ADDENDUM

The factor $\cos\theta$ which is introduced in the expression for the exposure time on p 3.9 and mentioned on p 1.5 is incorrect and should be deleted.

This factor was derived assuming that the photoresist is a two dimensional recording medium. It is in fact a three dimensional medium, in which case, the exposure is independent of θ . In most cases in this work, the error introduced by this factor is negligible.

AN INVESTIGATION OF THE PRODUCTION AND PROPERTIES OF HOLOGRAPHIC DIFFRACTION GRATINGS

- by -

G. W. GRIME, B.A.

Thesis submitted for the degree of Doctor of Philosophy in the University of Oxford.

Keble College

BODLEIAN LIBRARY OXFORD
DEPOSITED
THESIS
24.7.75

February, 1975

TABLE OF CONTENTS

Abstract

Acknowledgements

Chapter 1 INTRODUCTION

1.1 Background

1.2 The production of holographic diffraction gratings

1.2.1 Forming the fringes

1.2.2 Recording the fringes

1.3 The properties of holographic diffraction gratings - comparison with ruled gratings

1.3.1 Groove profile

1.3.2 Efficiency

1.3.3 Resolving power

Chapter 2 THE FORMATION OF THE FRINGES

2.1 The effect of wavefront curvature on the interference pattern

2.1.1 Discussion

2.1.2 Detailed calculation

2.2 Light source and stability requirements

2.2.1 Light source

2.2.2 The effects of environmental changes on the interference pattern

2.2.3 Checking the stability of the system

2.3 The fringe forming system

2.3.1 Comparison of systems

2.3.2 Description of the system used

2.3.3 Alignment of the system

CHAPTER 3 RECORDING THE FRINGES

3.1 Photoresist

3.2 Photoresist Techniques

3.2.1 Coating the blanks

3.2.2 Problems encountered in coating the blanks

3.2.3 Post bake

3.2.4 Exposure

3.2.5 Development

3.3 Metallising

3.4 The influence of exposure and development on groove profile

3.4.1 The response of the photoresist

3.4.2 Groove profile

CHAPTER 4 THE EXPERIMENTAL DETERMINATION OF GRATING PROPERTIES - APPARATUS AND TECHNIQUE

4.1 The measurement of efficiency

4.1.1 Efficiency at constant wavelength

4.1.2 Efficiency in constant deviation mounts

4.1.3 Efficiency in zero order

4.2 The measurement of scattered light

4.2.1 The total flux of scattered light as a function of incidence

4.2.2 The angular distribution of scattered light

CHAPTER 5 THE EFFICIENCY OF HOLOGRAPHIC DIFFRACTION GRATINGS

5.1 Wood's Anomalies

5.2 Scalar theory

5.3 Electromagnetic theory

5.3.1 Infinite conductivity theory

5.3.2 Finite conductivity theory

CHAPTER 6 SURFACE PLASMON EFFECTS IN DIFFRACTION GRATINGS

6.1 The Surface Plasmon Resonance

6.1.1 Surface plasmons

6.1.2 Surface plasmons on periodic surfaces

6.2 Surface plasmons and efficiency

6.2.1 Efficiency at constant wavelength

6.2.2 Efficiency in Littrow mountings

6.3 Surface plasmons and scattered light

6.3.1 The total intensity of scattered light

6.3.2 The angular distribution of scattered light

CHAPTER 7 APPLICATIONS OF HOLOGRAPHIC DIFFRACTION GRATINGS

7.1 Radiation resistant gratings for use with synchrotron radiation

7.2 Gratings for tunable lasers

7.2.1 Dye lasers

7.2.2 CO_2 lasers

7.3 A wide range multichannel spectrometer using a composite HDG

7.4 Holographic zone plates

APPENDIX I

APPENDIX II

APPENDIX III

APPENDIX IV

REFERENCES

ABSTRACT

The techniques for producing high quality holographic diffraction gratings are investigated. These gratings are formed by recording the fringes of an optical interference pattern in such a way as to give grooves in a highly reflecting surface, and have been found to be superior to ruled gratings in terms of resolving power, although the sinusoidal groove profile means that the efficiency cannot be controlled easily.

In order to produce gratings of high quality, care must be taken in the design of the apparatus used to form the fringes. It is shown that the two interfering wavefronts must be identical in curvature and accurately plane ($r \gg 20m$) in order to form an accurate plane grating.

The effects of the light source and environmental changes on the quality of the pattern are investigated, and the apparatus is designed to minimise these effects. Stability tests of the fringe system show that it is satisfactory.

The fringes are recorded in positive photoresist (Shipley AZ1350), which becomes soluble on exposure to light. The techniques for preparing the blanks by spin coating and developing the exposed gratings are discussed.

The exposure characteristics of the photoresist are measured and the groove profile for various combinations of exposure and development time is calculated. It is found that the minimum distortion

from a sinusoidal groove profile results when the gratings are given a small, constant exposure to the fringes and the groove depth is changed by varying the development time. The resist surface is metallised to give a high quality reflection grating.

The efficiency of the gratings as a function of angle of incidence, wavelength and groove depth is investigated in various configurations. It is found that for coarse gratings ($\lambda/d \lesssim 0.25$) with highly reflecting surfaces, simple scalar theory is sufficient to explain the observed behaviour.

For gratings with larger values of λ/d, there are fewer propagating orders and a corresponding increase in the intensity of the high orders. This introduces Wood's anomalies, which are rapid variations of efficiency occurring when one of the orders passes off the grating. These are polarisation dependant and cannot be explained on simple theories.

