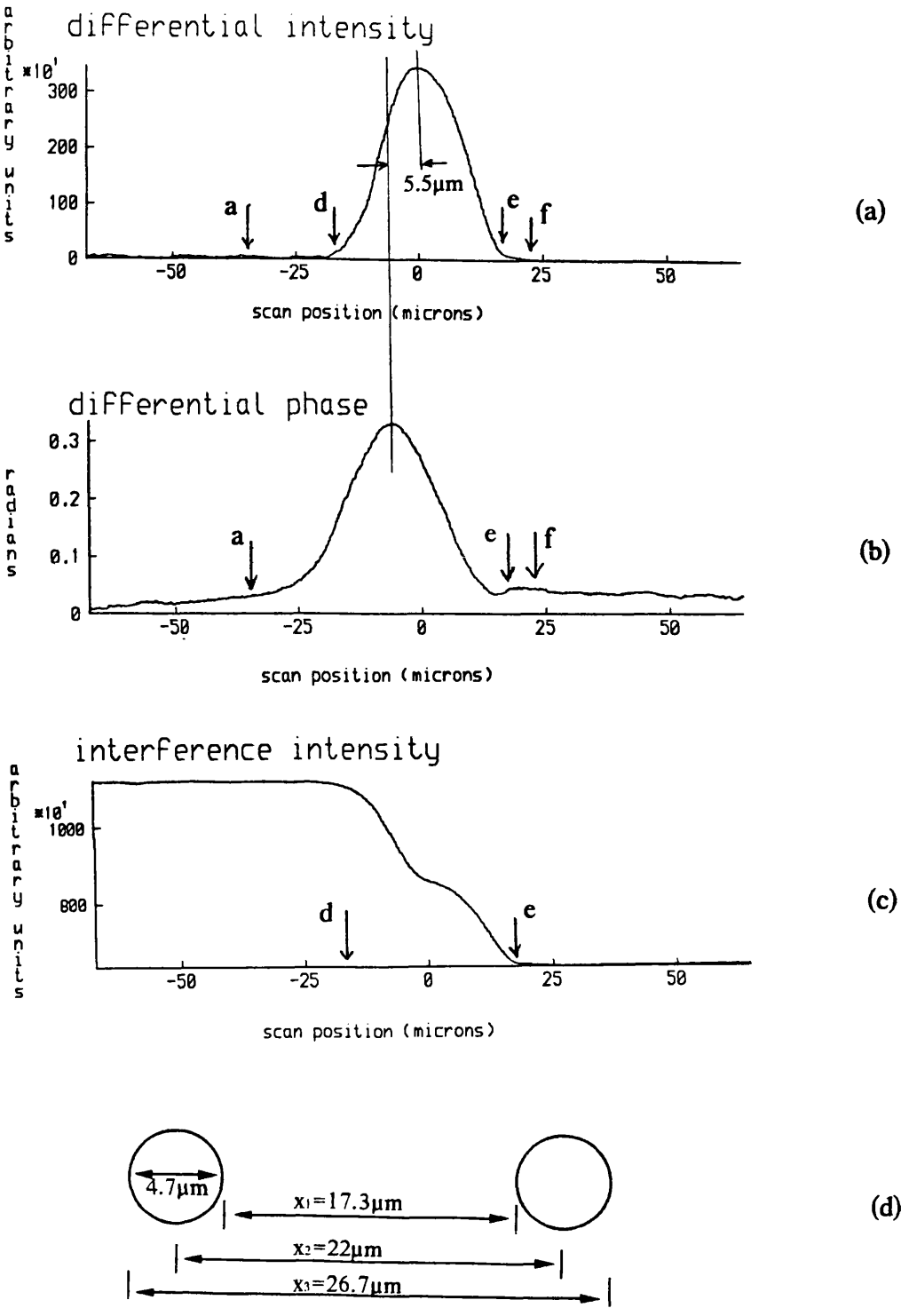
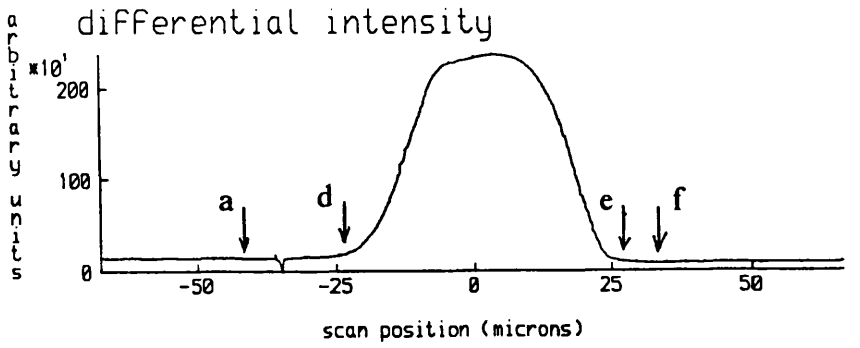
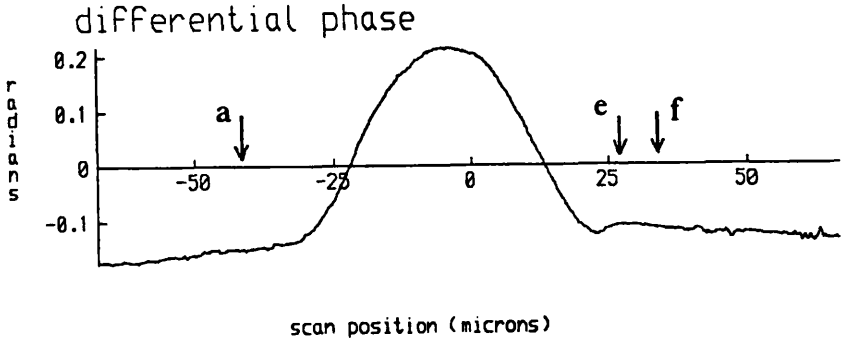


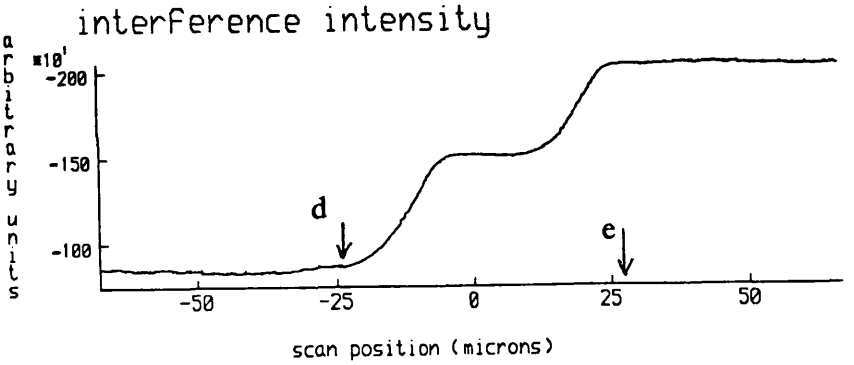
Figure 6-10 Differential optical phase and intensity line traces measured by two first order system across the aluminium/ silicon step sample - $R_{ss} = 4.7$. (a) differential intensity, (b) differential phase, (c) interference intensity, and (d) dimensions of probing beams.



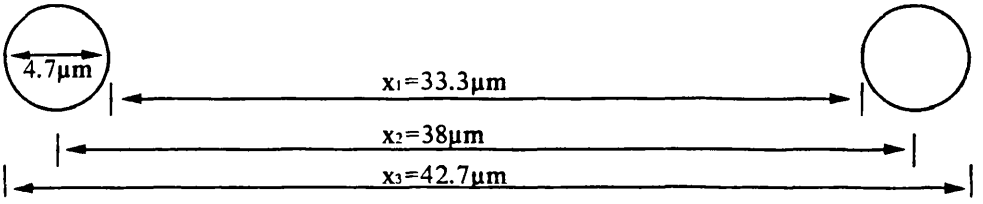
(a)



(b)



(c)



(d)

Figure 6–11 Differential optical phase and intensity line traces measured by two first order system across the aluminium/ silicon step sample - R_{ss} = 8.1 . (a) differential intensity, (b) differential phase, (c) interference intensity, and (d) dimensions of probing beams.

region 2. Since most intensity contrast occurs in this central region, there is a levelling in the rate of change between probe 2 leaving this area and probe 1 entering.

Using equations (6 - 1) and (6 - 3) with the value of $x_3 = 27\mu m$, labels *a*, and *f* have been located on the graphs of figures 6-10*a* and 6-10*b*.

The effect of a three fold increase in the probe separation distance causes considerable changes to both the differential phase and intensity responses. The width of the phase peak is now wider than the differential intensity because the latter is only significantly affected over region 2. Also it is apparent that the peak of the differential phase is displaced 5.5μm from the position at which the differential intensity response is maximum. This is an interesting consequence of the phase and intensity contrast being generated by different physical areas on the object.

In figure 6-11 the separation of the probe beams has been further increased so that x_3 is now equal to 43μm (see figure 6-11*d*). In a similar way to the results of figure 6-10, labels *d* and *e* may be located, followed by *a*, *b*, *c* and *f* using equation (6 - 1) and (6 - 3).

Looking first at the interference intensity (figure 6-11*c*), it can be seen that the horizontal section has now increased in length to about 20μm. The extent of this section would be expected to equal the minimum separation distance of the probe beams (x_1) minus the physical length of the sample which affects the interference intensity. Note that due to an artifact of software, the vertical axis of this graph has been inverted. Assuming that only region 2 will influence the interference intensity, this evaluates as ($x_1 - w_2$), ie., 25μm. The reason for this value being larger than the 20μm measured from the results can most likely be attributed to the diameter of the focussed beams being larger than the assumed values and also that interference may be affected over a distance greater than just region 2.

The differential intensity response is shown in figure 6-11 *a* . The effect of the further increase in the probe beam separation is that there is a flattening of the peak. In a similar way to the interference intensity response, it would be expected that the length of the top is equal to the scan distance for which the probe beams are either side of region 2, ie., 25μm. The sloped top is possibly caused by variations in reflectivity in regions 1 and/ or 3. A further comparison of figures 6-10 *a* and 6-11 *a* show that the sloped opposing sides remain unchanged in both sets of results. This is because in both cases, they are generated as each of the probe beams individually is scanned over region 2.

The differential phase result (figure 6-11 *b*) still has a curved top. The reason for this is that phase contrast occurs over a longer length on the sample (ie., regions 1, 2 and 3) than differential intensity contrast which occurs mostly in region 2.. Therefore it will be found that if the beam separation is further increased, this too will have a flattened top.

6.4 Summary of results

Results have been presented from three different experimental setups. The differential phase and intensity response from different types of structures have been demonstrated. In particular it has been seen how the simultaneous measurement of differential phase, intensity as well as interference intensity can enable complex structures to be understood. The versatility of the two first order arrangement has been demonstrated in its ability to change the ratio R_{ss} by a factor of 8 by purely electrical means (the optics remained unchanged throughout).

Areas where improvements are still required are in the modulations. It was also found that the differential phase signal to noise level in the last setup was at best about 60dB in a 300 Hz bandwidth. Work is required to improve this figure if the technique is to be used for imaging applications.

6.5 Description of experimental arrangement

This section will describe the experimental arrangement that was used to collect the most recent results shown in section 6.3 . The setting up of an interferometric optical system which includes so many separate elements is not at all easy without a well founded methodology. Description of the optical and electronic aspects of the system are divided into two parts. Starting with the optics, a full description is given of the system, including an extremely practical but to my mind, very useful guide to the alignment of the system. The electronics is described in the second part.

6.5.1 Description of experimental arrangement: the optics

Figure 6-12 shows details of the experimental setup. Only the central axis of each beam is drawn but it is straightforward to determine whether the light within any beam is converging, parallel, or diverging. The first unit in the system is the optical isolator which, for all the results presented in this chapter, comprised a quarter wave plate plus two polarizers. This type of isolation relies on there being no birefringent elements within the rest of the system. In practice there can be birefringence of the Bragg cell medium, the object under study and it can also arise as a result of in-built stresses in the lenses or beam splitters. The resulting isolation was extremely poor, between 10 and 20dB. A Faraday rotator type isolator is very much preferred, and recent trials have shown that an isolation of around 40dB can be achieved with this type of device.

Three lenses are needed so that the zeroth order beam generated on the first pass through the Bragg cell may be removed. The first two lenses, L_1 and L_2 furthermore serve as an extremely useful beam expander to ensure that the full aperture of the objective lens, L_3 is illuminated. Each lens is positioned as shown, so the distance between each is the sum of the respective focal lengths. Aperture 1 is used to block the zeroth order beam from the Bragg cell. If the focal length (f_1) of lens L_1 is small, the zeroth and first order beams

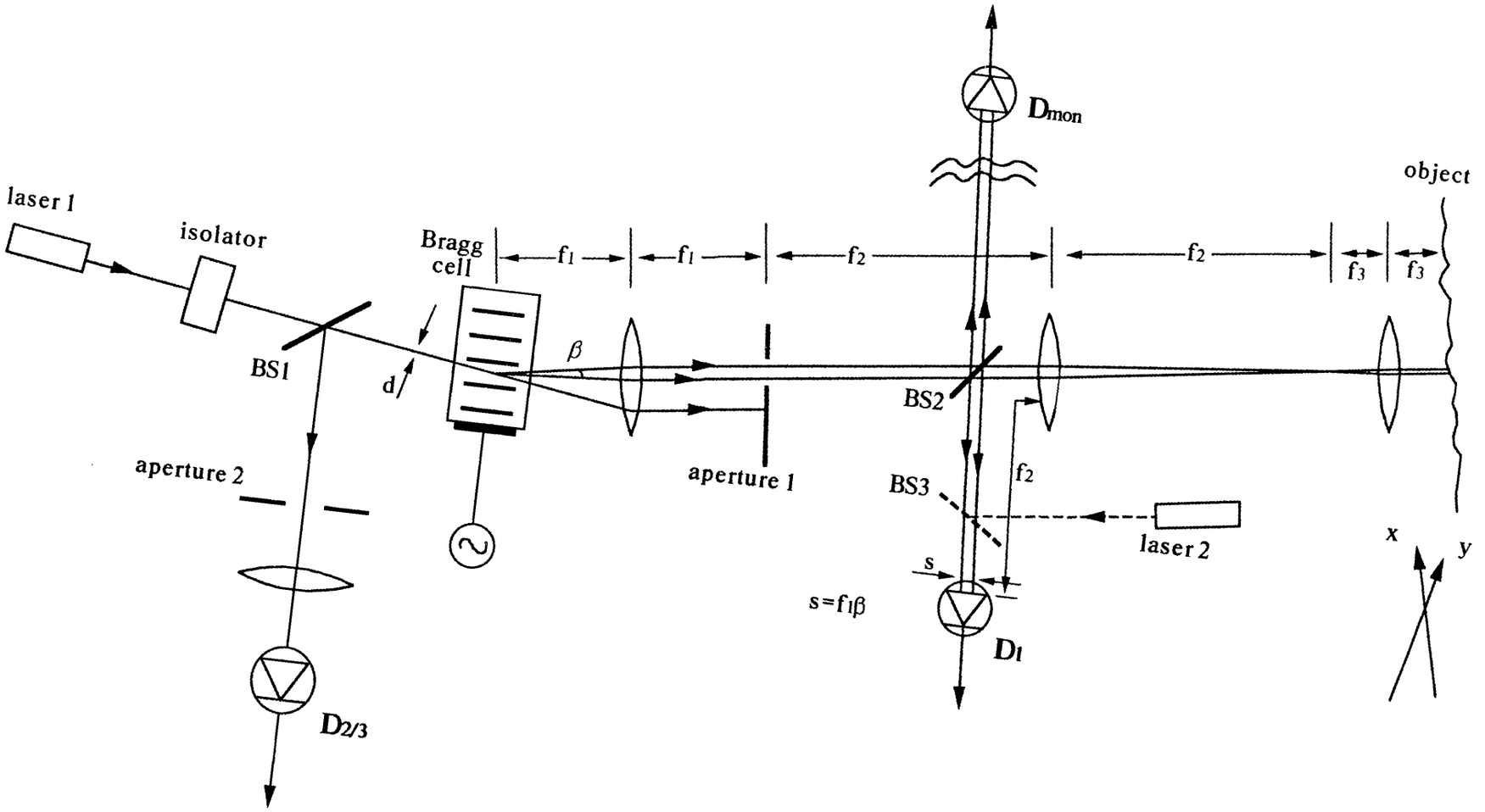


Figure 6-12 The two first order system : the layout of the optical components

become very close at aperture 1 and it is physically very difficult to block the zeroth order without affecting the two first order beams. The focal length f_1 must therefore be sufficient, but exactly how large will depend upon the precision with which the aperture can be translated. For the results of section 6.3 a focal length of f_1 = 5cm was used. The mean divergence of the zeroth and first orders at the Bragg cell was 13mrad, and so the consequential separation of the zeroth and first orders at Aperture 1 was 0.6mm.

Having chosen f_1, the focussed beam diameter p, and separation on the object surface q, are determined by the focal lengths f_2, f_3, the diameter of the beam entering the Bragg cell (d) and the angular divergence of the beams leaving the Bragg cell (β) such that,

$$p = \frac{1.22\lambda}{d} \cdot \frac{f_1 f_3}{f_2} \qquad \ldots(6-4)$$

and

$$q = \frac{f_1 f_3}{f_2} \beta \quad . \qquad \ldots(6-5)$$

The ratio of beam separation to focussed spot size is,

$$R_{ss} = \frac{\beta d}{1.22\lambda} \, . \qquad \ldots(6-6)$$

After a second passage through the Bragg cell, the light passes through Aperture 2 and is detected at detector D2/ D3. The purpose of the aperture is to remove the zeroth order beams (ζ_2 and ζ_3 - see figure 3-2), and also μ_2 and μ_3 if the detector is to function as detector D2. When $R_{ss} < 1$, ie., imaging mode, it is not physically possible to remove μ_2 and μ_3 , and even if $R_{ss} > 1$, it can still be difficult to remove these side beams. Since it is not even essential for differential phase measurement to remove them,

it was deemed (in the experiments already performed) more desirable to collect all three beams μ_1, μ_2 and μ_3. A lens is used in front of the detector, and in order to avoid problems arising from interference between intermodulation orders within μ_1, μ_2 and μ_3 (as described in section 5.3.4), the detector and lens should be separated by one focal length.

Detector D1 is most conveniently located as shown in figure 6-12, between aperture 1 and lens L2. It should be positioned so that it is more or less one focal length (f_2) from L_2. This will mean that any tilt of the specimen will only change the angle of incidence of the light on this detector. The separation of the two focussed beams on the detector surface is $f_1 \beta$ (this defines the minimum detector size).