The basic properties of rigorous theories which assume a perfect metal are considered. These are found to be inaccurate in the region of the S polarised (E perpendicular to the grooves) anomaly.

The recent theories which take account of the optical properties of the metal are shown to predict accurately the observed variation of efficiency even in the region of the anomalies. The dependance of the grating properties on the metal coating has been shown to be due to an interaction between the radiation field and the electrons of the metal coating. This surface plasmon resonance is investigated in detail. The phenomena caused by the resonance are found to be:

i) A net absorption of energy by the grating;

ii) Redistribution of energy between the orders of the grating;

iii) An increase in the scattered light from the grating;

iv) A well defined angular distribution of scattered light.

The phenomena only occur with S-polarised incident light.
The conditions under which the resonance occurs are investigated and the effects are observed. The dependence of the resonance on groove depth is also investigated. It is shown how observations of the surface plasmon resonance may provide information concerning the properties of electron waves on periodic surfaces.

Finally, some applications of the holographic gratings made during this work are discussed. These are:

i) Radiation resistant gratings for use in synchrotron radiation;

ii) Gratings for tunable lasers;

iii) A multichannel spectrometer using a composite grating;

iv) Holographic zone plates as imaging elements.

ACKNOWLEDGEMENTS

I wish to acknowledge the advice and encouragement of my supervisor, Dr.A.Corney, during the course of this work. I should also like to thank Mr.C.W.Goodwin, who produced all the metal coatings, Mr.C.W.Band of the Clarendon Laboratory photographic department for many helpful discussions on the problems of high resolution photography and Mr.H.S.Lock for technical advice and assistance.

Rank Precision Industries Ltd., Margate, provided some of the optical components used in the project, which was supported by the Science Research Council under the C.A.P.S. Scheme.

Finally I wish to thank Toni Tattersall for typing the manuscript and my wife for her patience.

Chapter 1

INTRODUCTION

1.1 Background

1.2 The production of holographic diffraction gratings

1.2.1 Forming the fringes

1.2.2 Recording the fringes

1.3 The properties of holographic diffraction gratings - comparison with ruled gratings

1.3.1 Groove profile

1.3.2 Efficiency

1.3.3 Resolving power

In recent years, it has become possible to produce high quality diffraction gratings by photographically recording the fringes of an optical interference pattern. Apart from the convenience of production, these holographic diffraction gratings (HDG's) can show substantial improvements in performance over ruled gratings, and it seems likely that for certain applications they will replace ruled gratings as a dispersing element.

The work presented in this thesis is in two sections. The first part (Chapters 2 and 3) is an investigation of the techniques of producing HDG's of high quality.

In the second part (Chapters 5 and 6) a study is made of the efficiency of HDG's under various conditions. This leads to an investigation of the processes taking place at the surface of the grating. An interaction is found between the incident light and the conduction electrons in the metal.

Chapter 4 describes the experimental techniques of measuring the performance of diffraction gratings and Chapter 7 outlines some of the applications of HDG's.

In this chapter we summarize the background to this work and outline the principles of the production of HDG's. In the final section the basic properties of HDG's are considered.

1.1 Background

The technique of producing gratings by recording an optical interference pattern was first developed by Burch and Palmer in 1958 (1,2,3). Using thermal light sources and silver halide emulsions, the gratings they produced were of poor quality in terms of spectroscopic performance but as they were designed for metrological use, this was unimportant. With the development of the laser and the renewed interest in holography, the possibility of making spectroscopic gratings by this method became apparent. In 1966, George and Matthews (4) produced "holographic" diffraction gratings by recording fringes produced by a laser in silver halide emulsion, and despite some efforts to adopt the more descriptive term "interference gratings" this nomenclature has become accepted.

The use of silver emulsion as a recording medium meant that the gratings were still of poor optical quality. The efficiency was low and because of the grain structure, there was a high level of scattered light. The technique was improved by the use of grainless recording media. In 1968, Shankoff and Curran (5) used dichromated gelatine to record high resolution gratings.

In the same year, Rudolf and Schmahl (6) produced reflection gratings by metallising HDG's recorded in photoresist. In 1970 (7) it was shown that the performance of HDG's recorded in photoresist could compare favourably with that of ruled gratings. In particular, the stray light properties were superior. Recent research into HDG's has fallen into two categories.

One field concerns the production of gratings with non-parallel grooves, formed by the interference of two point sources. These gratings are able to improve the performance of concave grating

monochromators by correcting some of the large geometrical oberrations present in classical mountings. The positioning of the point sources to optimise the performance of a particular monochromator is a complicated task requiring the use of computers. This work is being carried out mainly by Pouey (8) and Namioka et al (9).

The second category of research concerns the dependance of the efficiency of the grating on its parameters. A large amount of theoretical work has recently been carried out by Petit et al (31) on the relationship between groove profile and efficiency. HDG's, which have a sinusoidal profile and can be conveniently produced with any groove spacing and groove depth, are used to verify the predictions of this theory (10).

HDG's are also convenient for the investigation of the surface plasmon resonance (11, Ch.6 of this thesis). In addition to the production of HDG's with sinusoidal grooves, the possibility of producing HDG's with triangular groove profiles, thereby creating blaze, is also being investigated (12,18).

1.2 The production of holographic diffraction gratings

The parallel fringes used in the formation of plane HDG's are produced by the interference of two plane wavefronts (figure 1.1).