The effect of object tilt will be to move the light laterally across detector D2/ D3. Even if this results in a small loss of light, it is not serious as the interference phase measurement will be unaffected. It should be noted that only object defocus (and not tilt) will have any effect on the light returning through aperture 1

We will now consider the alignment of the system. It was found that the best method was to treat the lenses as afocal pairs, and the following method was used:

i) Mark out on the bench the approximate positions for the Bragg cell and lenses. Use a mirror in place of the sample. This should be mounted so that it can be tilted in two directions and also translated along the optical axis for focussing. It further helps in the alignment procedure if it can be moved to one side, completely out of the way of the beam.

ii) Drive the Bragg cell with a single RF frequency which is the mean frequency of those to be used for the two first order probe beams. Rotate it so that there is maximum intensity in the first order and set the mirror normal to this beam. The tilt of the mirror should be left fixed throughout the rest of the alignment procedure.

iii) Position lens L_1 as accurately as possible (with a ruler). Similarly locate lens L_2, and place BS2 in the way of the beam so that the front and back surfaces are nearly perpendicular to the incident light. L_2 may now be accurately positioned by moving it back and forth so that the light leaving this lens is dead parallel. The orientation, vertical and horizontal positions of all the lenses should be set so that they are exactly symmetrically placed about the optical axis. This is managed by looking at the reflections from the lens surfaces. It is assumed that none of the lenses have planar surfaces.

iv) Use a second laser as shown in figure 6–12 arranged so that when this light passes through L_2, it is as near as possible coincident with the optical axis of the lenses already aligned. This can be achieved by comparing the light from each laser beyond lens L_2. It is useful if the beam splitter BS3 is mounted permanently on a translation stage. This second laser is required every time the objective lens is changed.

v) Turn off the first laser. Lens L_3 can now be accurately located by ensuring that the light returning back to laser 2 is well collimated. Again by looking at the light reflected from the front and back surfaces, this lens should be positioned exactly symmetrically about the optical axis.

vi) Turn off the second laser, and move the beam splitter BS3 to one side. It would help in any future alignment if this beam splitter may re-positioned in exactly the same place.

vii) It was found that the most difficult element to know how to position was the Bragg cell. Indeed it did not seem at all critical exactly how it was positioned : there was no noticeable change in performance if it was moved perhaps over a distance of 2cm along the optical axis. If the distance between the Bragg cell and lens L_1 is not exactly one focal length then it can be seen that the light returning back to the Bragg cell will be laterally displaced, but *the direction of beam propagation will remain unchanged*. This is most important and means that small misplacements of the Bragg cell are easily tolerated. It is also interesting to note that this kind of

misalignment will result only in a change in the *orientation* of the light at aperture 1, and will in no way affect the way in which the aperture blocks the light.
The Bragg cell is best positioned by maximizing the interference signal at detector D2/D3.

viii) It is sometimes found that light reflected from the surfaces of elements in the system is collected by detector D2/D3 where interference is generated at the same frequency as that of the differential phase response. This is overcome by slightly misaligning (eg rotating, tilting, translating at 90° to the optical axis etc.) the offending element.

In chapter 5 the use of an extra monitoring detector was discussed which measures the total power of the incident light. This detector may most easily monitor the outgoing light from beam splitter BS1. However unlike the light entering detector D1, the light within each beam on the other side of BS1 will be diverging. Further lenses will therefore be needed in order to collect all of this light onto detector D_{mon}. Also it should be remembered when choosing a detector that the accuracy of the differential intensity response is dependent on an even sensitivity across the detector surfaces.

6.5.2 Description of experimental arrangement: the electronics

Figure 6-13 shows the electronics used to drive the Bragg cell and the detection electronics. Two phase locked synthesizers provide RF carriers at frequencies ω_1 and ω_2. Each of these signal are divided into two by summers S_1 and S_2 (these are 3dB couplers), and half of each is directed into the R ports of the HP 10514A mixers. It was found that these mixers had particularly low distortion. A low frequency unit (LFU) generates two identical low frequency modulation signals in phase quadrature and these are used to amplitude modulate the carriers. Both modulated carriers are added at the summer S_3 and the signal is then amplified by an RF power amplifier before being used

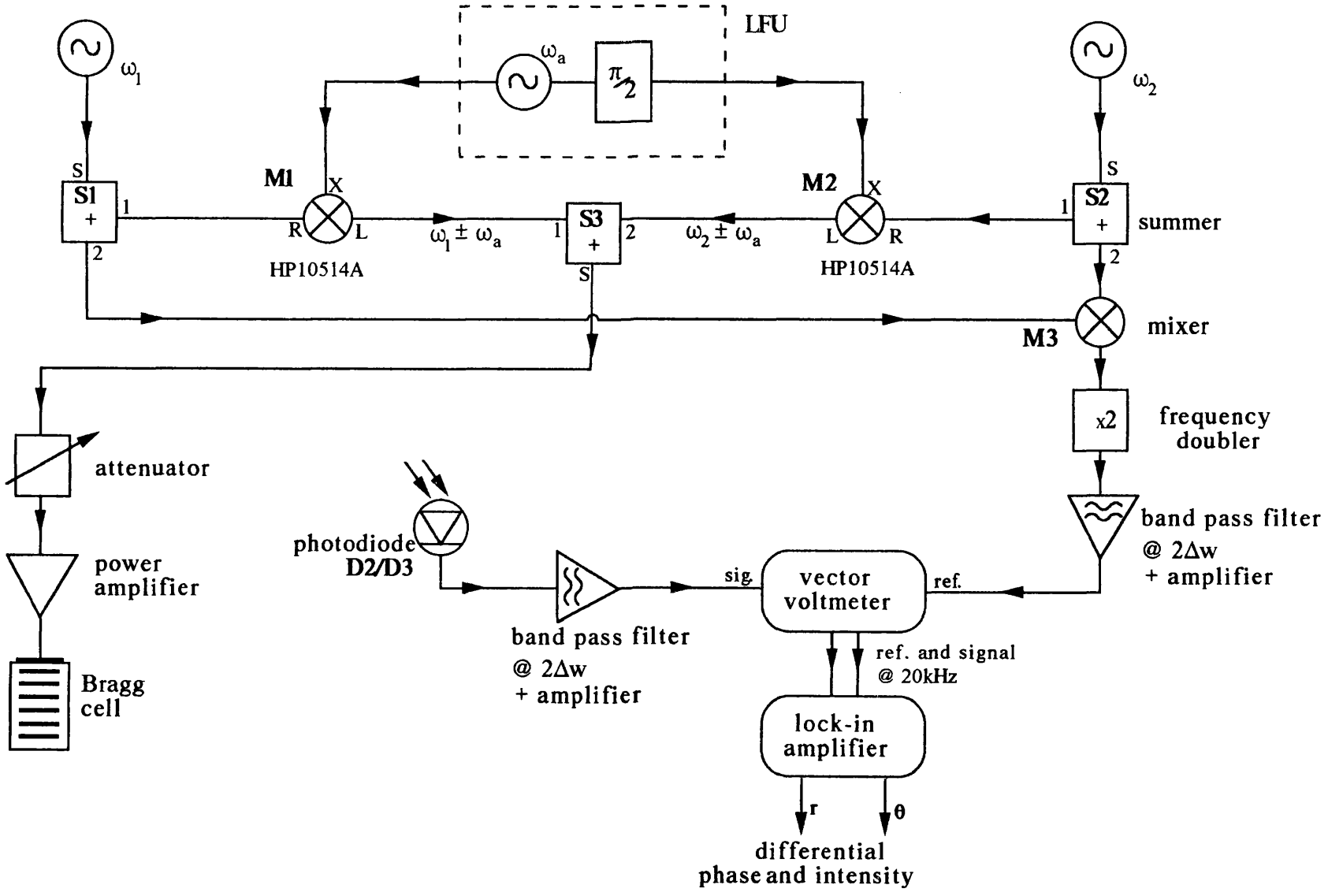


Figure 6-13 The two first order system : the electronics

to drive the Bragg cell. The variable attenuator is used to adjust the overall level of both modulated carriers. The level of each modulated carrier may be affected individually by altering the outputs of the synthesizers or the LFU.

The output from the second port of the summers, S_1 and S_2 are mixed by M_3, frequency doubled, and the $2\Delta\omega$ frequency component is filtered and amplified. This is used as the reference for the phase sensitive detection. A detector D2/ D3 measures the optical interference, and the output of this is filtered and amplified. Both reference and signal are beaten down to 20kHz by a vector voltmeter, and phase sensitive detection is performed by a lock-in amplifier. The r and θ outputs are used to measure the interference intensity and differential phase responses respectively. It is not strictly necessary to filter the output from the photodiode, but it was found that this reduces the noise level.

Altogether three different version of the low frequency modulation unit have been designed and built. The last unit has complete and independent control over the amplitude and the phase relation of each output signal. Unfortunately lack of time has not enabled results to be taken with this newly completed unit. Furthermore, in order to perform feedback with optimum accuracy a different design is required, and for this reason the unit will not be described here.

The differential intensity measurement is made from detector D1 and performed at frequency $2\omega_a$. This is most easily done using a lock-in amplifier, taking a reference from the low frequency modulation unit.

One last measurement that may be made above dc is absolute intensity at frequency $4\omega_a$ from detector D2/ D3.

CHAPTER 7

MIRROR SCANNING IN A DIFFERENTIAL SYSTEM

7. Mirror scanning in a differential system

A Scanning optical system may take its measurements in one of two ways, by moving the light over the object or translating the object with the light remaining fixed. The prime attraction of the latter method concerns the effects of lens aberrations and small misalignment of optical elements. Although these inevitably degrade system performance, every point in the field of view is affected in an identical manner. This is different from the situation for a light scanning method where the response is most likely to change in some way at the edge of the optical field.

Although well suited to metrology, object scanning is rather cumbersome for imaging applications, and is most likely to be the one factor limiting the image acquisition time. Scanning of the light may on the other hand be managed by a small light weight mirror enabling much faster data collection.

Previously developed two beam differential techniques have until now used only object scanning. In this chapter we will look at the application of a scanning mirror to a differential system. The first section discusses the layout of the optical system, and outlines problems which will result from lens aberration effects. Analysis presented in the second section shows that in the absence of aberrations, the scanning mirror will not cause any errors in the differential measurements. Section three considers how to model the effects of lens aberrations and misalignments of optical elements, and the fourth section presents a ray tracing method. This implicitly ignores diffraction effects, and the validity of the method is discussed in the fifth section. Much of the material discussed is applicable to non-differential as well as differential optical profilometers.

7.1 Application of a scanning mirror to a differential optical system

The standard scanning optical microscope (SOM) measures the intensity of light reflected from the object, and the use of a mirror in such a system is relatively straightforward. In a non-differential interferometric system the mirror is used to scan the probe beam, with the interferometer reference arm remaining stationary. The choice of position for the mirror must ensure that there is no false phase contrast as the mirror is scanned. A suitable position is therefore at the back focal plane of the objective lens, where a change in the orientation of the mirror will cause lateral movement of light across the sample. The probe beam will scan across the object a distance which is proportional to the mirror rotation angle.

A mirror may be similarly applied to the differential technique described in this thesis. Since differentiation of the surface features is in the direction adjoining the focussed probing beam centres, it is better to raster the beams back and forth in the same direction. This is to ensure proper registering of differentiated points in consecutivelines of the image. The slow scan which is perpendicular to the raster direction, is at a fraction of the raster speed, and may be performed quite adequately (and with less complication) by a mechanical stage.

With ideal components, there will be no phase variation between the probing beams as the mirror is rotated (this is proved in the next section). As far as the effects of aberrations are concerned, there are two important considerations. First there should be negligible variation of phase *between* the probe beams, and second, the light returning through the Bragg cell should be well collimated and at the same angle as it emerged on the first passage. For the differential intensity measurement made at detector D1, providing that all the reflected light is collected, the effects of lens aberrations are unimportant. Since this is the favoured detector for differential intensity metrology, only the differential phase response of detectors D2 and D3 need be considered.

The effects of lens aberrations in a differential system is in certain ways less severe than in a non-differential system. Since both beams follow nearly identical paths, the wavefronts of each will become aberrated in very similar ways. So if the two beams could be well interfered, then a substantial amount of the aberration effects will cancel. The problem however lies in achieving this interference. If the light returning to the Bragg cell is not well collimated then there will be incomplete recombination of the two beams because light which is off the Bragg angle has a lower diffraction efficiency. Providing that this reduction is identical for both beams, then aberrations again may be unimportant. However to be sure that so many things are identically 'not quite right' for both beams is far from a reliable method to achieve consistent and accurate measurements. Furthermore, although there will be partial cancellation of aberration effects, the overwhelming increase in sensitivity of the differential technique (by virtue of it being common path) means that aberrations must still be taken into account.

It is therefore important to quantify the divergence of each reflected beam entering the Bragg cell, and means of assessing aberration effects will be discussed further in section 7.3 .

7.2 Mirror scanning in an aberration free system

This section discusses the mirror scanning arrangement shown in figure 7-1 and demonstrates that, in the absence of lens aberrations, it will allow error free differential phase measurement.

Light from the laser enters the beam splitting device (ie., the Bragg cell) where it is split into two diverging beams. Lenses 1 and 2 are separated from planes P_1, P_2 and P_2, P_3 by one focal length of the respective lenses (f_1, f_2) so that the point of divergence in plane P_1

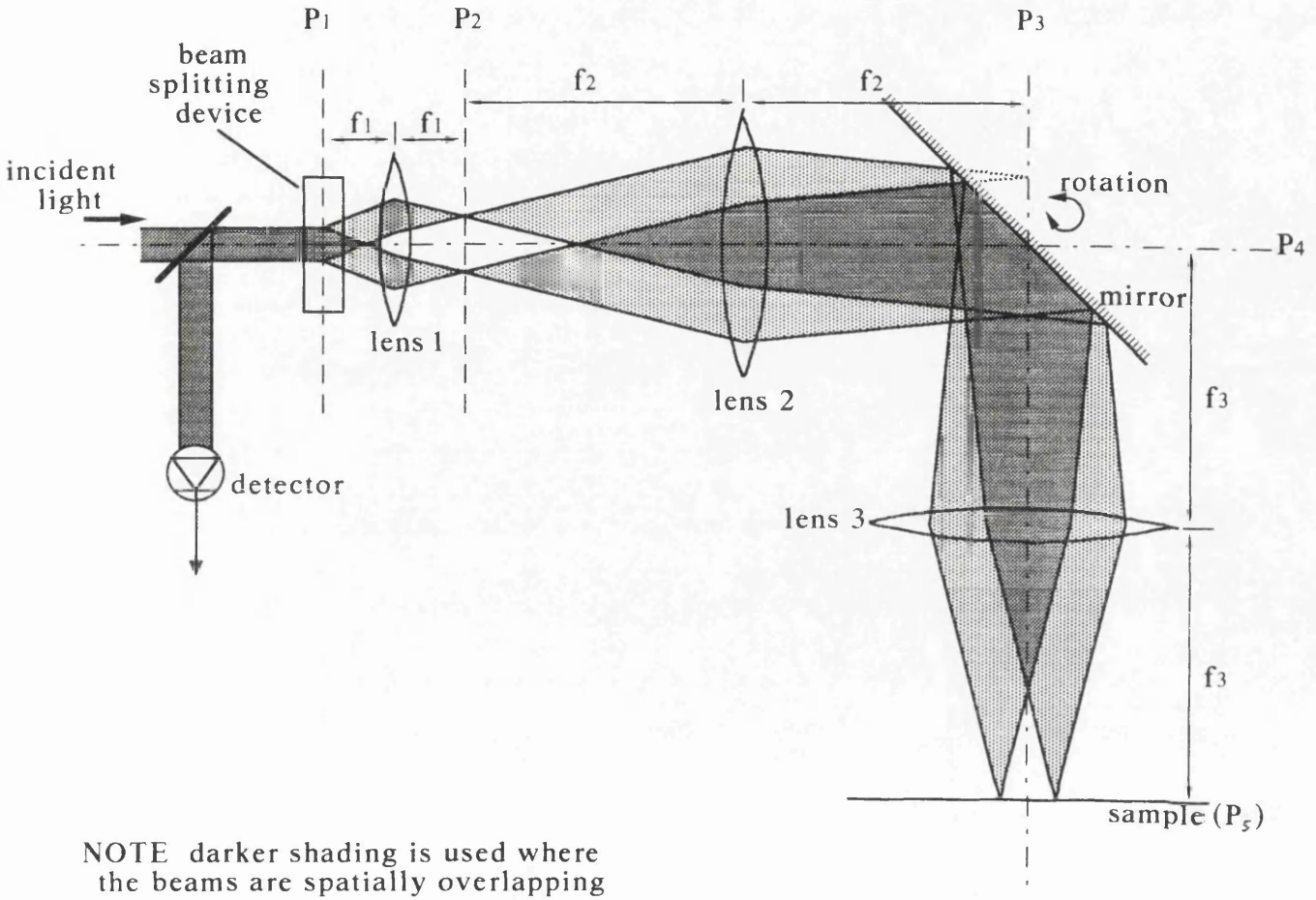


Figure 7-1 Schematic of mirror scanning arrangement

is imaged in plane P_3. Lens 3 is positioned with its optic axis coincident with P_3 and at a distance of one focal length (f_3) from plane P_4 and the sample plane P_5. The mirror is positioned so that its axis of rotation passes through the intersection point of P_3 and P_4.

Thus arranged,the two focussed beams at plane P_2 are imaged telecentrically at the sample (ie., with the central axis of each beam normal to P_5), and as the mirror is rotated, both focussed beams scan, with constant separation, across the object. The two beams are recombined by the beam splitting device and the measurement of the resulting interference at the detector is used to measure differences in optical path length due to sample features.

In this scanning arrangement, if diffraction effects and lens aberrations are ignored, the two beams will re-enter the beam splitting device with exactly the same collimation, orientation and lateral position as the outgoing beams. The wavefront of each beam is perpendicular to the propagation direction of each beam. In order for the differential phase response to be free from error, the phase of the wavefronts for each returning beam must be invariant upon rotation of the mirror. This can be proved using a simple ray model. First the optical path length variations at the lenses are considered, tracing a single ray through each lens, from plane P_1 to P_2, P_2 to P_3, P_4 to P_5, and then back to P_1 . The optical path length differences imposed by rotation of the mirror is then examined.

It is clear from figure 7-1 that light travelling outward between planes P_1 and P_3 will be unaffected by the rotating mirror so we will start by looking at the light travelling between planes P_4 , P_5 and back to P_4 .

7.2.1 Phase variation of a single ray traversing between planes P_4, P_5 and back to P_4

Figure 7–2 shows an enlarged view of a single rays's trajectory (FGHJK) between planes P_4 and P_5. The ray starts at F on plane P_4, with an orientation of θ to the normal and a displacement d from plane P_3. Upon reflection, the ray arrives back at K on plane P_4, with an orientation of θ' and displacement d'. The perpendicular distances between planes P_4, the lens and P_5 are both equal to the focal length of the lens. In this section we will show that for an aberration free lens the optical path length is constant for *any* ray which is reflected between planes P_4 and P_5.

The displacement and orientation of the outgoing and reflected rays are related such that,

$$d' = d \qquad \ldots(7-1a)$$

and

$$\theta' = \theta \qquad \ldots(7-1b)$$

On rotation of the mirror, both d *and* θ change, and the position of H is scanned over plane P_5.

The variation in optical path length along FGHJK versus θ is most easily considered by comparing the path to a reference along $FR_1R_2R_3K$ which intersects plane P_4 normally at F and K.

If we denote the optical path length difference of the two rays FR_1R_2 and FGH as '*a*', and the difference in length of the two rays R_2R_3K and HJK as '*b*', the ray FGHJK differs in length from the reference ray by a distance p_l, such that,

$$p_l = b - a \qquad \ldots(7-2)$$

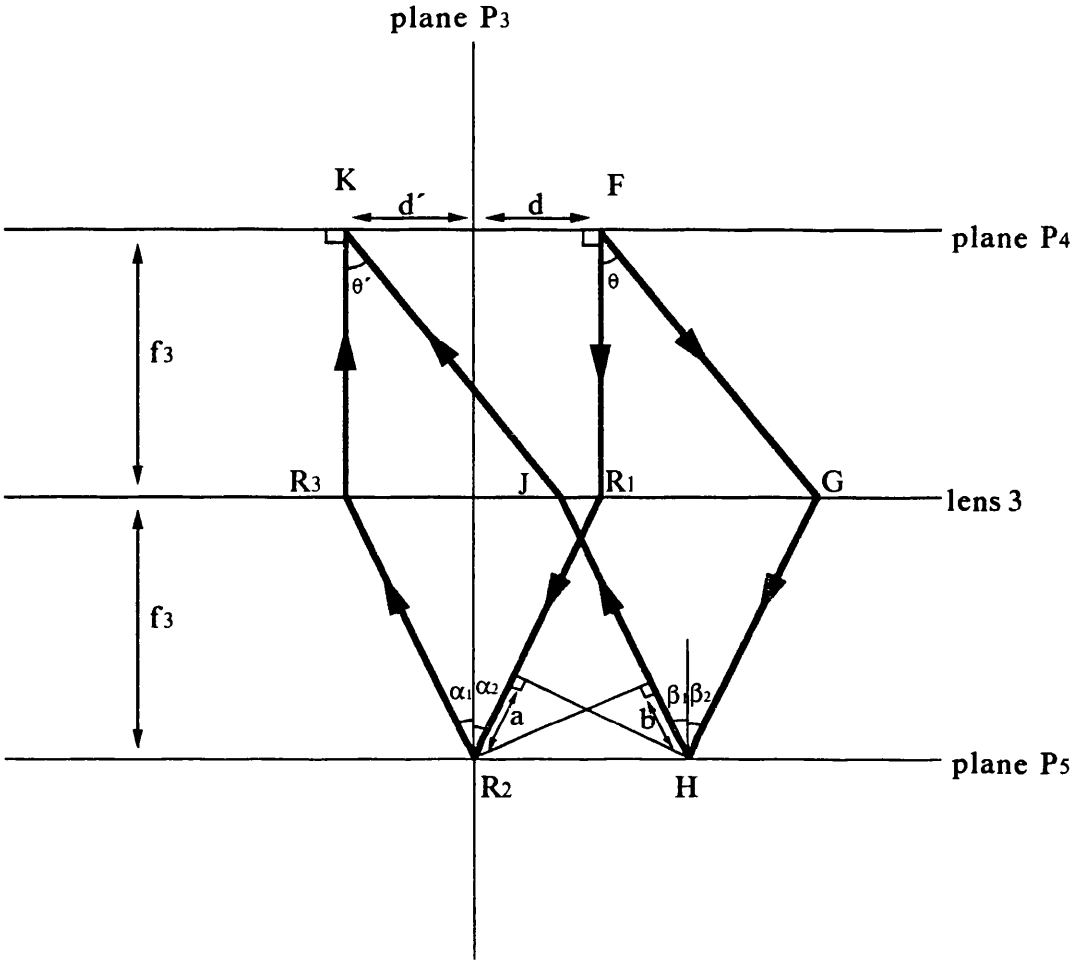


Figure 7-2 Tracing a ray between planes P4 and P5

Since rays R_1R_2 and GH are parallel, α_1 equals β_1 (see figure 7–2). The angle of reflection is equal to that of incidence so,

$$\alpha_2 = \alpha_1, \text{ and } \beta_2 = \beta_1 \qquad ...(7-3)$$

Hence triangles X_1R_2H and X_2HR_2 are identical, and

$$a = b \qquad ...(7-4)$$

Substituting this into equation (7 – 2), it follows that the optical path length is the same for any ray reflected between planes P_4 and P_5 which are launched with *the same* displacement from the optical axis (ie., d), but arbitrary angle of incidence (ie., θ).

Figure 7–3 shows two reference rays $FR_1R_2R_3K$ and $R_4R_5R_2R_6R_7$ reflected between planes P_4 and P_5. The first reference ray originating at F is the same as that shown on figure 7–2, and the second reference ray originating at R_4 has an arbitrary displacement from the optical axis. Since both rays are brought to the same focus, and both originate from the back focal plane of the lens, the optical path lengths FR_1R_2 and $R_4R_5R_2$ will be equal. Similarly the optical path lengths of the reflected rays, R_2R_3K and $R_2R_6R_7$ will also be equal.

Hence it follows that all reference rays have equal optical path lengths.

The above conclusions from figure 7–2 may now be revised as follows : the optical path length is the same for any ray reflected between planes P_4 and P_5 which are launched with arbitrary angle of incidence (θ) *and displacement* (d) from the optical axis.

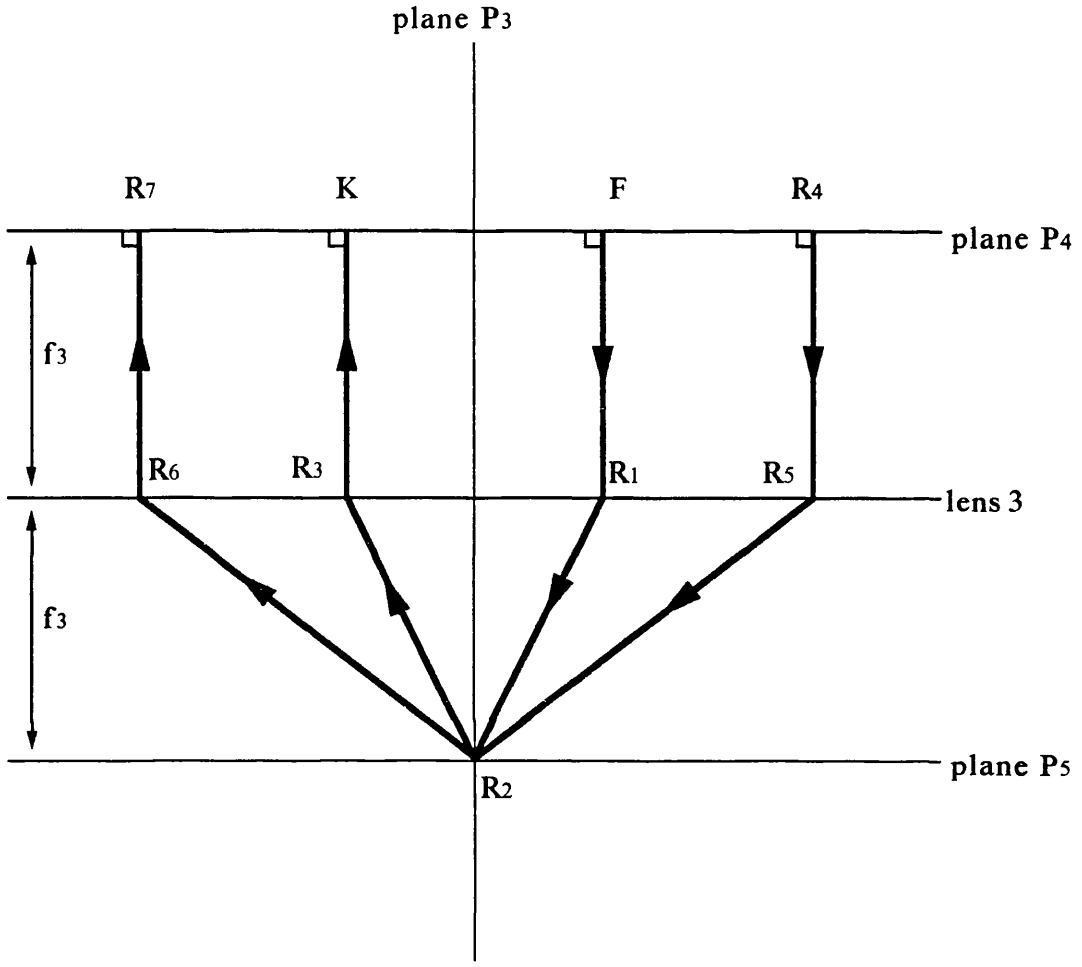


Figure 7-3 Tracing two reference rays between planes P4 and P5

7.2.2 Optical path length variation at rotating mirror

Figure 7–4 shows the interaction of an outgoing and returning ray at the mirror. The outgoing ray hits the mirror at S and is reflected, intersecting the plane P_4 at K. The ray reflected from the object would, in the absence of the mirror, intercept plane P_4 at F such that,

$$KT = TF \qquad ...(7-5a)$$

Upon reflection from the mirror surface, it will cross the plane P_3 at E where,

$$LT = TE \qquad ...(7-5b)$$

Since the incident and reflected rays from the object are parallel, it follows that the triangles SLT, RET, and SKT, RFT are identical, and hence,

$$SL = RE \quad \text{and} \quad FR = SK \qquad ...(7-6)$$

The optical path length imposed by the mirror for the outward and return rays is,

$$p_m = -SL + SK - FR + RE \qquad ...(7-7)$$

Therefore using (7 - 6), the total optical path length due to the mirror is zero for any rotation angle.

7.2.3 Optical path length variation around entire system

It has been shown that, on rotation of the mirror, there is no variation in optical path length of light reflected between planes P_4 and P_5 , and also at the mirror. The mirror will furthermore have no effect on the position and orientation of light returning from plane P_3 to P_1 . It can therefore be concluded that in an aberration free system, the optical arrangement shown in figure 7–1 will allow the measurement of differential phase at detector D2/ D3, without error.

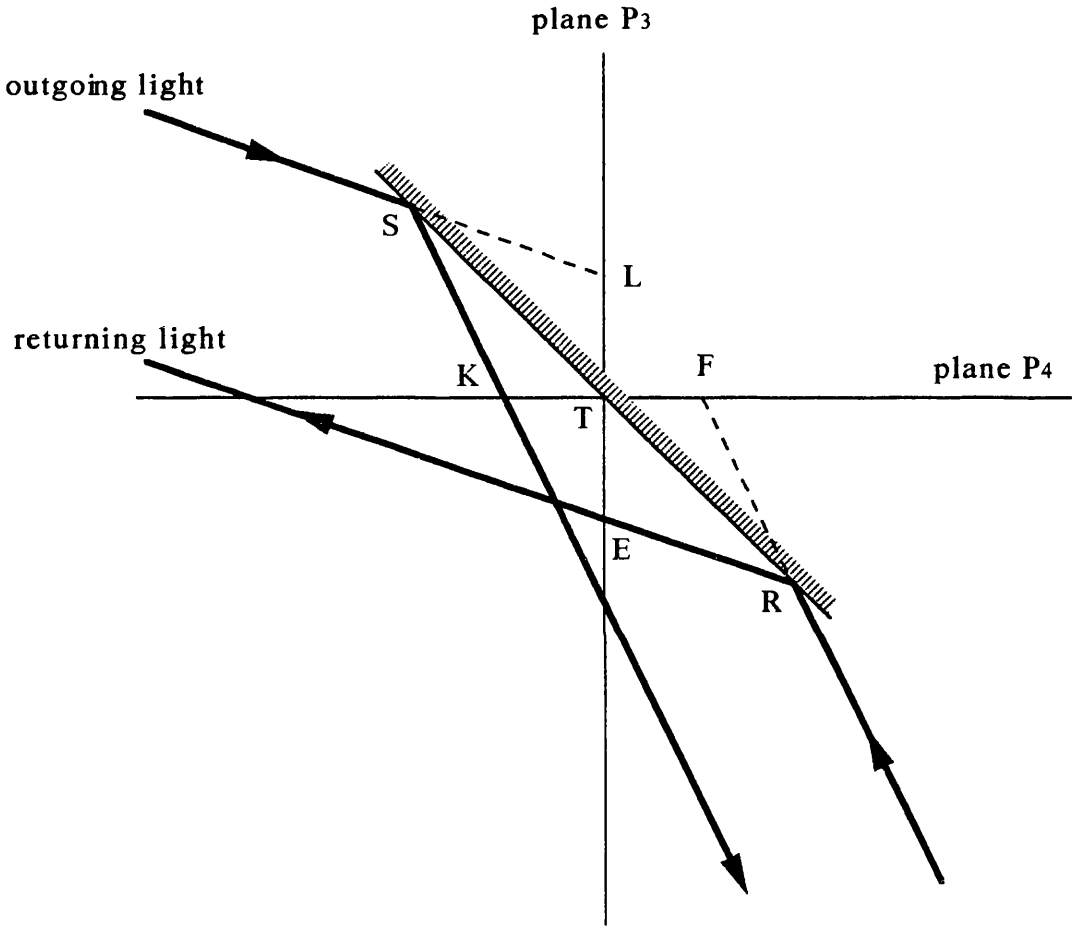


Figure 7-4 Ray tracing at the mirror

7.3 Modelling the effects of lens aberrations

Taking full account of lens aberrations is considerably more complicated than the previously discussed aberration free system. The reason why there was no overall phase error is because any variations in orientation or path length of the light on the outward passage is equal and opposite to the variations effected as it returns through the system. Aberrations remove this symmetry.

The element which has the largest amount of aberration is the objective lens (lens 3). However it is insufficient to view this one element in isolation because, even though aberrations are relatively unimportant for the other lenses, the light will not re-enter the other elements with exactly the same orientations and positions as on the first passage.

It is therefore necessary to propagate the light through every element in turn. The method favoured is a ray approach in which individual rays are traced around the system. Starting at plane P_1 (see figure 7-5) a ray is first traced via lens 1 to plane P_2. From here it is propagated through lens 2 to the point where it would intersect plane P_3 (the mirror is assumed transparent). The effect of the mirror is to transform the ray between planes P_3 and P_4. The ray is then propagated via lens 3 to the object (P_5) from where it is reflected back to plane P_1, past each element in turn.

The effect of aberrations on the collimation of the light returning into the Bragg cell is calculated by tracing several parallel rays around the system. The mirror must then be rotated by a small amount to see how the collimation varies with scan position.

In order to ascertain the effects of lens aberrations on the differential phase response, the difference in the total optical path length $\phi(y_d, \alpha; \beta)$ of two rays 1 and 2 (see figure 7-5) must be calculated. Here, y_d is the initial displacement of rays 1 and 2 at plane P_1 along the y axis, α is the mirror rotation angle, and β is the initial divergence between the two beams at plane P_1.

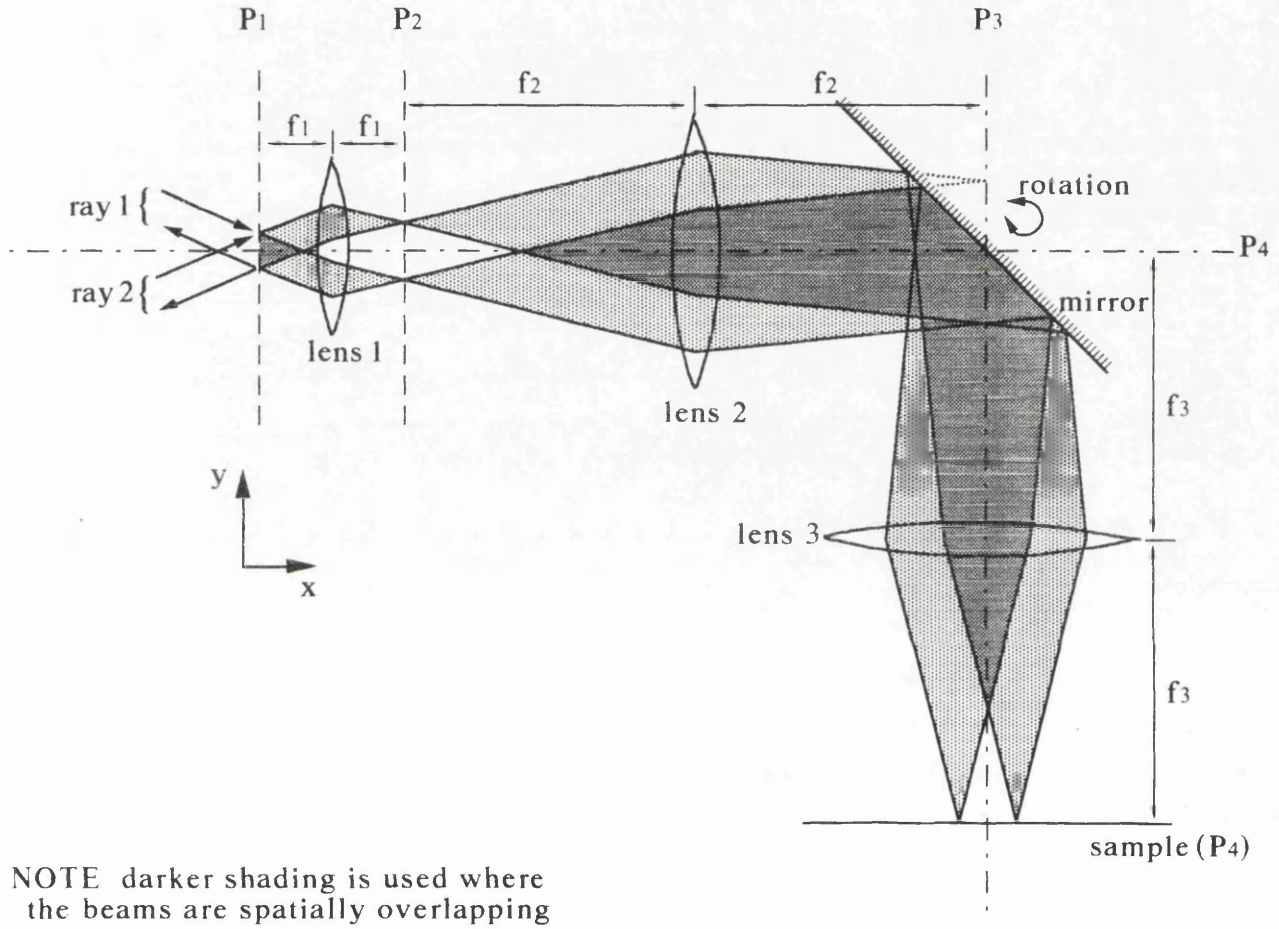


Figure 7-5 Ray tracing in a mirror scanning system

The differential phase response $\phi_m(\alpha)$ is a weighted sum of the phase difference of each pair of rays so that,

$$\phi_m(\alpha) = \frac{\int_{-\frac{d}{2}}^{\frac{d}{2}} \gamma(y_d)\,\phi(y_d, \alpha; \beta)\ \mathrm{d}y_d}{\int_{-\frac{d}{2}}^{\frac{d}{2}} \gamma(y_d)\ \mathrm{d}y_d} \qquad \ldots(7-8)$$

where d is the diameter of the beams emerging from the Bragg cell, and the initial intensities of all rays at plane P_1 are assumed equal.