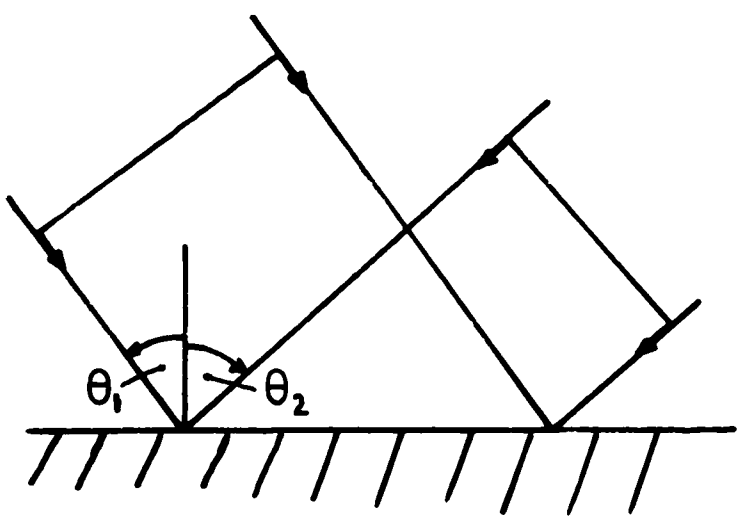


Fig. 1.1

Assuming that the wavefronts are coherent and of equal intensity over the area of the blank, an interference pattern is formed in the plane of the blank whose intensity is given by

$$I(x) = I_{max} \cos^2 \frac{\pi x}{d}$$

with $$d = \frac{\lambda}{\sin \theta_1 + \sin \theta_2}$$

where λ is the wavelength of the light.

Here d is the groove spacing of the HDG. It is seen that the minimum occurs when $\theta_1 = \theta_2 = 90^\circ$, when $d = \lambda/2$.

Since the normal intensity on the blank is proportional to $\cos\theta$, however, a practical lower limit is set by the speed of the recording medium; if long exposures are necessary, instabilities may destroy the fringe pattern.

In order to produce wavefronts which are coherent over a large area and which are sufficiently intense to keep the exposure times short, a high power laser must be used as a light source.

In addition to ensuring that the fringe pattern remains stable during the exposure, precautions must be taken to produce fringes which are straight and parallel to the required degree of accuracy.

The design of the fringe forming system is discussed in detail in chapter 2.

1.2.2 Recording the fringes

The medium used to record the fringes is positive working photoresist (Shipley AZ1350). This consists of a resin, which when placed in developer is etched at a rate dependent on the exposure received. Thus when the blank is exposed to a fringe pattern and

subsequently developed, grooves are formed in the surface of the resist. The blank is then metallised to produce a reflection grating.

The films of resist are typically 1 μm thick and are produced by coating an optical flat with a thin layer of resist solution and allowing the solvent to evaporate. This is done by spin coating.

The resist characteristic is approximately exponential, so to avoid the distortion which would result if the exposure was increased to increase the groove depth, all the gratings are given a constant exposure to the fringes and the groove depth is changed by varying the development time. This is shown to reduce the distortion to a low level.

The techniques of recording the fringes are fully discussed in chapter 3.

1.3 The properties of holographic gratings - a comparison with ruled gratings

1.3.1 Groove profile

The groove profile of a diffraction grating is difficult to observe directly. Electron microscope techniques are destructive, time consuming and difficult to interpret, and the most convenient method seems to be the use of a modified "Talystep" machine (19).

In this work, the groove profile of the HDG's is assumed to be accurately sinusoidal, with a groove depth linearly proportional to the development time. This assumption is supported by the close agreement between the observed efficiencies and theoretical calculations for sinusoidal profiles.

The fundamental difference between HDG's and ruled gratings is the groove profile. Ruled gratings are produced with carefully

controlled triangular profiles designed to concentrate most of the incident light into a particular order.

1.3.2 Efficiency

The efficiency of a grating (the fraction of the incident energy which is diffracted into a given order) is determined largely by the groove profile. The carefully controlled blaze of a ruled grating ensures that gratings are available with high efficiencies for all wavelengths, but the symmetrical profile of a HDG means that the efficiency is relatively low, especially when many orders are propagating.

High efficiencies may be obtained with HDG's if they are operated in the region where only the order of observation and the zero order are propagating, for now the energy is distributed between only two orders. Efficiencies of up to 90% have been obtained in this way.

The disadvantage of using a grating in the single order region is that the dispersion is very high, and this may be unsuitable for some applications.

A prominent feature of the efficiency variation of HDG's are the Wood's anomalies. These rapid variations of efficiency occur when one of the orders is diffracted at 90° and passes off the grating. They are particularly prominent in HDG's because they are not masked by blaze effects as in ruled gratings. The intensity of the anomalies in HDG'sis convenient from a theoretical standpoint, as a study of the anomalies can yield interesting information concerning the diffraction process, but may be a disadvantage to the user. Associated with the anomalies there is found to be an interaction between the incident radiation and surface plasmons in the grating metal.

As described in chapter 6, this causes several interesting phenomena, including a large drop in efficiency and a marked increase in the scattered light from the grating.

This interaction implies that the efficiency of the grating can only be predicted using theories which take into account the optical properties of the metal. This is supported by the close agreement between observed efficiencies and finite conductivity theories.

1.3.3 Resolving power

The resolving power of a spectrometer system is determined largely by the optical quality of the grating. Errors in groove position and imperfections on the grating surface will tend to spread the image of the spectral line and so degrade resolution.

In this respect, HDG's are superior to ruled gratings because all the grooves are formed simultaneously and so the problems of controlling the position of the groove over a period of several days, as in a ruling engine, is avoided.

In addition the surface of the groove is smoother and so the general level of scattered light is reduced.

The low level of general scattered light from a HDG facilitates the observation of the scattered light from surface plasmons. With ruled gratings, this can be masked by scattering from the grating imperfections.

Chapter 2

THE FORMATION OF THE FRINGES

2.1 The effect of wavefront curvature on the interference pattern

2.1.1 Discussion

2.1.2 Detailed Calculation

2.2 Light source and stability requirements

2.2.1 Light source

2.2.2 The effects of environmental changes on the interference pattern

2.2.3 Checking the stability of the system

2.3 The fringe forming system

2.3.1 Comparison of systems

2.3.2 Description of the system used

2.3.3 Alignment of the system.

One of the advantages of HDG's over ruled gratings is that the groove accuracy can be made very high with relatively little effort compared with that required to control a ruling engine during the period of ruling, which may be several days. In particular, periodic errors, which are caused by errors in rotating components, can be eliminated.

In order to achieve this accuracy, the recording wavefronts must satisfy certain conditions, which in turn will influence the design of the system used to form the fringes. In addition, the effects of instabilities during the exposure must be considered.

In this chapter, the effects on the fringe pattern of wavefront curvature are considered. This is followed by a discussion of the effects of system instabilities on the fringe system and the final section is a description of the fringe forming system used here.

2.1 The effect of recording wavefront curvature on the interference pattern

2.1.1 Qualitative Discussion

An ideal plane grating has grooves which are parallel and equally spaced over the full extent of the grating. In order to make such a grating holographically it is obvious that perfectly plane wavefronts must be used.

In practice, however, the recording wavefronts will have a slight curvature due, for example, to diffraction at the apertures of the system, or errors in adjusting the collimator.

The effect of wavefront curvature may be investigated by assuming that the interference pattern is due to two coherent point sources A and B. The condition for constructive interference at any point P is that

$$[AP] - [BP] = n\lambda$$

where $[\ ]$ denotes optical path

n is an integer

λ is the wavelength of the sources

Thus the resulting interference pattern in space is a family of hyperboloids of rotation with foci A and B and axis AB.

The two dimensional interference pattern recorded on the blank is the intersection of the hyperboloids with the plane of the blank, and so several arrangements are possible (figure 2.1)

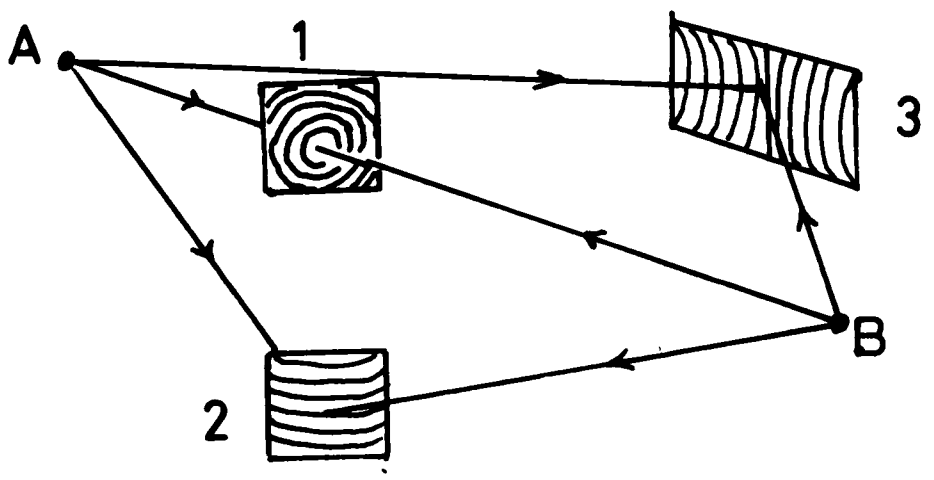


Fig. 2.1

Position 1: The centre of the blank lies on AB. A series of concentric ellipses are formed, which become circles when the blank is normal to AB. Such a grating will act as a zone plate.

Position 2: This is similar to 1, but now only arcs of ellipses are recorded.

Position 3: The recording plane is parallel to AB and intersects the zero order fringe. In this case, the fringe pattern is a set of hyperbolae which are symmetrical about the straight zero order fringe.