The function $\gamma(y_d)$ describes the diffraction efficiency of the second Bragg interaction for the rays reflected from the object. It is dependent on the collimation of the light returning back to the Bragg cell, and ranges in value between unity for perfect collimation (ie., no lens aberrations) to zero for no collimation. Determination of this function is rather complicated. In addition to ray tracing, it requires computation of the acousto-optic interaction for an optical beam with an arbitrary phase front. This information may be available from the manufacturer of the Bragg cell or else it is possible to do the analysis as outlined in chapter 4.

If the collimation of the returning light remains within certain limits determined by the Bragg cell design, it may be possible to assume that $\gamma(y_d) = 1$ for $-d \leq y_d \leq d$. In other words, the returning light is always at Bragg incidence for any mirror rotation angle.

However, even if the function $\gamma(y_d)$ is not easily evaluated, the ray tracing approach will give at the very least a clearer insight into any problems which may arise in this mirror scanning arrangement. It would be useful for example to calculate how the phase difference of pairs of rays across the beams varies with the position (y_d) of the rays in the incident beam. The effect of mirror rotation can also be investigated by varying α.

7.4 Application of ray tracing to mirror scanning system

Appendices D and E analyze a ray propagating through a lens and reflected from the mirror surface. These will be briefly summarized.

Ray tracing through a lens

The lens is modelled as a thick lens. Referring to figure 7-6, a ray is propagated between two planes (P1, P2) which are a distance s_o and s_i from the principal points of the lens. The lens is defined by the radii of curvature (R_1 , R_2) of the two surfaces, thickness (t) and refractive index (n). The incident ray is at an orientation σ_1 and displacement d from the optical axis.

Expressions are derived enabling calculation of:

i) the orientation (σ_2) of the ray leaving plane P2 ;

ii) the displacement (d_2) from the optical axis of the ray leaving plane P2 ;

iii) the total optical path length between the two planes.

Ray tracing at mirror

The mirror is modelled as shown in figure 7-7. The point of rotation is a distance t behind the front surface (t is usually above zero for off the shelf mirror assemblies), and the mirror rotation angle is α. The incident ray is defined by its orientation (σ_1) and displacement (d_1).

Expressions are derived enabling calculation of:

i) the orientation (σ_2) of the reflected ray relative to the x-axis ;

ii) the displacement (d_2) of the reflected ray along the x-axis ;

iii) the total optical path length imposed by the mirror.

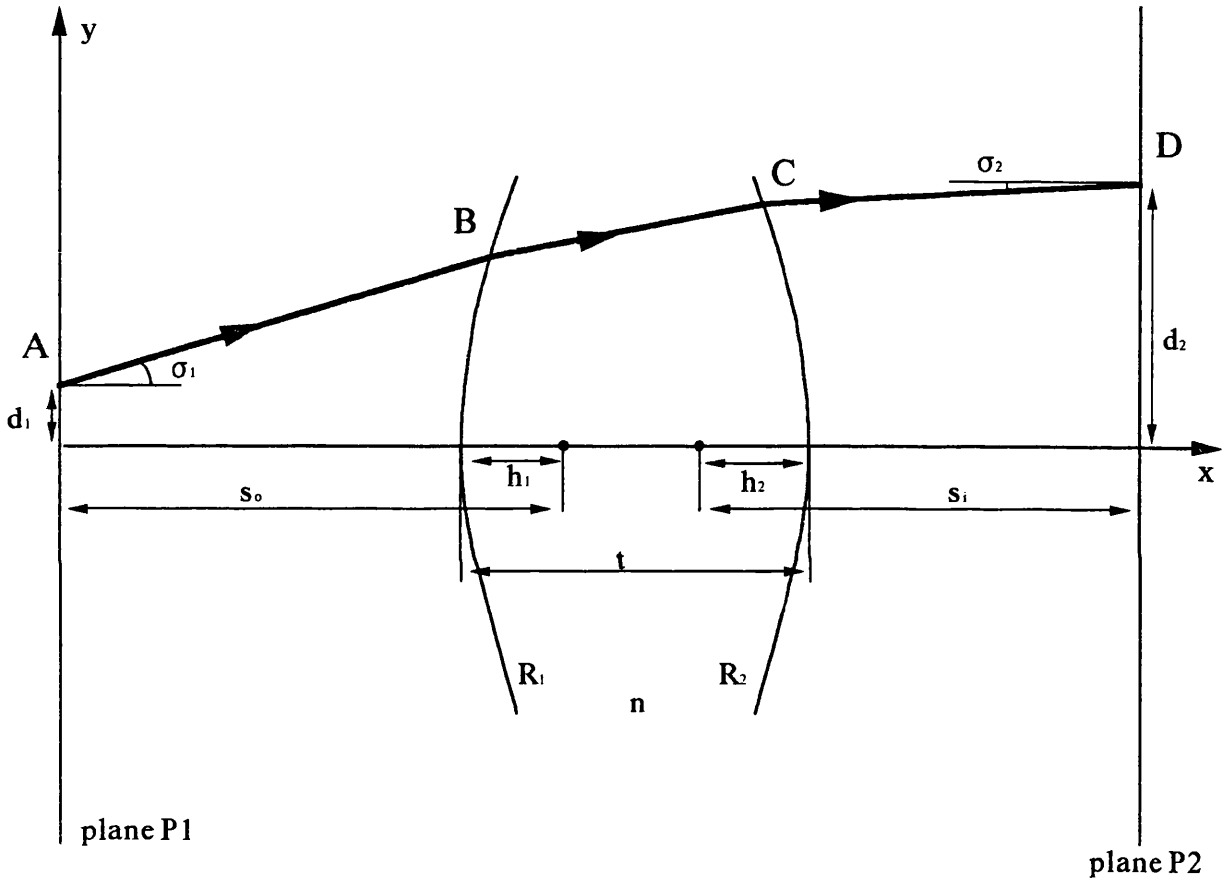


Figure 7-6 Tracing a single ray through lens

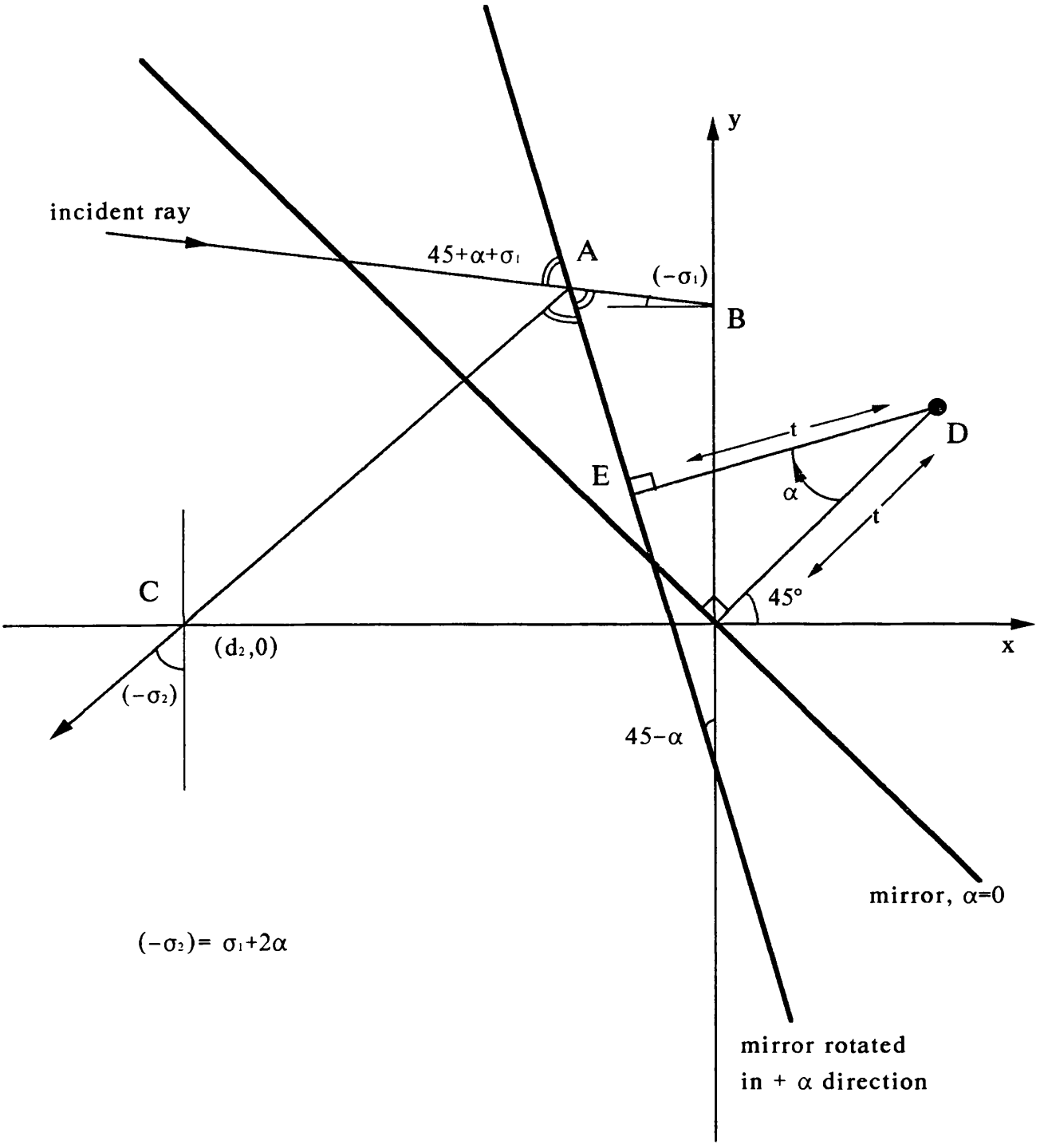


Figure 7-7 Tracing a single ray at rotated mirror

The algebraic expression derived for each set of calculations are best evaluated numerically. Each set can form the basis of a computer subroutine, enabling an entire optical system to be modelled by sequentially calling each routine. The distance between the different elements is determined by the s_o and s_i parameters.

The above description of the lens analysis defines the lens in terms of the radii of curvature of the surfaces, thickness and refractive index. These are related to the focal length according to the thick lens equation[1],

$$\frac{1}{f} = (n-1)\left[\frac{1}{R_1} + \frac{1}{R_2} - \frac{(n-1)\,t}{n\,R_1\,R_2}\right] \qquad ...(7-9)$$

The lens thickness must be sufficient so that the lens has a large enough aperture and does not block any light. Referring to figure 7–8, this shows a lens which has surfaces with radii of curvature of R_1 and R_2 . E and A are a distance of R_1 and R_2 respectively from the front and back surfaces of the lens. The thickness (t) of the lens is equal to BD and the lens aperture, R_L is equal to CF.

Therefore, $$AE = R_1 + R_2 - t \qquad ...(7-10)$$

AE is divided into two in the ratio x: (x– 1) so that,

$$AC = x\,.\,(AE) \qquad ...(7-11)$$

$$CE = (1-x)\,.\,(AE) \qquad ...(7-12)$$

Using Pythagoras in triangles ACF and CEF,

$$R_1^2 = R_L^2 + (1-x)^2\,(AE)^2 \qquad ...(7-13)$$

$$R_2^2 = R_L^2 + x^2\,(AE)^2 \qquad ...(7-14)$$

---

[1] "Optics" by E. Hecht and A Zajac, first edition, Addison-Wesley, 1974 (chapter 6).

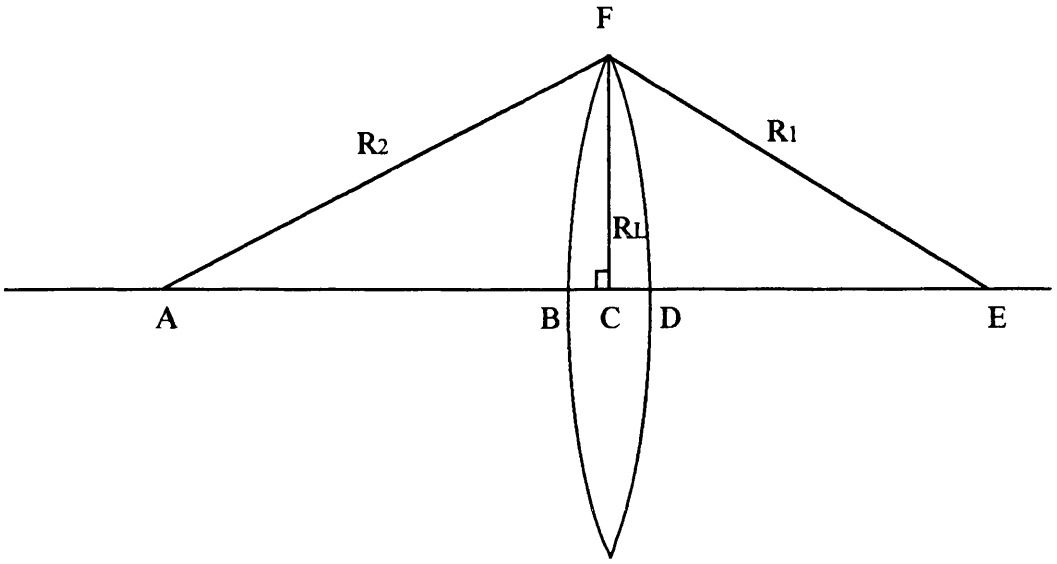


Figure 7-8 Calculating the lens aperture

Hence eliminating x, the radius of the lens aperture is,

$$R_L = \sqrt{R_1^2 - \frac{\left(R_2^2 - R_1^2 - (AE)^2\right)^2}{4\,(AE)^2}} \qquad \ldots(7-15)$$

This expression may be used to ensure that the lens aperture is always sufficiently large.

A ray tracing computer program has been written which implements the algorithms presented in appendix D and E. Lack of time however has not permitted completion of this work and the results from the program are not shown. There has also been no time available to implement these ideas in the laboratory.

7.5 Diffraction effects

Ray optics is useful only as long as lens aberrations are less than diffraction effects. After propagating a ray through a lens or past a mirror, the accuracy of values calculated for the orientation and displacement of the emerging ray is determined by diffraction. This is governed by the size of the beam at the element. So for example, if the diameter of the beam entering the element is d then the accuracy of the orientation calculated by ray optics will be no better than $\sim \lambda/d$, where λ is the optical wavelength.

The decision whether this accuracy is high enough will depend upon the specifications of the Bragg cell. If after tracing a ray around the whole system it is estimated that the overall accuracy of orientation is $\delta\sigma$, then a variation in orientation of the light entering the Bragg cell equal to this value should cause no perceivable change in the Bragg scattered light.

If it is found that the accuracy of the ray model is insufficient, then diffraction effects must be included. This requires propagating not a single ray around the system, but the full optical field. In order to take into account ray aberrations, a hybrid model may be formed. First using the ray tracing technique, phase factors can be found which describe the aberrations of each lens. These can then be used in diffraction theory such as that outlined in reference [2].

7.6 Summary

In this chapter we have discussed the application of a scanning mirror in a differential system. Only measurements made from detectors D2 or D3 are likely to be adversely affected by this kind of scanning arrangement. Chapter 5 concluded that differential intensity was best measured at detector D1. Therefore the effects of a scanning mirror on the differential phase response only from detectors D2 and D3 need be considered.

The scanning mirror must be placed at the back focal plane of the objective lens. In the absence of lens aberrations such an arrangement will not cause false differential phase contrast. Lens aberrations reduce the collimation of the light returning for a second time through the Bragg cell. This will affect the interference at detectors D2/ D3, which in turn will cause variations in the phase response as the mirror is rotated.

Lens aberrations can be modelled using ray tracing. This method has been implemented in computer software. There has however been insufficient time to complete the work, and no results have been shown. Considerable time has nevertheless been afforded developing this software and work is continuing in this area.

---

[2]J.W. Goodman, "Introduction to Fourier Optics", Physical and Quantum Electronics Series, McGraw Hill, 1968.

CHAPTER 8

CONCLUSIONS AND FURTHER WORK

8 Conclusions and further work

8.1 Conclusions

The aim of this project has been to develop an optical technique which can simultaneously measure small changes in the phase and intensity of light reflected from an object. This thesis has been concerned with defining the fundamental design criteria which are necessary in any practical implementation of the technique.

From the foregoing chapters we may draw the following conclusions:

i) The technique is a heterodyne interferometer in which both interfering beams are reflected from the object. The interference response is differential, equal to the phase difference of each reflected beam. This is a true common path technique, and phase noise due to microphonics is eliminated by performing the measurement above dc.

ii) A Bragg cell is used to divide the incident light into two closely spaced probing beams. It introduces a relative frequency shift in each beam which is necessary for the phase measurement and is further used to amplitude modulate the two probe beams in phase quadrature at frequency ω_a. Detection of the $2\omega_a$ frequency component of the total reflected light from the object yields the differential intensity response. This is equal to the difference between the intensity reflection coefficients of the two probed areas.

iii) By driving the Bragg cell with a single RF frequency, a single zeroth and first order beam are produced which may then be used as probe beams. This implementation of the technique is called the zeroth/ first order system. The variable R_{ss} was introduced to describe the ratio of the probe beam separation to focussed beam diameter. For imaging R_{ss} should be less than unity and for metrology applications, it should be greater than unity. It is unlikely that R_{ss} may be reduced below a value of 3 in the zeroth/ first order system, limiting it for the use in metrology only. This problem may be overcome by driving the Bragg cell with two separate RF frequencies, each amplitude modulated with identical signals which are in exact

phase quadrature. By changing the difference frequency, R_{ss} may be varied from a value much less than unity to well above. This arrangement therefore has the flexibility to be used for both imaging and metrology.

iv) A mathematical framework has been constructed relating the optical modulation of each probe beam, $g_1(t)$ and $g_2(t)$ to the differential intensity and phase response. A thorough description has been given of the acousto-optic interaction and a series expansion is given describing the two frequency plane wave Bragg interaction. This is then combined with the modulation theory to provide a complete description of the system, from the electrical drive to the detector responses. The effects on system performance of using triangular and sine/ cosine electrical modulation has been extensively investigated.

v) The responses of detectors at three positions around the system have been considered. Detector D1 which monitors the light before a second passage through the Bragg cell is suitable only for the measurement of differential intensity. Detector D2 responds to light after a second pass through the Bragg cell which is collinear with the initial laser light (ie., μ_1). This detector is good for both differential phase and intensity measurement. In imaging mode it is not possible to remove the close side beams, μ_2 and μ_3 . Therefore a further detector D3 must be considered. This detector however is only suitable for differential phase measurement.

vi) Two other responses above dc, interference intensity and absolute intensity, are also available from detectors D2 and D3. Interference intensity is the amplitude of the differential phase signal, and absolute intensity is proportional to the absolute reflectivity of the object, and is measured at frequency $4\omega_a$.

vii) The effects of imperfect modulation has been extensively investigated. Imperfect modulation refers to the situation whenever the so called differential intensity response is no longer exactly proportional to the difference in reflectivities of the two probed areas. The response will then include a proportion of absolute intensity information which reduces the differential intensity resolution considerably. There are three parameters which may be used to control the modulations. These are, the orientation of the Bragg cell relative to the light, and the phase and amplitude of the electrical drive signals. It was shown that in order to achieve a differential intensity resolution $\Delta R/R$ of better than 5×10^{-4} from detector D1, the electrical modulations

must be in phase quadrature with an accuracy of at least 0.5mrad, the factional difference in the electrical modulation amplitudes must be less than 1×10^{-3}, and the splitting difference of the light due to the orientation of the Bragg cell must be less than 0.02% in intensity.

viii) In order to fulfil the stringent modulation conditions, we have seen that an extra detector may be used to constantly monitor the total intensity of the light incident on the object, and in conjunction with feedback, continuously make any necessary adjustments. The feedback method is only suitable for use with detector D1.

ix) The differential phase sensitivity for measurement from both detectors D2 and D3, assuming shot noise limited performance, is 1×10^{-6} rad in a 1Hz bandwidth, using a laser with a power of 100μW.

x) The effect of intermodulation orders generated in the Bragg interaction have been studied for triangular and sine/ cosine electrical modulations. It was found that an error less than 1% will be introduced into the differential intensity response of detector D1. The differential phase response of detector D2 will have an error of less than 2%. Detector D3 will have a similar associated error providing there is no mixing between the beams μ_1, μ_2 and μ_3.

xi) The technique has been applied to the examination of three specimens, one with pure topography detail, the second with primarily reflectivity variation and a third with a mixture of phase and topography. For the latter specimen, a variation of the ratio R_{ss} between unity and a value above 8 was achieved. It was demonstrated how the differential phase, differential intensity as well a interference intensity information may be applied to the understanding of complex structures.

xii) Mirror scanning in a differential system has been discussed. It was shown that, in the absence of lens aberrations, the use of a scanning mirror at the back focal plane of the objective lens will cause no reduction in differential phase or intensity performance. Lens aberrations are most detrimental to the differential phase measurements made from detector D2 or D3, and will make little difference to the differential intensity response of detector D1. Ray tracing analysis was presented which can model the effects of lens aberrations.