2.1.2 Detailed calculation

In order to quantify the pattern, consider figure 2.2

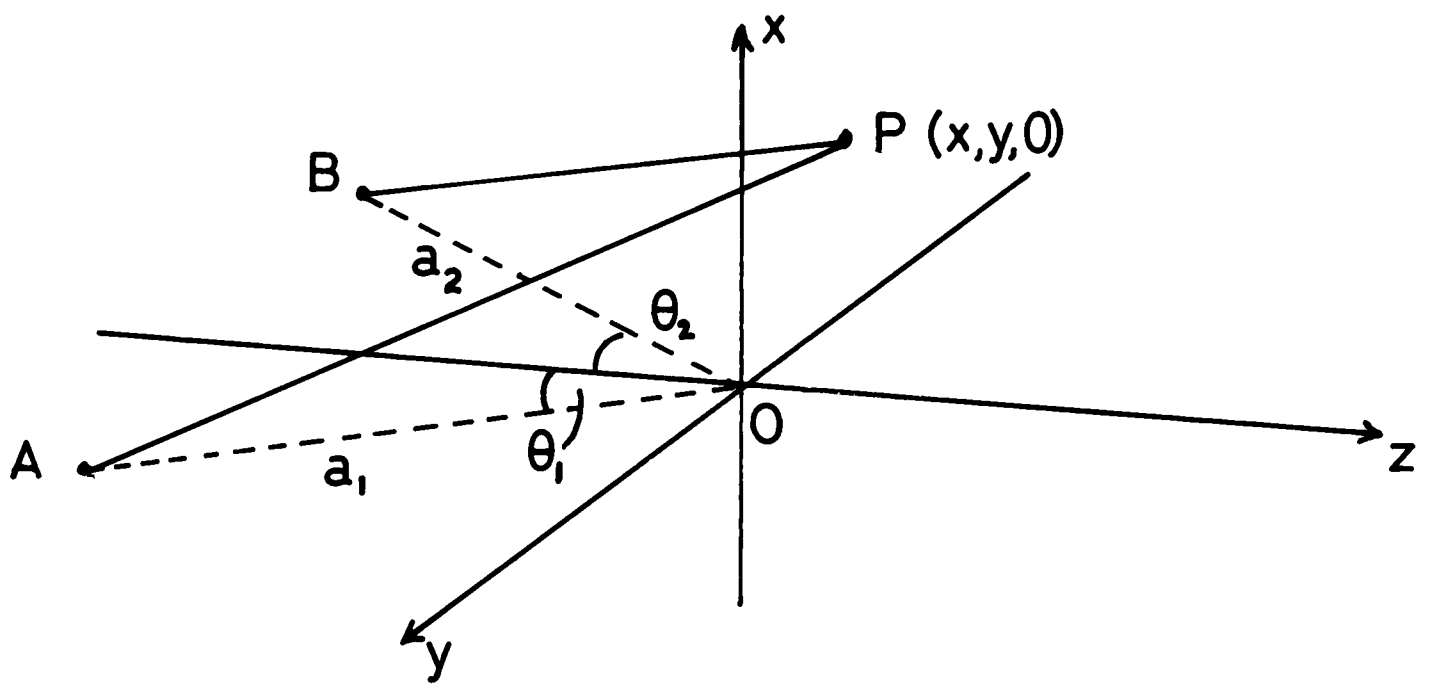


Fig. 2.2

The recording plane is the xy plane and the sources A and B lie in the yz plane. If it is assumed that there is a bright fringe at the origin, then P will also lie on a bright fringe if the path difference $[AP]-[BP]$ increases by $n\lambda$ as P goes from O to P. Thus:

$$([AP]-[BP]) - ([AO] - [BO]) = n\lambda$$

or

$$\left[x^2 + (y - a_1 \sin\theta_1)^2 + a^2_1 \cos^2\theta_1\right]^{\frac{1}{2}}$$

$$-\left[x^2 + (y-a_2 \sin\theta_2)^2 + a^2_2 \cos^2\theta_2\right]^{\frac{1}{2}}$$

$$- (a_1 - a_2) = n\lambda$$

Thus at any point (x,y) on the recording plane, the groove number, n, relative to the centre of the grating may be calculated.

A perfect grating would consist of a family of lines parallel to the x axis given by

$$y = md$$

where d is the groove spacing

m is an integer

Thus if the actual fringe number at (x,y) is n, the fringe displacement may be written

$$\Delta(x,y) = nd - y$$

This is the difference between the position of the fringe and its ideal position. This may be written as a fractional fringe error

$$\epsilon(x,y) = \frac{\Delta(x,y)}{d} = n - \frac{y}{d}$$

For a perfect grating, $\epsilon(x,y)$ should be zero for all values of x and y, but since in practice, the ancillary optics of spectroscopic systems rarely have wavefront errors less than $\lambda/10$, a realistic limit would be $\epsilon(x,y) < 0.1$ over the area of the grating.

The value of ϵ was calculated for the following set of conditions:

wavelength:	457.9 nm
grating period:	763 nm
a_1 :	-5×10^5 cm
a_2 :	various values between 5×10^6 cm and -4.5×10^5 cm

The beams were incident symmetrically on the blank (i.e $\theta_1 = -\theta_2 = 17.46^\circ$) The value chosen for a_1 is approximately the radius of curvature produced by diffraction at an aperture of 5 cm diameter.

The results are summarised in figure 2.3 for x = 2.5 cm and y = 0 to 2.5 cm

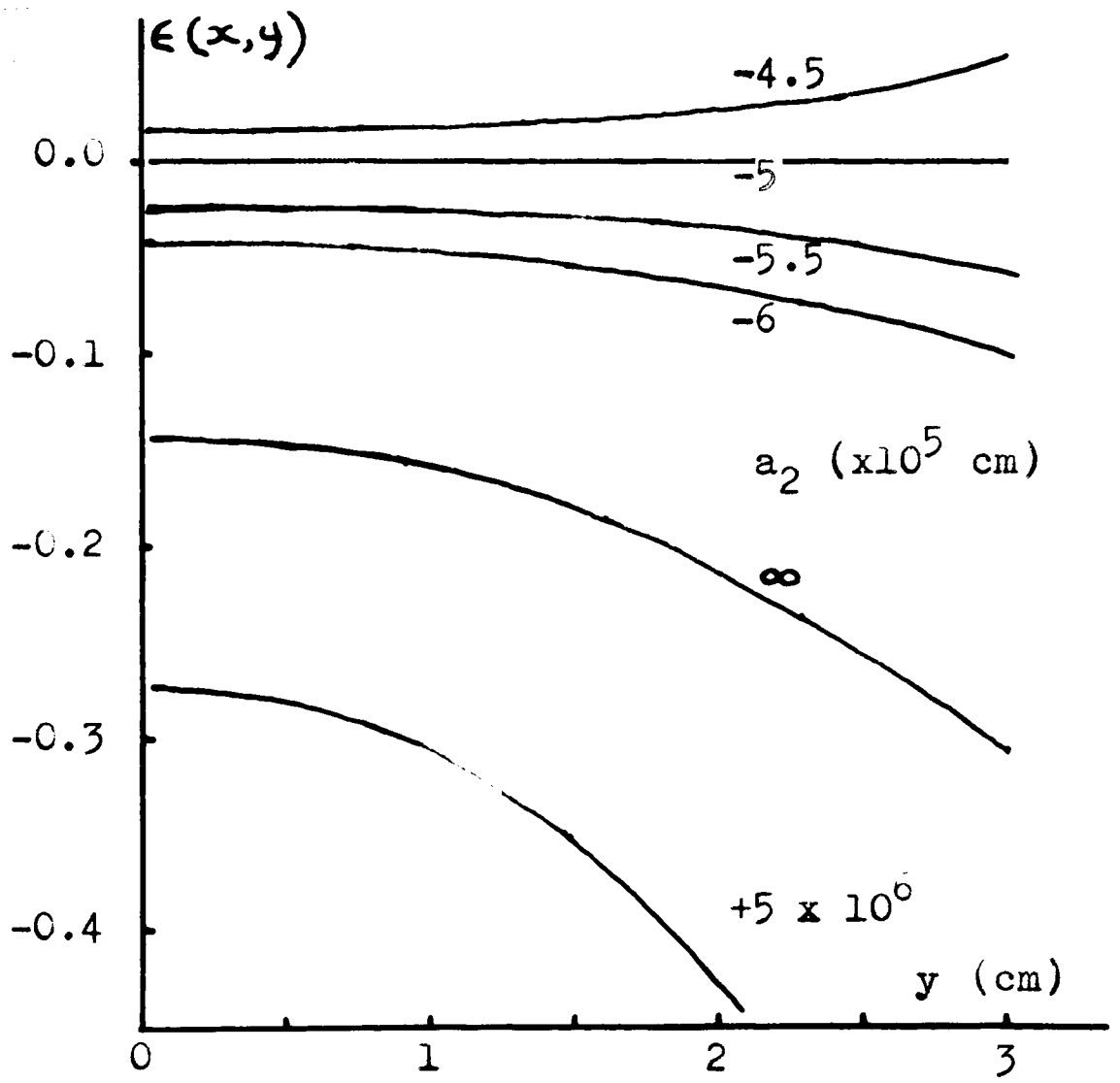


Fig. 2.3

The curves demonstrate that ϵ is zero when $a_2 = a_1$, but increases rapidly as the difference between them increases. This arrangement with $a_2 = a_1$ and $\theta_2 = -\theta_1$ is the symmetrical arrangement (position 3) of figure 2.1.

If the restriction $\epsilon < 0.1$ is taken, then acceptable gratings may be made with a_2 between -7×10^5cm and -3×10^5cm. If it is assumed that this curvature is due to the misalignment of the focus of a collimator, then it can be shown that for a 50 cm focal length lens, the condition requires the focus to be aligned to an accuracy of 0.1 mm.

If the value of ϵ is calculated with $a_1 = a_2 = a$ for several values of a, it is found that even in the symmetrical case, the fringe error is not zero, but has a small value. If the condition $\epsilon < 0.1$ is again taken, then it may be shown that a must be greater that 2×10^3 cm. This is easily achieved.

So far, the effect of local aberrations in the beam forming optics has been neglected. The effect of these will be to cause local variations in the curvature of the recording wavefronts, with a consequent distortion of the fringe pattern. Providing the aberrations are small, their effect may be minimised if the recording wavefronts are identical and are overlapped so that the distorted areas coincide on the blank.

2.2 Light Source and Stability Requirements

In the treatment of the preceding section, it is assumed that the sources A and B are coherent point sources with a single stable frequency and that the optical path length between the sources and the recording plane remains constant during the exposure.

In practice, however, the source will have a finite spread of frequencies, causing the visibility of the fringes to vary across the blank, and instabilities may cause the fringe system to move during the exposure.

These effects must be studied in detail so that measures may be taken to minimise them.

2.2.1 Light source

The only possible source for this application is a laser. This is because the photoresist used as recording medium is relatively insensitive compared with photographic emulsion, and the laser is the only source with sufficient power in a small bandwidth to keep the exposure time within reasonable limits. The choice of laser is determined largely by the spectral response of the photoresist. This requires that the laser should operate in the deep blue or near u.v. region of the spectrum.

The laser chosen for this work was a 4 watt c.w. argon ion laser (Coherent Radiation Inc. Type 52B). The transition at 457.9 nm was used. The power available at this wavelength was 150 mW in a single longitudinal mode.

2.2.1(a) Longitudinal Mode Structure

The spectrum of a gas laser consists of a number of longitudinal cavity modes within the gain profile of the transition which is oscillating. The separation of the modes is C/2L Hz where L is the length of the cavity. Thus the coherence length of the output depends on the number of modes oscillating.

It may be shown (13) that the visibility of the interference fringes, defined as

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$$

where I_{max} = peak intensity in pattern

I_{min} = minimum intensity in pattern

is given by

$$V = |\mu_T(\delta)|$$

where $\mu_T(\delta)$ is the degree of temporal coherence of the light as a function of path difference, δ. (The two beams are assumed to be equal in intensity and polarised normal to the plane of the beams). For a laser whose spectrum consists of N modes of equal intensity with a cavity length L, the temporal coherence may be written (14):

$$|\mu_T(\delta)| = \left| \frac{\sin (N\pi\delta/2L)}{N \sin (\pi\delta/2L)} \right|$$

and so we have the same expression forthe visibility as a function of δ

The variation of V with N and δ is shown in figure 2.4, assuming a cavity length of 100 cm.