8.2 Further work

8.2.1 Mirror scanning

As has been discussed in chapter 7, the use of a mirror to scan the light across the object is very attractive for imaging applications of the differential technique. A ray tracing model was described which implicitly takes into account the effects of lens aberrations, but insufficient time prevented the algorithms being fully implemented in software. This model can initially be used to study the performance of a differential mirror scanning system using a singlet objective lens. It will then be interesting to investigate the effects on performance when this is changed for a compound lens design such as those given in reference [1]. This work should be coupled with a practical implementation in the laboratory.

8.2.2 Indirect interference method

Modulation, which is an essential requirement for the differential intensity response plays no part whatsoever in the differential phase measurement. Indeed the very presence of the modulation is detrimental to the phase response : for optimum sensitivity, both beams must be 'on' the entire time, in other words, no modulation.

Even so, the shot noise limited phase sensitivity calculated in chapter 5 of 1×10^{-6} rad is a not at all poor. In experiments however, for some as yet unquantified reason, a signal to noise ratio for the interference could not be obtained better than about 60dB in a 300 Hz bandwidth (at least another 20dB would be expected). To improve this, as well as enable continuous interference to occur, an extension is proposed to the existing design.

---

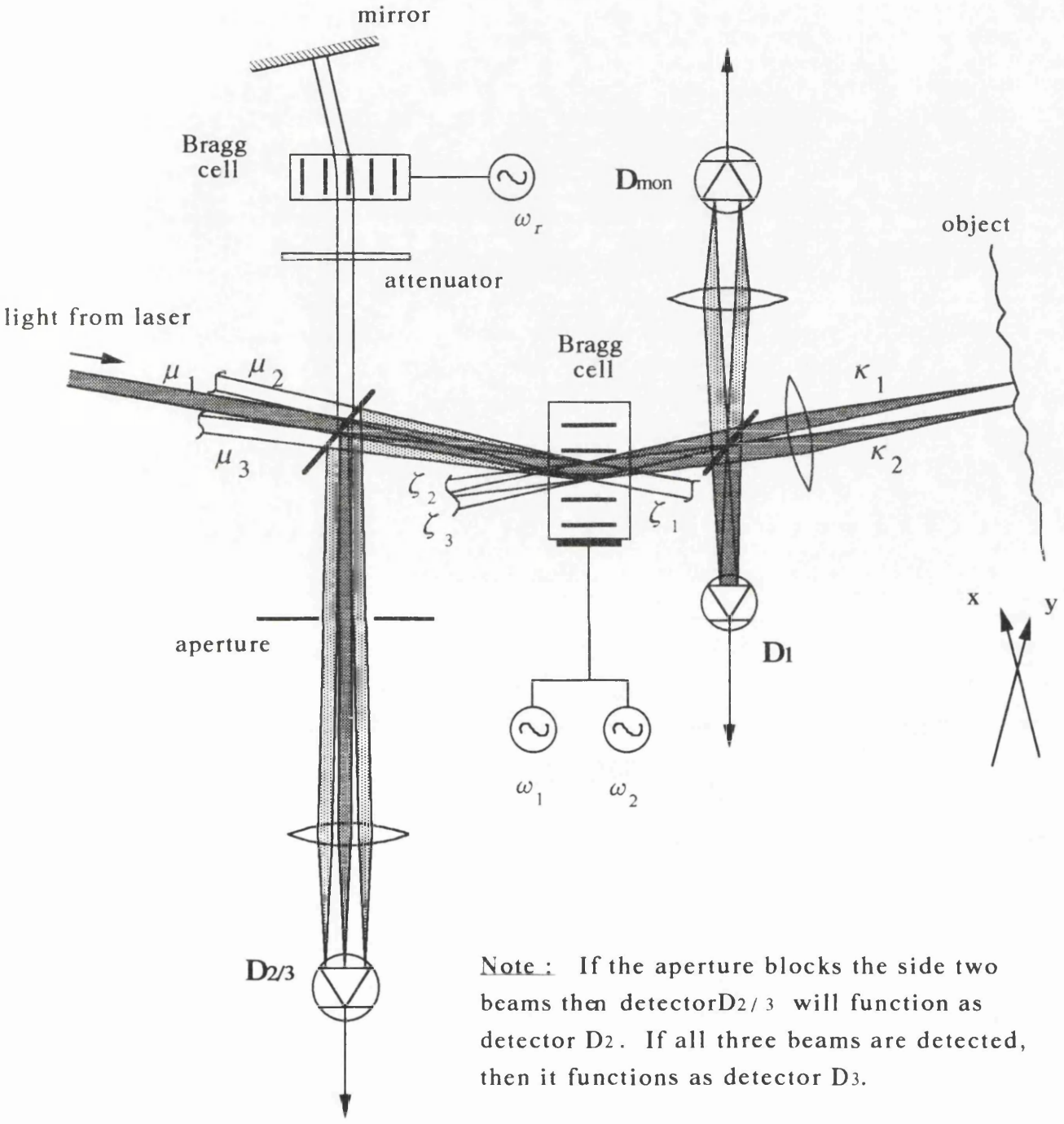
[1] "A system of optical design", by Arthur Cox, Focal press (London and New York).

Referring to figure 8-1, the light which was initially discarded at the first beam splitter, is now frequency shifted by $2\omega_r$ and reflected back to interfere with light from the sample. The system can now be considered as two superimposed Michelson interferometers, each with the same reference arm. Measurement of the phase of the interference signals from detector $D_{2/3}$ at frequencies, $2[\omega_1 - \omega_r]$ and $2[\omega_2 - \omega_r]$ gives the absolute phase variation of each probe beam. Since the reference and probe beams follow different paths, these phase responses will be severely limited by microphonics noise. The purpose of the $2\omega_r$ frequency shift in the reference is to lower the frequency of each of the above interference terms. This is not strictly necessary, but simplifies the detection electronics.

A most important feature of the optical layout is that both probe beams follow very similar paths and that the same reference beam is used for each interferometer. This means that, even though the individual interference signals will be noisy, the microphonics noise is identical for each of the interferometers. Hence by electronically mixing the two interference signals, the noise is eliminated and the differential phase can be measured at frequency $2[\omega_1 - \omega_2]$.

This method of running two heterodyne interferometers in tandem, obtaining two interference signals, and then electronically mixing to remove any common noise, we have called *indirect* interference. It should be emphasized that this is *our* term, and rarely used by others. The technique has previously been used by Massie et al., Huang and Offside et al. in absolute phase interferometers, and these methods are described in sections 2.1 and 2.2.3 (II, IV). This is the first application of the indirect interference method to this kind of differential system.

The resulting differential phase response is very similar to that from the original system described throughout the thesis. The advantage of this indirect interference method is that it is much easier to satisfy the requirements of both differential phase and differential intensity, ie., switching of the power between the probe beams whilst maintaining an interference signal. It is now possible to use square wave modulation, where the power is



Note : If the aperture blocks the side two beams then detector $D_{2/3}$ will function as detector D_2. If all three beams are detected, then it functions as detector D_3.

Figure 8-1 The two first order system with an additional reference arm for indirect interference

alternately switched between each probe beam. This type of modulation does not require the complicated analogue electronics which was required for sine/ cosine and triangular modulations, and it should also be easier to implement the feedback technique (using the signal from detector D_{mon} , as outlined in chapter 5, section 5.4).

The indirect interference method will result in larger interference signals, and hence greater phase sensitivity. The disadvantage of the method is that two interference signals are now detected, and so the shot noise will also increase. This will reduce the final phase sensitivity. In the next section, the phase sensitivity of the indirect interference system is derived.

The phase sensitivity of the indirect interference method

Using the formalism set out in section 3.2 of chapter 3, the amplitudes of the optical fields at detector D2/ D3 due to the two probe beams and the extra 'reference' beam are,

$$E_a = \sqrt{\rho_1}' E_o a_1 A_1 g_1^2(t)\, r_1 \exp i\left(\left[2\omega_1 + \omega_o\right] t + (2\theta_1 + \alpha_1) + \phi_1 + \phi_{m1}\right) \quad \ldots(8-1a)$$

$$E_b = \sqrt{\rho_1}' E_o a_2 A_2 g_2^2(t)\, r_2 \exp i\left(\left[2\omega_2 + \omega_o\right] t + (2\theta_2 + \alpha_2) + \phi_2 + \phi_{m2}\right) \quad \ldots(8-1b)$$

and

$$E_r = \sqrt{\rho_r} E_o \exp i\left(\left[2\omega_r + \omega_o\right] t + \phi_r + \phi_{m3}\right) \quad \ldots(8-1c)$$

where ϕ_{m1}, ϕ_{m2} and ϕ_{m3} describe the phase noise due to microphonics ; ϕ_1, ϕ_2 and ϕ_3 are phase constants ; ω_1, ω_2 and ω_r are frequency shifts due to interaction in one or other Bragg cell, and ρ'_1 , ρ_r are constants which account for any losses of light at lenses and beam splitters etc. (they are equivalent to ρ_1, ρ_2 and ρ_3 used previously, and refer to intensity, hence the square root).

Since the two probe beams have a common path, their associated phase noise due to microphonics will be the same ie., $\phi_{m1} = \phi_{m2}$. Interference between the different beams at detector D2/ D3 will generate several frequency components in the electrical response, of which two are of particular interest :

$$I_{2(\omega_1 - \omega_r)} = 2\sqrt{\rho_1' \rho_r}\ \zeta P_L\ a_1 A_1\ |g_1^2(t)|_{dc}\ r_1 \times \cos\left\{2\left[\omega_1 - \omega_r\right] t + (2\theta_1 + \alpha_1) + (\phi_{m1} - \phi_{m3}) + (\phi_1 - \phi_r)\right\} \quad \ldots(8-2a)$$

$$I_{2(\omega_2 - \omega_r)} = 2\sqrt{\rho_1' \rho_r}\ \zeta P_L\ a_2 A_2\ |g_2^2(t)|_{dc}\ r_2 \times \cos\left\{2\left[\omega_2 - \omega_r\right] t + (2\theta_2 + \alpha_2) + (\phi_{m2} - \phi_{m3}) + (\phi_2 - \phi_r)\right\} \quad \ldots(8-2b)$$

These frequency components are separated using a bandpass filter, and mixed. Remembering that $\phi_{m1} = \phi_{m2}$, the resulting signal is,

$$I_{2(\omega_1 - \omega_2)} = 2\rho_1' \rho_r\ \zeta^2 P_L^2\ a_1 A_1\ a_2 A_2\ |g_1^2(t)|_{dc}\ |g_2^2(t)|_{dc}\ r_1 r_2 \times \cos\left\{2\left[\omega_1 - \omega_2\right] t + 2\delta\theta' + \phi'\right\} \quad \ldots(8-3)$$

where $\delta\theta' = \delta\theta + \dfrac{\delta\alpha}{2}$ is the differential phase, and ϕ' is a phase constant ($\phi' = \phi_1 - \phi_2$)

It can be seen that the noise due to microphonics (ie., ϕ_{m1} , ϕ_{m2} and ϕ_{m3}) has cancelled in the mixing process, and that the differential phase may be measured in a similar way as previously, using phase sensitive detection. The important difference

between the direct and indirect interference response is now apparent :

the $|g_1^2(t)\, g_2^2(t)|_{dc}$ term for direct interference (equation 3 – 46) has been replaced by $|g_1^2(t)|_{dc}\, |g_2^2(t)|_{dc}$ for indirect interference, as shown in the above equation. It is this difference which allows the use of square wave modulation, with the two probe beams intensity modulated in anti-phase.

The signal to noise calculation is complicated by the fact that each frequency component is separately detected. Since these are in different frequency bands, it can probably be assumed that the associated noise of each signal is uncorrelated (the worst case scenario, and probably the correct one). The expression for shot noise limited coherent detection is then (modifying equation 5 – 25 with an extra factor of 2),

$$S/N = \frac{E_c^2\,(2\delta\theta')^2}{2\,e\,\overline{I_D}\,B\,F} \qquad \ldots (8-4a)$$

where B is the bandwidth of the measurement, F is the noise figure of the detector pre-amplifier, and E_c is the magnitude of the $2(\omega_r-\omega_1)$ or $2(\omega_2-\omega_1)$ frequency component of the current measured at detector D2/D3. This is obtained from equation (8 – 2*a*) :

$$E_c = \frac{\sqrt{\nu}}{4}\,\zeta P_L\,|g_1^2(t)|_{dc} \qquad \ldots (8-4b)$$

$\overline{I_D}$ is the time averaged current at the detector :

$$\overline{I_D} = \frac{|E_a|^2 + |E_b|^2 + |E_r|^2}{|E_o|^2} \times \zeta P_L$$

and using equations (8 – 1a, b, c),

$$\overline{I_D} = \frac{\zeta P_L}{16} \times \left(\overline{g_1^4(t)} + \overline{g_2^4(t)} + 4\nu \right) \qquad \ldots (8-4c)$$

In writing down equations (8 – 4*b,c*), the following assumptions have been made :

i) $a_1 \approx a_2 \approx A_1 \approx A_2 \approx 1$

ii) $r_1 \approx r_2 \approx 1$

iii) 50:50 beam splitters are used, and the attenuation of the reference beam is ν (ν can range between zero for full attenuation and unity for none).

Hence $\rho'_1 = {}^1/_{16}$ and $\rho_r = {}^{\nu}/_4$.

iv) there are no losses of light in the Bragg cell or lenses.

The limiting sensitivity is when the signal to noise ratio is equal to unity.

Hence, $$(\delta\theta')_{\text{res}} = \sqrt{\frac{e\,\overline{I_D}\,BF}{2E_c^2}} \qquad ...(8-5)$$

and using expressions (8 – 4*b,c*),

$$(\delta\theta')_{\text{res}} = \sqrt{\frac{e\ BF}{2\zeta P_L} \times \frac{\overline{g_1^4(t)} + \overline{g_2^4(t)} + 4v}{v\,|g_1^2(t)|_{dc}^2}} \qquad ...(8-6)$$

It follows from this expression that maximum sensitivity is when $v = 1$. The values used previously in section 5.3.2, of F = 2, B=1Hz, P_L=100μW and ζ= 0.4083 can be used to evaluate the constant term,

$$\frac{e\ BF}{2\zeta P_L} = 3.924\text{x}10^{-15}$$

We will now consider two types of modulation regimes.

CASE I triangular or sine/ cosine modulation as used in chapter 5

Using computer simulations as described in chapter 5 (with the drive parameters v_1, v_1 varying between ±3.1), the following values were calculated : $\overline{g_1^4(t)} + \overline{g_2^4(t)} = 0.589$, and $|g_1^2(t)|_{dc}^2 = 0.165$. Hence the phase resolution is $3.30\text{x}10^{-7}$ rad in a 1 Hz bandwidth with a laser power of 100μW.

CASE II square modulation

The two modulations (g_1 ,g_1) are two square waves with equal mark space ratios, varying between $v = 0$ and 3.1 , and with a 180° phase shift. between each modulation. We then have, $\overline{g_1^4(t)} + \overline{g_2^4(t)} = 1.0$, and $|g_1^2(t)|_{dc} = 0.5$.

Hence the phase resolution is $2.8\text{x}10^{-7}$ rad in a 1 Hz bandwidth with a laser power of 100μW.

Summary

From the above calculations, indirect interference results in a five fold improvement in sensitivity over the direct method (sensitivity of detector D2 from chapter 5 is 1.6×10^{-6} rad). Since there is now continuous interference with the reference beam, it is no longer necessary to have both beams on at the same time, and the use of square wave type modulation leads to a further small improvement in sensitivity. More important however is that TTL type modulation may be controlled with far higher precision than the analogue modulations considered throughout this thesis. An additional bonus of this type of modulation will be that intermodulation beams are now totally absent. It therefore seems that, to meet the stringent modulation requirements necessary for the differential intensity measurement, further work in this area is certainly worth pursuing.

Appendix A

Response of the beam deflector differential system

Appendix A Response of the beam deflector differential system

In this appendix, we consider the heterodyne response of the beam deflector differential system. The system is shown in figure A – 1, and is described fully in section 2.2.2 –I, to which the results derived in this appendix also relate.

We will first define the following symbols which are used in the analysis.

- ω_o optical frequency of light from laser;
- ω_B frequency shift per passage through the Bragg cell;
- ω_s deflector frequency;
- r amplitude reflection coefficient of probed area.

It is assumed throughout that there is no variation in conductivity over the object and the reflectivities are always real.

Since the beam is moving over the sample, the reflectivity, r, experienced by the probing beam will vary with time. So we can write,

$$r = r_o \left[1 + \frac{\delta r}{r_o} \cos(\omega_s t + \phi_1) \right] \qquad (A-1)$$

where r_o is the mean reflectivity experienced by the beam, δr is the variation in reflectivity at the extremes of deflection, and ϕ_1 is a constant.

We will assume that the reflectivity variation (δr) is small, ie., $\left[\frac{\delta r}{r_o} \right] << 1$ (A – 2)

The electric field amplitude of the reference beam at the detector is,

$$E_r = A_r \exp\left[i((\omega_o + 2\omega_B) t + \phi_r) \right] \qquad (A-3)$$

where A_r and θ_r are constants.

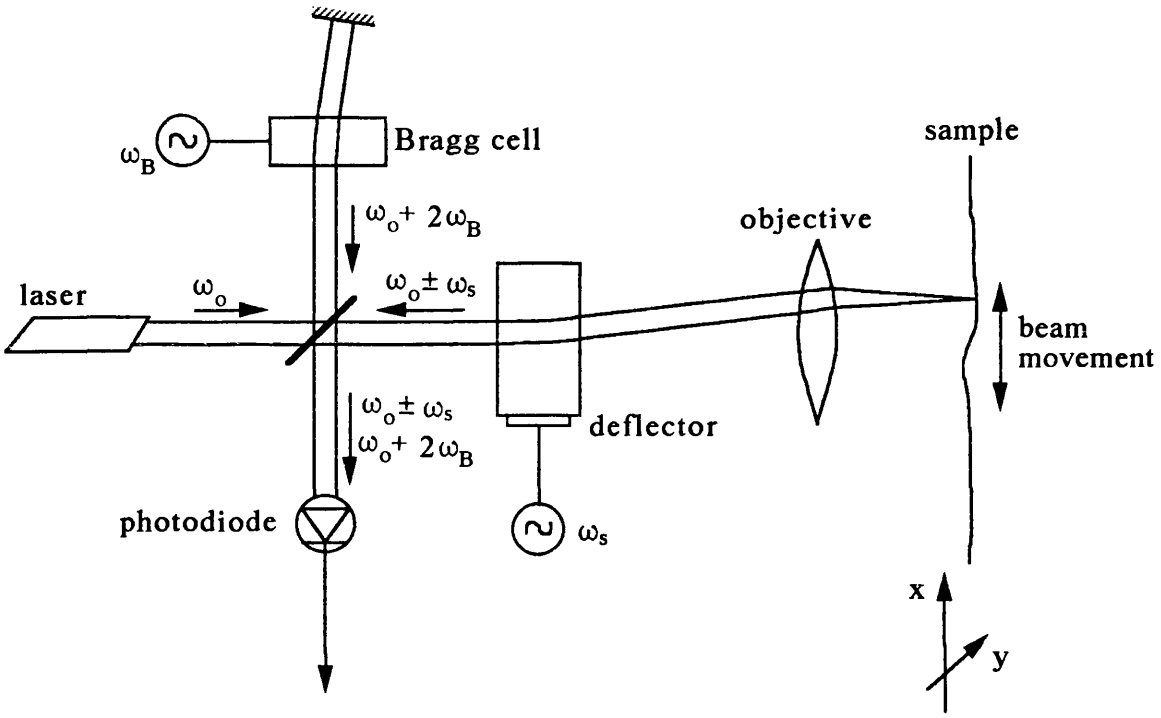


Figure A-1 Differential phase contrast optical microscope using beam deflector

The electric field amplitude of the probe beam at the detector is,

$$E_p = A_p\, r\, \exp\left[i(\omega_0 t + \delta\theta \cos(\omega_s t + \phi_1) + \phi_p) \right] \qquad (A-4)$$

where $\delta\theta$ is the phase variation due to object topography and conductivity of the object, and A_p , θ_p are constants.

The response of the photodiode is,

$$I \propto (E_r + E_p)(E_r + E_p)^*$$

$$\text{ie.,} \quad I \propto |E_r|^2 + |E_p|^2 + E_r E_p^* + E_r^* E_p \qquad (A-5)$$

Examining each term in turn,

i) $|E_r|^2$ contributes at dc only ;

ii) $$|E_p|^2 = A_p^2\, |r|^2 = A_p^2\, r_o^2 \left[1 + \frac{\delta r}{r_o} \cos\left[\omega_s t + \phi_1 \right] \right]^2 \qquad (A-6)$$

and using the assumption of (A - 2),

$$|E_p|^2 = A_p^2\, r_o^2 \left[1 + 2\frac{\delta r}{r_o} \cos\left[\omega_s t + \phi_1 \right] \right] \qquad (A-7)$$

Hence there is a response at frequency ω_s , and the amplitude of this frequency component is proportional to $(r_o \delta r)$. This the *differential intensity* response.

iii) The cross terms due to interference between the two beams,

$$E_{xx} = E_r E_p^* + E_r^* E_p$$

$$= A_r A_p r_o \left[1 + \frac{\delta r}{r_o} \cos\left[\omega_s t + \phi_1\right]\right] \cos\left[2\omega_B t - \delta\theta \cos(\omega_s t + \phi_1) + \phi_c\right]$$

...(A – 8)

where $\phi_c = \phi_r - \phi_p$ is a phase constant. ... (A – 8*a*)

With the assumption of small phase variations, ie., $\delta\theta << 1$, we can write,

$$E_{xx} = A_r A_p r_o \left[1 + \frac{\delta r}{r_o} \cos\left[\omega_s t + \phi_1\right]\right]$$

$$\times \left\{\cos(2\omega_B t + \phi_c) + \delta\theta \cos(\omega_s t + \phi_1) \sin(2\omega_B t + \phi_c)\right\}$$

...(A – 9)

Neglecting terms in $\delta\theta\,\delta r$,

$$E_{xx} = A_r A_p r_o \left\{\cos(2\omega_B t + \phi_c) + \frac{\delta r}{r_o} \cos(\omega_s t + \phi_1) \cos(2\omega_B t + \phi_c)\right.$$

$$\left. + \delta\theta \cos(\omega_s t + \phi_1) \sin(2\omega_B t + \phi_c)\right\}$$

$$= A_r A_p r_o \left\{\cos(2\omega_B t + \phi_c) + \frac{1}{2}\left[\frac{\delta r}{r_o} \cos(sum) + \delta\theta \sin(sum)\right]\right.$$

$$\left. + \frac{1}{2}\left[\frac{\delta r}{r_o} \cos(diff) + \delta\theta \sin(diff)\right]\right\}$$

...(A – 10)

where $$sum = (2\omega_B + \omega_s)\,t + \phi_c + \phi_1 \qquad (A-11a)$$

and $$diff = (2\omega_B - \omega_s)\,t + \phi_c - \phi_1 \qquad (A-11b)$$

Therefore, the voltage level of each sideband is proportional to,

$$\sqrt{\left[\frac{\delta r}{r_o}\right]^2 + (\delta\theta)^2} \qquad (A-12)$$

This response is a mixture of differential phase and differential intensity.

Any microphonics noise can be included in the ϕ_c term. If either of the sidebands in (A – 10) are mixed with the carrier ($2\,\omega_B$), then the resulting component at frequency ω_s will have an amplitude proportional to (A – 12) whilst the ϕ_c term is eliminated. This is the method used experimentally to obtain a response containing differential phase information *without* any microphonics noise.

Appendix B

Response of two beam differential phase system to a constant phase gradient

Appendix B Response of two beam differential phase system to a constant phase gradient

In this appendix, we consider the response of the two beam differential phase interferometer which is the subject of this thesis. The expression derived (equation B - 4) relates to the discussion in section 2.2.2 - section IV.

Consider a phase object with transmissivity,

$$t(x) = \exp(i\phi x) \qquad \text{(B - 1)}$$

where ϕ is the phase gradient.

The fourier transform of this is,

$$T(m) = \delta\left(m - \frac{\phi}{2\pi}\right) \qquad \text{(B - 2)}$$

It is shown in reference [B-1] that the intensity response of an interferometer system is,

$$I(x) = \int_{-\infty}^{\infty}\int C(m; p)\{T_1(m) + T_2(m)\}\{T_1^*(p) + T_2^*(p)\}\exp 2\pi i(m-p)x \ \mathrm{dm\ dp} \qquad \text{...(B - 3)}$$

where T_1 and T_2 are the Fourier transforms of $t(x)$ encountered by the sample and reference beams of the interferometer respectively.

Clearly for this type of differential interferometer, the sample and reference beams are not distinguished, the only difference being that one beam experiences a transmissivity of $t(x+\varepsilon)$, whilst the other experiences $t(x-\varepsilon)$, where ε is the separation of the two beams on the object surface. In Fourier space, T_1 is thus replaced by $T(m)\exp 2\pi i m\varepsilon$ and T_2 is replaced by $T(m)\exp -2\pi i m\varepsilon$. A further phase

B-1 C.J.R. Sheppard and T. Wilson, "Fourier imaging of phase information in scanning and conventional microscopes", Phil. Trans. Roy. Soc. Lond. **A295**, pp.513-536, (1980).

factor is introduced into T_2 by virtue of the fact the interferometer is heterodyne so that there is a continuous relative phase change between the two beams. This means that for a heterodyne interferometer, T_2 is replaced with $T(m) \exp(-2\pi i m\varepsilon) \exp(i2\Delta\omega_B t)$, where $2\Delta\omega_B$ is the total frequency difference between the two beams.

Substituting into (B – 3) gives,

$$I = 2C\left(\frac{\phi}{2\pi}; \frac{\phi}{2\pi}\right)\{1 + \cos[\lambda 4\omega_B t - 2\phi\varepsilon]\} \qquad \text{(B – 4)}$$

Appendix C

The SNR for phase measurement using coherent detection

Appendix C The SNR for phase measurement using coherent detection (referenced in section 5.3.2)

The analysis presented here follows that set down in reference [C-1].

Consider the heterodyne frequency component,

$$I(x, y; t) = E_c \cos\ (\omega t + \theta + \phi) + i_n(t) \qquad \ldots(C-1)$$

where ω is the measurement frequency, $\theta(x,y)$ is the phase variation due to the object, $i_n(t)$ is the noise current, and E_c, ϕ are constants.

Coherent detection can be performed by a lock-in amplifier. This operates by taking the reference input and dividing it into two parts in phase quadrature, $E_o \cos(\omega t + \phi)$ $E_o \sin(\omega t + \phi)$. The signal current is also divided into two and then each part is mixed with one of the quadrature references.

The resulting signals are,

$$I_s = \frac{1}{4} E_c E_o \left[\sin\theta + \sin(2\omega t + \theta + 2\phi)\right] + \frac{1}{2} E_o\, i_n(t)\ \sin(2\omega t + \phi) \qquad \ldots(C-2a)$$

and

$$I_c = \frac{1}{4} E_c E_o \left[\cos\theta + \cos(2\omega t + \theta + 2\phi)\right] + \frac{1}{2} E_o\, i_n(t)\ \cos(2\omega t + \phi)\,. \qquad \ldots(C-2b)$$

The first terms within the square brackets are the required phase information at dc, and the higher frequency $2\omega t$ terms are removed by filtering.

C-1 J. Brown and E.V.D. Glazier, 'Telecommunications', pub. by Chapman and Hall, 1964, p.157

The noise current can be expressed in terms of its quadrature components,

$$i_n(t) = x_n(t)\cos(2\omega t + \phi) + y_n(t)\sin(2\omega t + \phi) \qquad \ldots(C-3a)$$

where
$$\overline{i_n^2(t)} = \overline{x_n^2(t)} + \overline{y_n^2(t)} \qquad \ldots(C-3b)$$

and
$$\overline{x_n^2(t)} = \overline{y_n^2(t)} \qquad \ldots(C-3c)$$

Substituting equation (C – 3*a*) into (C – 2*a,b*) gives,

$$I_s = \frac{1}{4}E_cE_o\sin\theta + \frac{1}{4}E_o\left[y_n(t) + x_n(t)\sin(2\omega t + 2\phi) - y_n(t)\cos(2\omega t + 2\phi)\right] \qquad \ldots(C-4a)$$

$$I_c = \frac{1}{4}E_cE_o\cos\theta + \frac{1}{4}E_o\left[x_n(t) + x_n(t)\cos(2\omega t + 2\phi) + y_n(t)\sin(2\omega t + 2\phi)\right] \qquad \ldots(C-4b)$$

If the final measurement bandwidth is B, then $x_n(t)$ and $y_n(t)$ will contain frequency components only within the baseband, ie., 0 to $\frac{1}{2}B$ Hz. A low pass filter is used to filter out all the noise terms except the first in each of the square brackets.

Hence,

$$I_s = \frac{1}{4}E_cE_o\sin\theta + \frac{1}{4}E_o\,y_n(t) \qquad \ldots(C-5a)$$

$$I_c = \frac{1}{4}E_cE_o\cos\theta + \frac{1}{4}E_o\,x_n(t) \qquad \ldots(C-5b)$$

The lock-in amplifier will detect each of the quadrature components separately, and derive a response proportional to the phase θ and the carrier amplitude r:

$$I_\theta = \alpha_1 \left\{ \theta + \frac{\sqrt{\overline{x_n^2(t)}}}{E_c} \right\} \qquad \ldots(C-6a)$$

$$I_r = \alpha_2 \left\{ E_c + \sqrt{\overline{x_n^2(t)}} \right\} \qquad \ldots(C-6b)$$

where α_1 and α_2 are constants.

The signal to noise ratio (in power) for phase measurement is then,

$$(S/N)_\theta = \frac{E_c^2 \theta^2}{\overline{x_n^2(t)}} \qquad \ldots(C-7)$$

or using equation (C – 3),

$$(S/N)_\theta = \frac{E_c^2 \theta^2}{\frac{1}{2}\overline{i_n^2(t)}} \; . \qquad \ldots(C-8)$$

In a similar way, the signal to noise ratio for measurement of the carrier amplitude is,

$$(S/N)_r = \frac{E_c^2}{\frac{1}{2}\overline{i_n^2(t)}} \qquad \ldots(C-9)$$

The factor of a half in the denominators of equations (C – 8) and (C – 9) is a consequence of coherent detection and arises because one of the quadrature components of the noise is lost during detection.

Appendix D

Ray tracing through a lens

Appendix D Ray tracing through a lens (referenced in section 7.4)

This appendix traces a ray between two planes, P1 and P2, via a thick lens, as shown in figure D–1. The lens is defined by its radii of curvature, R_1 R_2 , thickness t, and refractive index n. The planes P1 and P2 are located at a distance of s_o and s_i from the principal points of the lens (see figure D–1). The ray crosses plane P1 with a displacement d_1 along the y axis and orientation of σ_1 , and crosses plane P2 with a displacement of d_2 and orientation of σ_2 .

The definition of the starting parameters is as follows :

σ_1 : orientation of incoming ray (+ve if ray has +y direction component) ;

d_1 : displacement of incoming ray along y axis ;

n : refractive index of lens ;

t : thickness of lens (ie., along central axis) ;

R_1 , R_2 : radii of curvature of the two lens surfaces. Each respective radius of curvature is defined as positive if the resulting lens surface is convex. Note for numerical computation it is easier to use $1/R_1$ and $1/R_2$;

s_o : distance between plane P1 and first principal point of lens ;

s_i : distance between second principal point of lens and plane P2 ;

The aim of the analysis is to find expression for the following :

σ_2 : orientation of ray leaving plane P2 (+ve if ray has +y direction component) ;

d_2 : displacement of ray leaving plane P2 along y axis ;

Path : the total optical path length along ABCD.

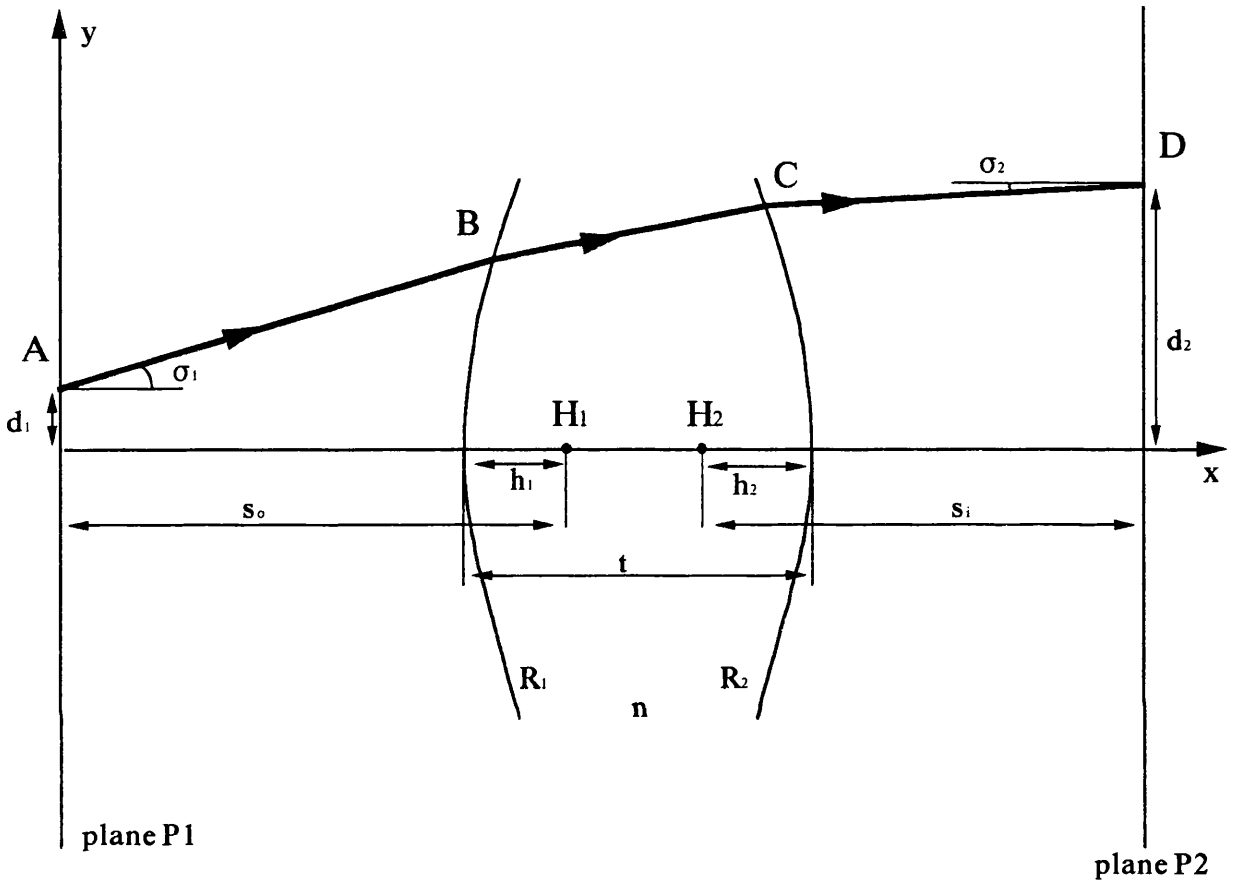


Figure D-1 Tracing a single ray through lens

The focal length of a thick lens is related to the lens curvatures, refractive index and thickness as follows[1] :

$$\frac{1}{f} = (n-1)\left[\frac{1}{R_1} + \frac{1}{R_2} - \frac{(n-1)\,t}{n\,R_1\,R_2}\right] \qquad \ldots(D-1)$$

The third term is equal to zero for a thin lens, and it can be seen that the effect of the lens thickness is to slightly increase the focal length.