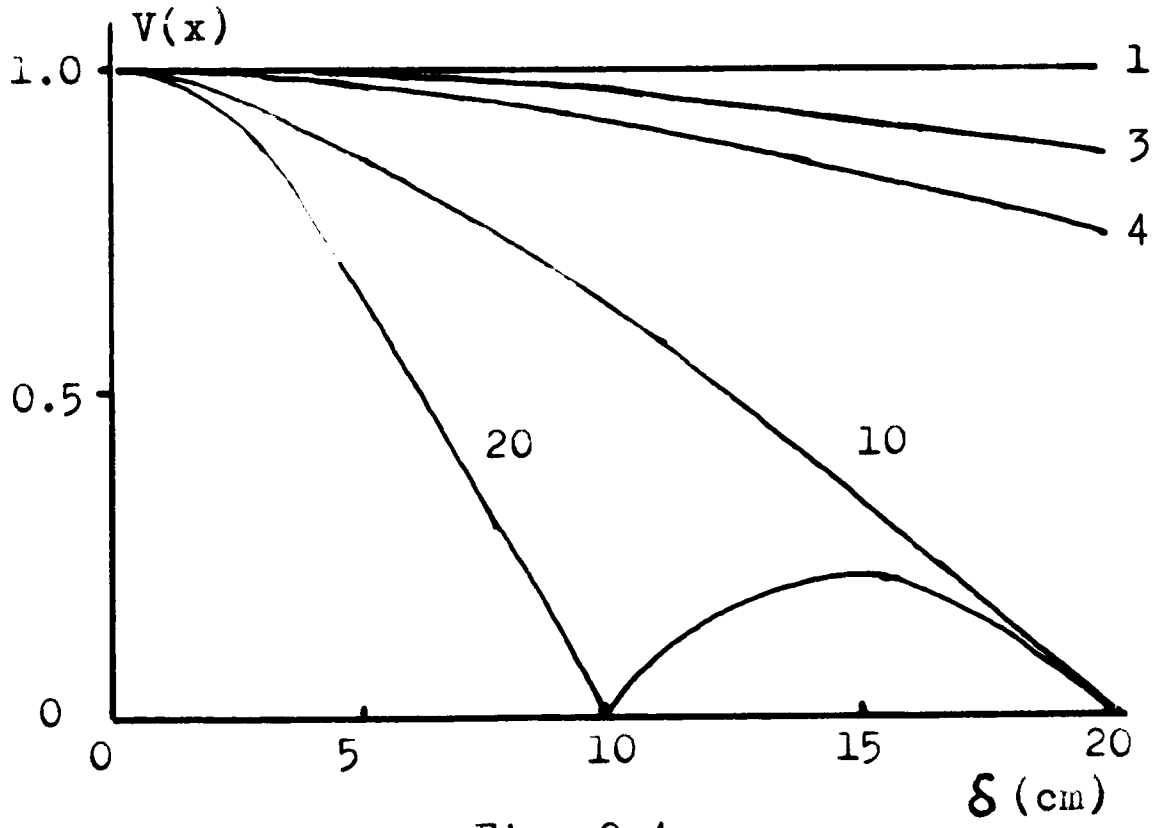


Fig. 2.4

The typical value of N for an argon ion laser is 20. This is due to the fact that the laser transition has a wide profile because of the high temperature of the discharge and the applied magnetic field. Hence it is seen from figure 2.4 that unless the number of modes is reduced, care will have to be taken to ensure equality of path in the interferometer if maximum visibility is to be achieved. In addition, the visibility may vary across the grating. The variation in optical path difference across the grating is given by

$$\Delta\delta = WG\lambda$$

where W is the width of the grating (mm)

and G is the spatial frequency (lines/mm)

For a 5 cm grating of 2400 lines/mm, the variation in path difference at 458 nm is 5.5 cm, and so if a large number of modes are oscillating, the visibility of the fringes may vary significantly over the area of the blank. This will lead to a grating with non-uniform groove depth. To avoid these effects, the number of oscillating modes must be reduced to one or two.

In the laser used in this experiment this is done by placing a solid quartz etalon in the cavity. The thickness of the etalon is controlled by means of a tilt adjustment and by thermal expansion until its passband coincides with a cavity mode at the peak of the gain profile.

In this way the laser can be made to oscillate on a single longitudinal mode.

2.2.1(b) Transverse mode structure

In order to give a uniform intensity over the area of the blank, the laser must operate in the lowest order (TEM_{oo}) transverse mode.

In this mode, the intensity by variation across the beam is given by (15):

$$I(r) = I(0) \exp(-r^2/w^2)$$

where r = distance from beam axis

W = distance at which intensity falls to $I(o)/e^2$

$I(0)$ = intensity on beam axis.

In the laser used here, w is 0.7 mm. Higher order modes could be avoided by means of an adjustable diaphragm concentric with the axis of the cavity.

2.2.1(c) Frequency stability requirements

If the beams are incident symmetrically on the blank at an angle θ with the normal, the fringe spacing is

$$d = \frac{\lambda}{2 \sin\theta}$$

Hence if the frequency of the laser varies by an amount $\Delta\nu$ it can be shown that the groove spacing will change by an amount Δd, where

$$\frac{\Delta d}{d} = -\frac{\lambda}{c} \cdot \Delta\nu$$

Assuming that the zero path difference position lies at the centre of the blank, a groove at the edge will shift by a distance

$$\Delta x = \Delta d \cdot \frac{W}{2d}$$

where W is the width of the blank. If Δx is specified, then we have

$$\Delta\nu \leq \frac{\Delta x}{d} \cdot \frac{2d.c}{W\lambda}$$

If $\Delta x/d < 0.1$, $d = 1\mu m$, $\lambda = 457.9$ nm and $W = 5$ cm

then $\Delta\nu \leq 2600$ MHz

The specified frequency stability of the laser is $\pm$ 75MHz long term, thus the laser is satisfactory in this respect.