The principal points of the lens are located at H_1 , H_2 which are a distance h_1 and h_2 from the lens surfaces as shown on figure D–1. h_1 and h_2 are given by the following expressions[1] :

$$h_1 = \frac{f(n-1)\,t}{R_2 n} \qquad \ldots(D-2a)$$

$$h_2 = \frac{f(n-1)\,t}{R_1 n} \qquad \ldots(D-2b)$$

We will now trace the ray between planes P1 and P2. This is done in three stages. Note that along the way, after an expression is first given, terms within it may be juggled around to produce an expression which is more accurate for numerical computation purposes. The reader should not be dismayed if occasionally expressions appear to be undergoing de-simplification !

I tracing ray from plane P1 to front surface of lens (figure D–2)

We will start by writing the equation of the line from the object to the lens as,

$$y = m_1 x + d \qquad \ldots(D-3a)$$

where $$m_1 = \tan \sigma_1 \qquad \ldots(D-3b)$$

[1] "Optics" by E. Hecht and A Zajac, first edition, Addison-Wesley, 1974 (chapter 6).

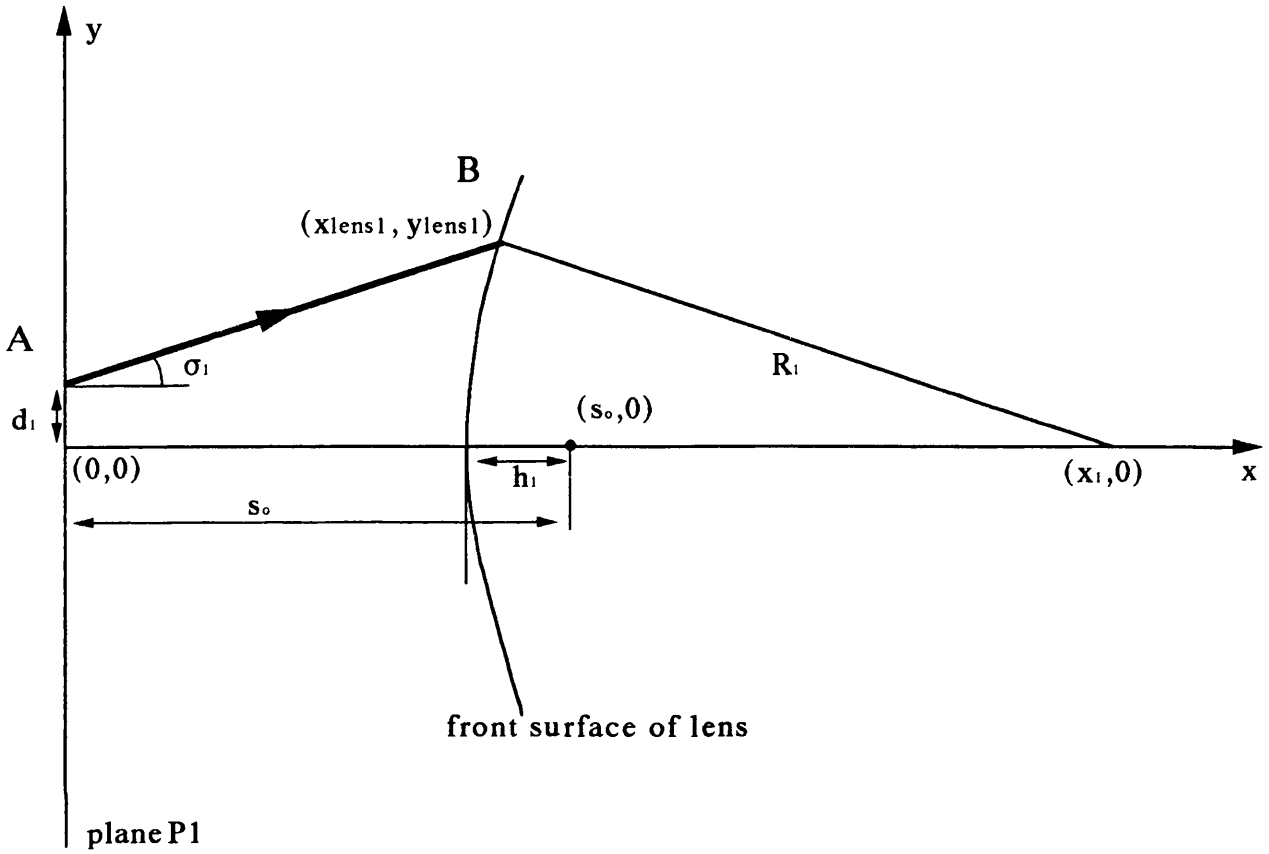


Figure D-2 Tracing a single ray to front surface of lens

The equation for the front surface of the lens is,

$$(x - x_1)^2 + y^2 = R_1^2 \qquad \ldots(D-4a)$$

where

$$x_1 = s_o + R_1 - h_1 \qquad \ldots(D-4b)$$

Eliminating y from the above two equations gives,

$$(x - x_1)^2 + (m_1 x + d_1)^2 - R_1^2 = 0$$

and expanding,

$$x^2\left(1 + m_1^2\right) + 2\left(m_1 d_1 - x_1\right)x + \left(x_1^2 + d_1^2 - R_1^2\right) = 0$$

$$\therefore \qquad x = \frac{(x_1 - m_1 d_1) \pm \sqrt{\left(m_1 d_1 - x_1\right)^2 - \left(1 + m_1^2\right)\left(x_1^2 + d_1^2 - R_1^2\right)}}{\left(1 + m_1^2\right)} \qquad \ldots(D-5)$$

The ray enters the lens at B which is located at (*xlens1*, *ylens1*). Using expressions (D – 3*a*) and (D – 5), the coordinates of B are,

$$xlens1 = \frac{(x_1 - m_1 d_1) - S_{R_1}\sqrt{-2m_1 d_1 x_1 - m_1^2 x_1^2 - d_1^2 + R_1^2 + m_1^2 R_1^2}}{\left(1 + m_1^2\right)} \qquad \ldots(D-6a)$$

and

$$ylens1 = m_1\,(xlens1) + d_1 \qquad \ldots(D-6b)$$

where

$$S_{R_1} = +1 \quad \text{for} \quad R_1 \geq 0\,,$$
$$= -1 \quad \text{for} \quad R_1 < 0 \qquad \ldots(D-6c)$$

For a planar surface, the radius of curvature becomes infinite. It is therefore more useful to re-write equation (D – 6*a*) in terms of the reciprocal of R_1 . So we will define the

variable,

$$x_{1\,recp} = \frac{x_1}{R_1} = (s_o - h_1).\left[\frac{1}{R_1}\right] + 1 \qquad \ldots(D-7)$$

so that the expression of xlens1 now becomes,

$$xlens1 = \frac{(x_{1\,recp}) - \left(\frac{1}{R_1}\right) m_1 d_1 - S_{root1}}{\left(1 + m_1^2\right)\left(\frac{1}{R_1}\right)} \qquad \ldots(D-8a)$$

with

$$S_{root1} = \sqrt{-2m_1 d_1 \left[\frac{1}{R_1}\right](x_{1\,recp}) - m_1^2 (x_{1\,recp})^2 - d_1^2\left[\frac{1}{R_1}\right]^2 + 1 + m_1^2} \qquad \ldots(D-8b)$$

In the situation of the incoming ray being parallel to the optical axis (ie., σ_1 =0), both the numerator and denominator of (D – 8 *a*) are equal to zero. In this case,

$$xlens1 = s_o - h_1 \qquad \ldots(D-9)$$

II tracing ray between lens surfaces **(figure D–3)**

The figure shows the ray crossing between B and C on the lens surfaces. The centre of curvature of the two surfaces is at E and F.

The incident ray subtends an angle of θ_i to the normal of the lens surface where,

$$\theta_i = \sigma_1 + \sin^{-1}\left[\frac{ylens1}{R_1}\right] \qquad \ldots(D-10)$$

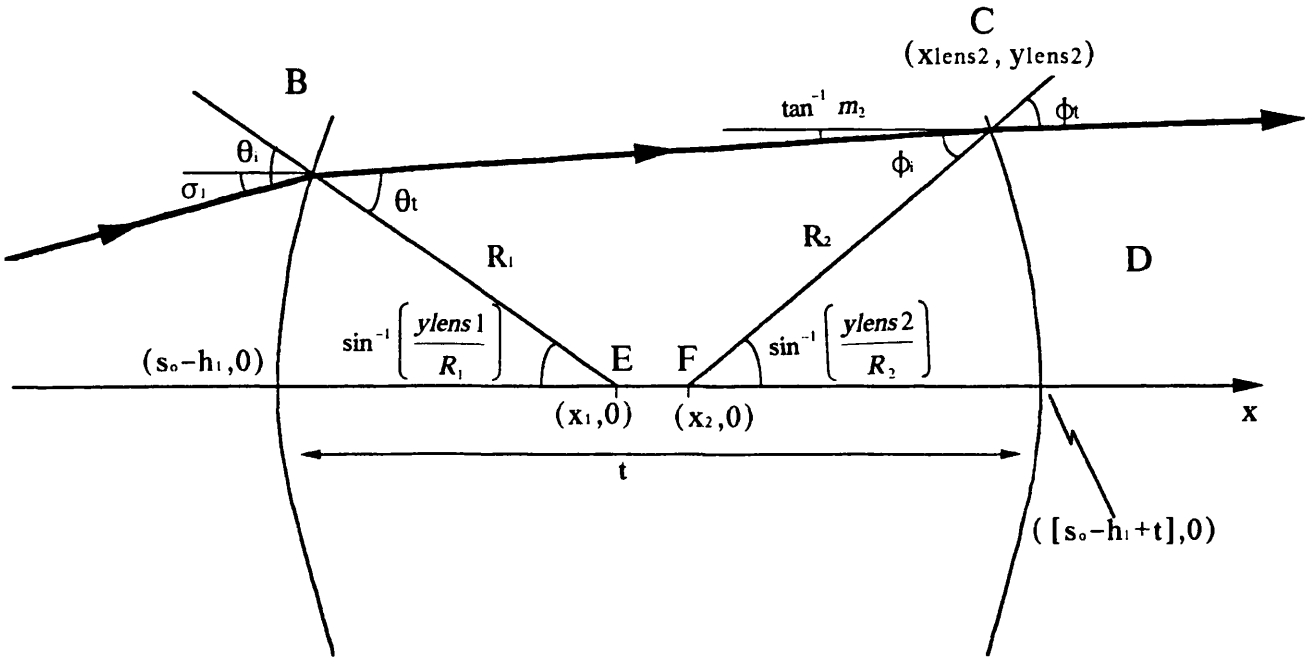


Figure D-3 Tracing a single ray between surfaces of lens

Using Snells law, the angle between the surface normal and refracted ray will be,

$$\theta_t = \sin^{-1}\left(\frac{1}{n}\sin\theta_i\right) \qquad \ldots(D-11a)$$

$$= \sin^{-1}\left\{\frac{1}{n}\sin\left[\sigma_1 + \sin^{-1}\left[\frac{ylens1}{R_1}\right]\right]\right\}$$

$$= \sin^{-1}\left\{\frac{1}{n}\sin\sigma_1\cos\left[\sin^{-1}\left[\frac{ylens1}{R_1}\right]\right] + \frac{1}{n}(\cos\sigma_1).\frac{ylens1}{R_1}\right\}$$

Hence

$$\theta_t = \sin^{-1}\left\{\frac{1}{n}\sin\sigma_1\sqrt{1-\left[\frac{ylens1}{R_1}\right]^2} + \frac{1}{n}(\cos\sigma_1).\frac{ylens1}{R_1}\right\}$$

$$\ldots(D-11b)$$

In deriving this expression, the following identity was used :

$$\cos\left(\sin^{-1}b\right) \equiv \sqrt{1-b^2} \qquad \ldots(D-12)$$

The reason for expanding the trigonometric expressions of equation (D – 11 *a*) to obtain (D – 11 *b*), is that the latter is prone to smaller computational errors.

We will write the equation of the line between the two surfaces as,

$$y = m_2 x + c_2 \qquad \ldots(D-13a)$$

where $$c_2 = (ylens1) - m_2\,(xlens1) \qquad \ldots(D-13b)$$

and $$m_2 = \tan(\theta_t - \theta_i + \sigma_1) \qquad \ldots(D-13c)$$

Expression (D – 13 *b*) can be simplified using equation (D – 3 *a*):

$$c_2 = (m_1 - m_2).xlens1 + d_1 \qquad \ldots(D-13d)$$

The equation for the second lens surface is,

$$(x - x_2)^2 + y^2 = R_2^2 \qquad \ldots(D - 14a)$$

where x_2 is the x-coordinate of F,

$$x_2 = s_o - h_1 + t - R_2 \qquad \ldots(D - 14b)$$

After substituting equation (D – 13*a*) into (D – 14*a*), we can calculate the point at which the ray crosses the second surface ie., the coordinates, (*xlens2*, *ylens2*) of C :

$$xlens2 = \frac{(x_2 - m_2 c_2) + S_{R_2}\sqrt{-2m_2 c_2 x_2 - m_2^2 x_2^2 - c_2^2 + R_2^2 + m_2^2 R_2^2}}{\left(1 + m_2^2\right)} \qquad \ldots(D - 15)$$

and

$$ylens2 = m_2\,(xlens2) + c_2 \qquad \ldots(D - 16)$$

In a similar way to the first surface of the lens, it is more useful to re-write the expression in terms of the reciprocal of the radius of curvature.

Hence if we define x_{2recp} such that,

$$x_{2recp} = \frac{x_2}{R_2} = (s_o - h_1 + t).\left[\frac{1}{R_2}\right] - 1 \qquad \ldots(D - 17)$$

then,

$$xlens2 = \frac{x_{2recp} - m_2 c_2\left(\frac{1}{R_2}\right) + \quad S_{root2}}{\left(1 + m_2^2\right)\left(\frac{1}{R_2}\right)} \qquad \ldots(D - 18a)$$

with

$$S_{root2} = \sqrt{-2m_2 c_2\,(x_{2recp}).\left[\frac{1}{R_2}\right] - m_2^2\,(x_{2recp})^2 - c_2^2\left[\frac{1}{R_2}\right]^2 + 1 + m_2^2}$$

$$\ldots(D - 18b)$$

and where, $S_{R_2} = +1 \quad \text{for} \quad R_2 \geq 0$,

$= -1 \quad \text{for} \quad R_2 < 0$ ...(D – 18 *c*)

In the situation of the ray between the two lens surfaces being parallel to the optical axis (ie., m_2 =0), both the numerator and denominator of (D – 18 *a*) are equal to zero. In this case,

$$xlens2 = s_o - h_1 + t \qquad \text{...(D – 19)}$$

Finally we must derive expression for the orientation of the light at the second lens surface.

The angle subtended by the ray and the surface normal at C is ,

$$\phi_i = \sin^{-1}\left(\frac{ylens2}{R_2}\right) - \tan^{-1} m_2 \qquad \text{...(D – 20)}$$

Using Snell's law the orientation of the light emerging from the lens surface is,

$$\phi_t = \sin^{-1}(n \sin \phi_i) \qquad \text{...(D – 21)}$$

Re-writing this using the above expression for ϕ_i ,

$$\phi_t = \sin^{-1}\left[n \sin\left\{\sin^{-1}\left(\frac{ylens2}{R_2}\right) - \tan^{-1} m_2\right\}\right] ,$$

and using the following identities,

$$\cos\left(\tan^{-1} b\right) \equiv \frac{1}{\sqrt{1 + b^2}} \qquad \text{...(D – 22 } a\text{)}$$

$$\cos\left(\sin^{-1} b\right) \equiv \sqrt{1 - b^2} \qquad \text{...(D – 22 } b\text{)}$$

$$\sin\left(\tan^{-1} b\right) \equiv \frac{b}{\sqrt{1 + b^2}} \qquad \text{...(D – 22 } c\text{)}$$

it follows that,

$$\phi_t = \sin^{-1}\left[n \frac{ylens2}{R_2} x \frac{1}{\sqrt{1 + m_2^2}} - n \sqrt{1 - \left[\frac{ylens2}{R_2}\right]^2} x \frac{m_2}{\sqrt{1 + m_2^2}} \right]$$

...(D – 23)

III tracing ray from second lens surface to plane P2 **(figure D–4)**

The ray will intercept plane P2 at (x_i, y_i) where,

$$x_i = s_o + s_i - h_1 + t - h_2 \qquad \text{...(D – 24)}$$

To find the value of y_i, we must write down the equation of the line from C on the lens surface to D on plane P2. This is,

$$y = m_3 x + c_3 \qquad \text{...(D – 25 a)}$$

where $$m_3 = \tan\left[\sin^{-1}\left(\frac{ylens2}{R_2}\right) - \phi_t\right] \qquad \text{...(D – 25 b)}$$

and $$c_3 = ylens2 - m_3\,(xlens2) \qquad \text{...(D – 25 c)}$$

Hence the y coordinate of C is,

$$y_i = m_3 x_i + c_3 \qquad \text{...(D – 26)}$$

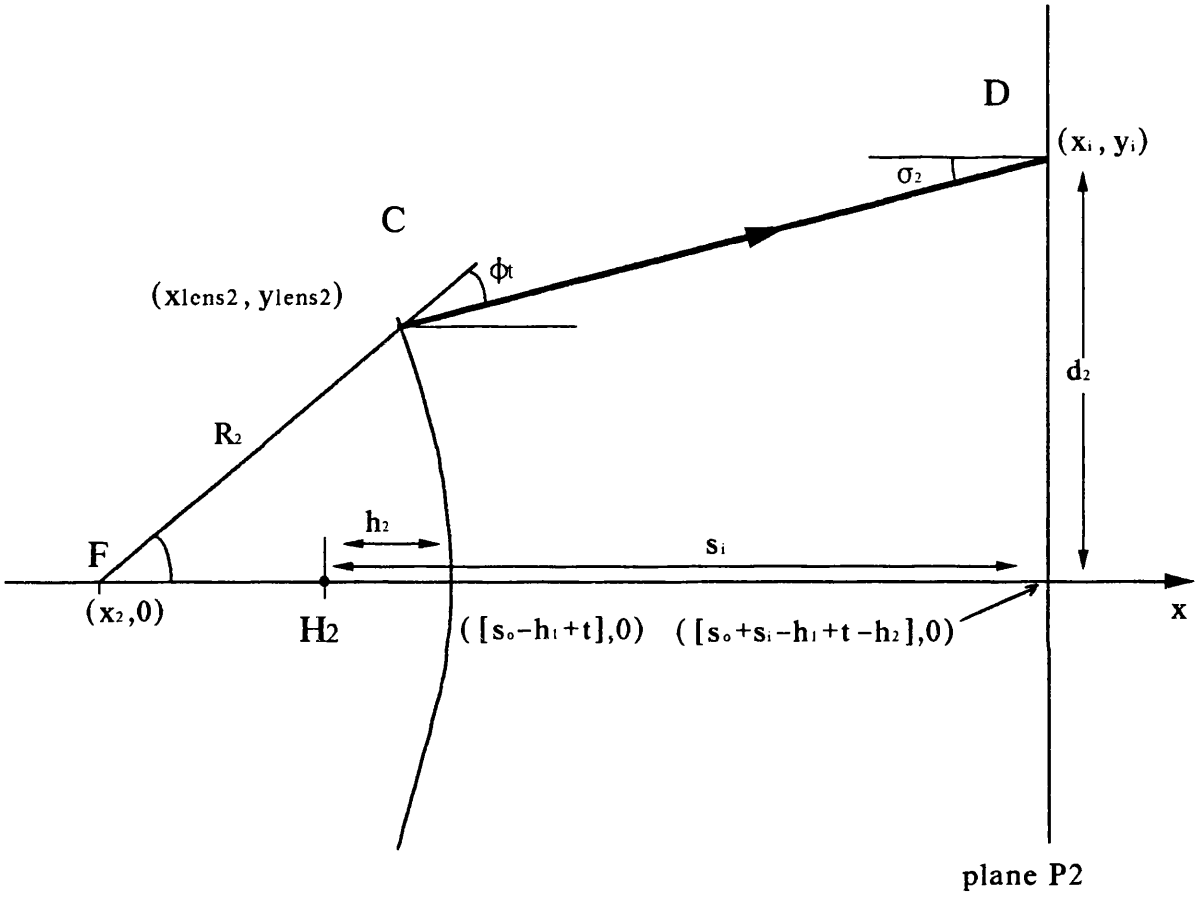


Figure D-4 Tracing a single ray from back surfaces of lens to plane P2.

IV Result (**figure D–5**)

The optical path lengths along AB, BC, and CD are as follows:

$$l_1 = \sqrt{(ylens1 - d_1)^2 + (xlens1)^2} \qquad \ldots(D-27a)$$

$$l_2 = \sqrt{(xlens2 - xlens1)^2 + (ylens2 - ylens1)^2} \qquad \ldots(D-27b)$$

$$l_3 = \sqrt{(x_i - xlens2)^2 + (y_i - ylens2)^2} \qquad \ldots(D-27c)$$

Hence the total optical path length is,

$$path = l_1 + l_3 + n.\, l_2 \qquad \ldots(D-28a)$$

The orientation of the ray leaving plane P2 is,

$$\sigma_2 = \tan^{-1} m_3 \qquad \ldots(D-28b)$$

and the displacement of the ray along the y-axis is,

$$d_2 = y_i \qquad \ldots(D-28c)$$

Equations (D – 28 *a,b,c*) can be evaluated numerically using the expressions derived in sections I, II and III.

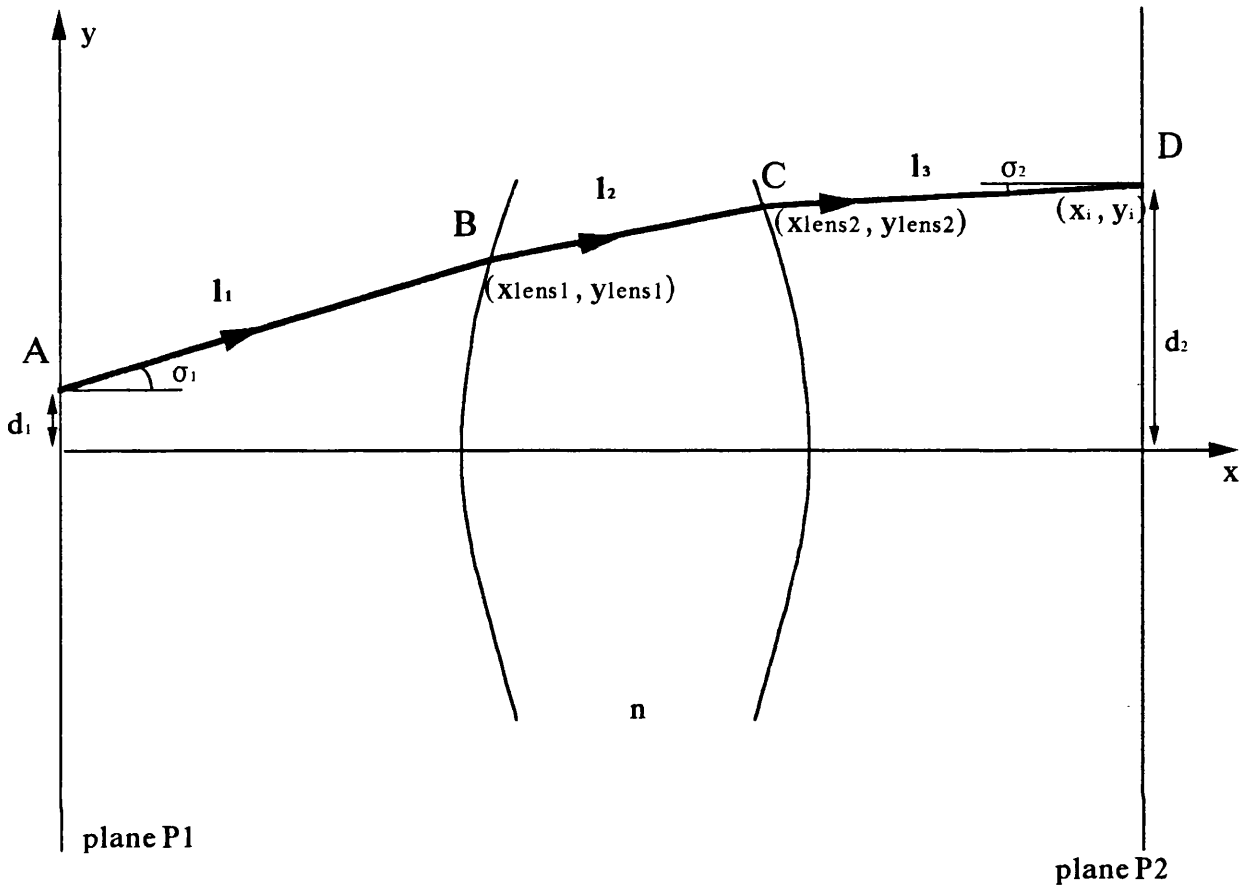


Figure D-5 The total optical path length between planes P1 and P2

Appendix E

Ray tracing at a scanning mirror

Appendix E Ray tracing at a scanning mirror (referenced in section 7.4)

This appendix traces a ray which is reflected from a scanning mirror, as shown in figure E-1. The mirror is pivoted about a point at a distance t behind the front surface. The optical axis of the elements preceding the mirror is along the x-axis, and the mirror is arranged so that when it is at zero rotation ($\alpha = 0$), the front surface of the mirror, the x-axis, and the y-axis all meet at a single point. The optical axis of the elements following the mirror is coincident with the y-axis (in the minus y direction).

Referring to figure E-1, a ray is incident on the mirror at A. As shown in the figure, the ray does not actually reach the y axis, but is instead reflected, and crosses the x axis at ($x = d_2$) with an orientation of ($-\sigma_2$). The basis of the analysis presented in this appendix is to consider the mirror as a transforming element. In the absence of the mirror, the ray would have intercepted the y-axis at ($y = d_1$), with an orientation of ($-\sigma_1$). Expressions are derived relating the sets of parameters, d_1 and σ_1 to d_2 and σ_2. An expression is also derived for the optical path length imposed by the presence of the mirror ie., (AC-AB) on figure E-1.

The definition of the starting parameters is as follows :

σ_1 : orientation of incoming ray (+ve if ray has +y direction component)
σ_1 must be within the range, $-\frac{\pi}{2} < \sigma_1 < \frac{\pi}{2}$;

d_1 : displacement of incoming ray along y axis at x = 0 ;

α : rotation angle of mirror (+ve if rotation is in clockwise direction)
α must be within the range, $-\frac{\pi}{4} < \alpha < +\frac{\pi}{4}$;

t : thickness of mirror ($t \geq 0$) ;

A further constraint is that σ_1 and α must always fulfil the condition, $-\frac{\pi}{2} < \sigma_1 + 2\alpha < \frac{\pi}{2}$.

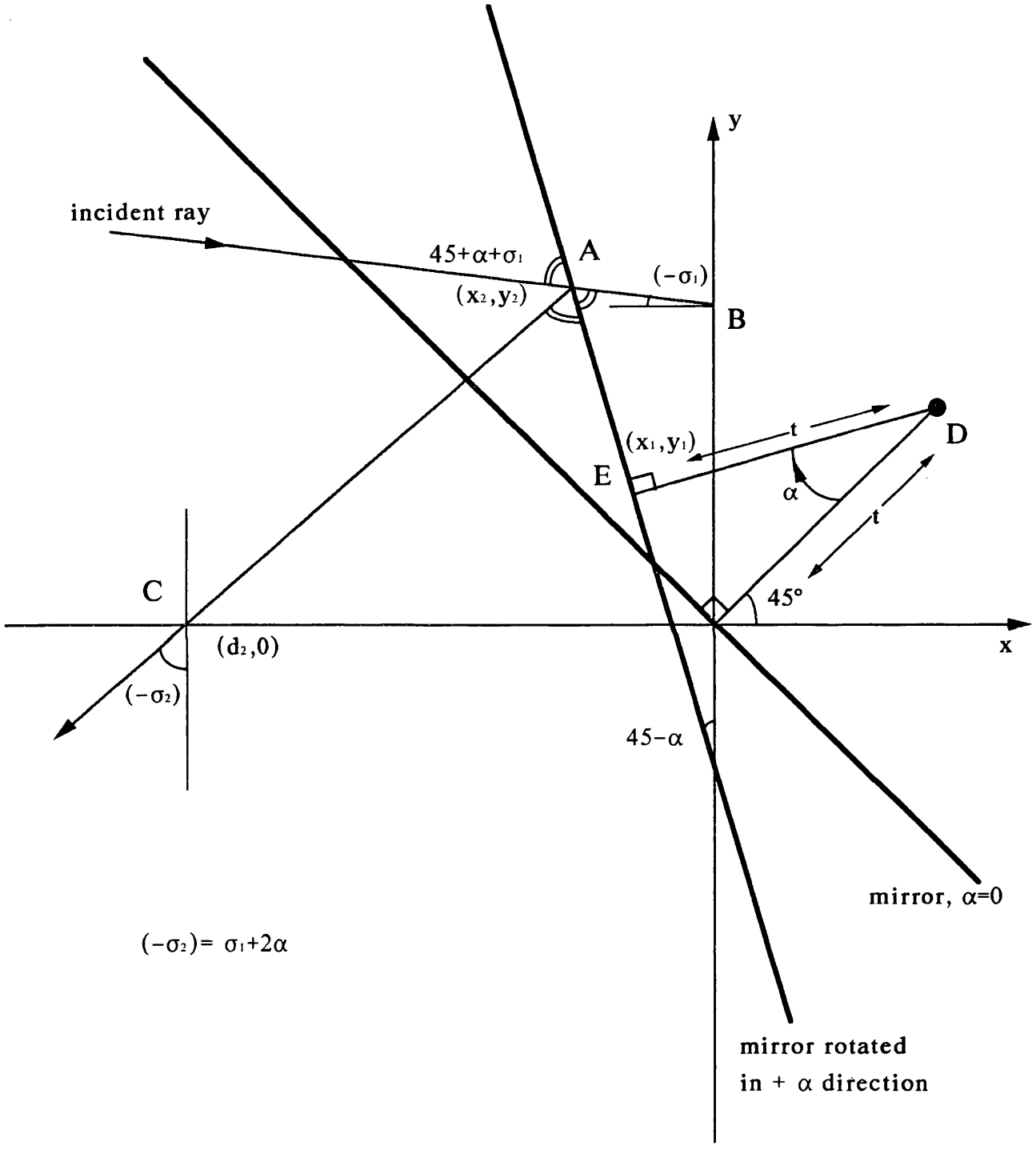


Figure E-1 Tracing a single ray at rotated mirror

The analysis finds expression for the following :

σ_2 : orientation of reflected ray ie., angle subtended with x = 0. σ_2 is positive if outgoing beam has a +x direction component ;

d_2 : displacement of reflected ray ie., intercept with x-axis at $y = 0$;

Path : extra path length imposed by mirror ie., (AC–AB).

The analysis is divided into 3 section. In the first section, the equation is found for the mirror surface (at arbitrary rotation). The remaining sections derive expression for the results as listed above.

I finding the equation of the mirror surface

Referring to figure E–1, the point of pivot of the mirror (D) is at coordinate $\left(\frac{t}{\sqrt{2}}, \frac{t}{\sqrt{2}}\right)$, a distance of t from the origin. E is on the mirror surface at a constant distance, t from D, and has coordinates, (x_1 , y_1).

Using these coordinates for D and E it follows that,

$$\left[x_1 - \frac{t}{\sqrt{2}}\right]^2 + \left[y_1 - \frac{t}{\sqrt{2}}\right]^2 = t^2$$

and expanding the brackets,

$$x_1^2 - \sqrt{2}\, t\, x_1 + y_1^2 - \sqrt{2}\, t\, y_1 = 0 \qquad \ldots(E-1)$$

The slope of the radial line from (x_1 , y_1) to $\left(\frac{t}{\sqrt{2}}, \frac{t}{\sqrt{2}}\right)$ is $\tan(45 - \alpha)$. The mirror surface is normal to this line, and hence has a slope, $-\cot(45 - \alpha)$.

We can write the equation of the mirror surface as,

$$y = a_1 x + a_2 \qquad \ldots(E-2a)$$

where

$$a_1 = -\cot(45 - \alpha) \qquad \ldots(E-2b)$$

To find a_2, we must solve equation (E – 1) with equation (E – 2) at (x_1 , y_1).

Substituting equation (E – 2) into (E – 1), and using the identity,

$$a_1 = -\frac{1}{\tan(45 - \alpha)} \equiv -\tan(45 + \alpha) \qquad \ldots(E-3)$$

we get,

$$x_1^2 - \sqrt{2}\, t x_1 + (a_1 x_1 + a_2)^2 - \sqrt{2}\, t (a_1 x_1 + a_2) = 0$$

and simplifying,

$$x_1^2 \left(1 + a_1^2\right) + x_1 \left(2a_1 a_2 - \sqrt{2}\, t - \sqrt{2}\, t\, a_1\right) + \left(a_2^2 - \sqrt{2}\, t\, a_2\right) = 0 \qquad \ldots(E-4)$$

The solution of this equation gives the x coordinate of E. However E is single valued, so both solutions are the same. It therefore follows that,

$$b^2 = 4ac \qquad \ldots(E-5a)$$

where a, b and c are the coefficients of each term,

ie.,

$$a = \left(1 + a_1^2\right) \qquad \ldots(E-5b)$$

$$b = \left(2a_1 a_2 - \sqrt{2}\, t - \sqrt{2}\, t\, a_1\right) \qquad \ldots(E-5c)$$

and

$$c = \left(a_2^2 - \sqrt{2}\, t\, a_2\right) \qquad \ldots(E-5d)$$

After some re-arrangement, and eliminating a_1 using equation (E – 2*b*), we get,

$$a_2^2 - \frac{2t \cos\alpha}{\sin(45 - \alpha)} - \frac{t^2 \sin^2\alpha}{\sin^2(45 - \alpha)} = 0$$

and factorising,

$$\left[a_2 - \frac{t\ (1 + \cos\ \alpha)}{\sin(45 - \alpha)}\right] \cdot \left[a_2 + \frac{t\ (1 - \cos\ \alpha)}{\sin(45 - \alpha)}\right] = 0 \qquad \ldots(E-6)$$

Hence $a_2 = + \frac{t\ (1 + \cos\ \alpha)}{\sin(45 - \alpha)}$ or $a_2 = - \frac{t\ (1 - \cos\ \alpha)}{\sin(45 - \alpha)}$

The constant a_2 is the intercept point of the mirror surface and y-axis. Study of the geometry of figure E–1 shows that for $t \geq 0$ and $-\frac{\pi}{4} < \alpha < +\frac{\pi}{4}$, then a_2 must always be less than or equal to zero.

Therefore the valid solution for a_2 is,

$$a_2 = - \frac{t\ (1 - \cos\ \alpha)}{\sin(45 - \alpha)} \qquad \ldots(E-7a)$$

Summarizing,

The equation of the mirror surface is

$$y = a_1 x + a_2 \qquad \ldots(E-2a)$$

where $$a_1 = -\cot(45 - \alpha) \qquad \ldots(E-2b)$$

and $$a_2 = - \frac{t\ (1 - \cos\ \alpha)}{\sin(45 - \alpha)} \qquad \ldots(E-7a)$$

The necessary conditions for this solution are,

$$t \geq 0 \quad \text{and} \quad -\frac{\pi}{4} < \alpha < +\frac{\pi}{4}$$

These conditions imply that,

i) $a_1 \leq 0$

ii) $a_2 \leq 0$.

II equation of incident and reflected rays

The equation of the incoming ray, incident on the mirror surface is,

$$y = b_1 x + d_1 \qquad \ldots(E-8a)$$

where $$b_1 = \tan \sigma_1 \qquad \ldots(E-8b)$$

The equation of the reflected ray is,

$$c_1 y = x - d_2 \qquad \ldots(E-9a)$$

where $$c_1 = -\tan \sigma_2 \qquad \ldots(E-9b)$$

and $$\sigma_2 = -(\sigma_1 + 2\alpha) \qquad \ldots(E-9c)$$

From the geometry of figure E-1, σ_1 and σ_2 must be within the limits,

$$-\frac{\pi}{2} < \sigma_1 < \frac{\pi}{2} \quad \text{and} \quad -\frac{\pi}{2} < \sigma_2 < \frac{\pi}{2}$$

III determination of the position of the reflected ray, and optical path length

From the preceding sections, three equations have been derived for,

the mirror surface : $y = a_1 x + a_2$

the incident ray : $y = b_1 x + d_1$

the reflected ray : $c_1 y = x - d_2$

where a_1, a_2, b_1, and c_1 are given by earlier expressions, and d_1 is an 'input' variable.

Solving the three equations gives,

$$x_2 = \frac{d_1 - a_2}{a_1 - b_1} \qquad \ldots(E-10a)$$

$$y_2 = \frac{a_1 d_1 - a_2 b_1}{a_1 - b_1} \qquad \ldots(E-10b)$$

and $$d_2 = x_2 - c_1 y_2 \qquad \ldots(E-10c)$$

The optical path length along AB is,

$$PathA = \sqrt{x_2^2 + (y_2 - d_1)^2} \qquad \ldots(E-11a)$$

(take positive root if $x_2 < 0$, else take negative root)

The optical path length along AC is,

$$PathB = \sqrt{(x_2 - d_2)^2 + y_2^2} \qquad \ldots(E-11b)$$

(take positive root if $y_2 > 0$, else take negative root)

Hence the extra optical path due to the mirror is,

$$Path = pathB - pathA \qquad \ldots(E-11c)$$