2.2.2 The effects of environmental changes on the stability of the fringe pattern

In addition to the factors already discussed, any real fringe forming system will be affected by changes in the ambient conditions. The effect of these will be to cause the pattern to move relative to the blank, either as a vibration about the mean position or as a slow drift. If this occurs during the exposure, the quality of the recorded grating will be adversely affected.

The causes of instability which must be considered are:

1) Mechanical vibration and creep of components
2) Thermal expansion of the system.
3) Refractive index changes caused by fluctuations of ambient temperature and atmospheric pressure.

2.2.2(a) Mechanical Instability

Relative motion of the optical components stems from two causes, room vibrations transmitted to the optical system and creep of the mounts due to the relaxing of the strains induced during the setting up procedure.

The effect of vibrations is greatly reduced by setting up the system on a massive, stiff base. The mass of the table reduces the displacement caused by the vibration, while the stiffness prevents relative motion of different parts of the system.

In addition, the optical components should be held in such a way that they cannot move relative to the base Should further vibration protection be required, the table must be isolated from the building by means of resilient suspensions such as air bearings.
This was found to be unnecessary for the system used here.

The base used for the system is a steel surface plate of dimensions 5' x 3' x 15" and mass 560 kg. Flexion of the base is avoided by means of bracing struts. The laser head is mounted on a separate table to prevent the transmission of any vibration it may cause and the laser power supply and autotransformer are mounted on rubber blocks to reduce the level of room vibration.

The mirror mounts are described in detail in the next section. They are made from aluminium alloy, which was found to be prone to creep. Creep is a serious form of instability which causes a steady drift of the fringes, and the only way of avoiding it is to allow the system to settle after setting up. This can take up to seventy two hours.

2.2.2(b) Thermal expansion

The temperature change required to produce an expansion of $\lambda/10$ in a 1m length of steel is of the order of $5 \times 10^{-3}{}^{o}C$. This indicates that the system may be very sensitive to temperature fluctuations.

Fortunately the large thermal capacity of the table (approximately 5×10^{5} J/°C) ensures that it responds slowly to ambient temperature changes. As an additional precaution, the temperature of the room is held within $0.1^{o}C$. Under these conditions exposures up to 15 minutes may be made, and no further precautions were taken.

It is important that the blank should be in thermal equilibrium with the system when the exposure is made, otherwise thermal stresses may cause the blank to distort and move under the fringe pattern. This is particularly important for large blanks which may have been removed from an oven shortly before exposure.

2.2.2(c) Optical path changes caused by variations in the refractive index of air

The refractive index of air changes slightly with temperature and pressure (16). Over a small enough range, this variation may be considered to be linear, and the respective gradients calculated from the information given in this reference. For dry air at 20°C and 760 mm Hg, these are

$$\frac{dn}{dp} = 3.78 \times 10^{-7} \text{ per mm Hg}$$

$$\frac{dn}{dT} = -9.6 \times 10^{-7} \text{ per } ^{\circ}C$$

The difference in optical path to a point on the grating may be written

$$d = n_1 l_1 - n_2 l_2$$

where $n_{1,2}$ is the mean refractive index

$l_{1,2}$ is the physical length of the respective paths.

Thus if the ambient conditions change the refractive indices by δ_1 and δ_2, the change in o.p.d. will be

$$\Delta = l_1\delta_1 - l_2\delta_2$$

To keep the fringe error less than 0.1, Δ must be less than $\lambda/10$

i.e. $$l_1\delta_1 - l_2\delta_2 < \lambda/_{10}$$

In the best case, the effects are equal in each arm, so that $\delta_1 = \delta_2 = \delta$

Hence

$$\delta < \frac{\lambda}{(l_1 - l_2) \times 10} \qquad \ldots\ldots\ldots \quad (i)$$

Taking a typical value of 10 cm for $(l_1 - l_2)$ and the above values of $^{dn}/_{dT}$ and $^{dn}/_{dp}$. This shows that

$$\Delta p < 0.12 \text{ mmHg}$$

$$\Delta T < 0.4^{\circ}C$$

if temperature and pressure changes occur independantly. If, however, the disturbance acts only on one beam (i.e. $\delta_2 = 0$) then the condition for stability is that

$$\delta_1 < \frac{\lambda}{10 l_1}$$

Since l_1 may be of the order of 1m, the effect is much greater.

This discussion illustrates the importance not only of having the path lengths equal (if $l_1 = l_2$ in equation (i) then Δ is independant of δ) but also of keeping the beams physically close together so that local fluctuations, in particular pressure variations due to draughts may act equally on each beam. Draughts are reduced by enclosing the system with polythene sheet and allowing the air within the enclosure to settle before an exposure is made.

The enclosure serves the additional purpose of protecting the optical components from dust and is flushed with clean air from a filter unit when the system is not in use.

2.2.3 Checking the overall stability of the system

The following method was used to test the stability of the system. The optical components were set up as a Michelson interferometer. This was arranged so that the arms were of different lengths and one arm contained the full length of the table. The interferometer was adjusted to give circular fringes and a detector masked by a pinhole was used to monitor the intensity of the central spot.

If the optical path difference between the two arms varies, the intensity will fluctuate periodically, each period corresponding to a change of $\lambda/2$. The intensity was monitored over a period of thirty minutes on three occasions after setting up. The results are shown in figure 2.5.

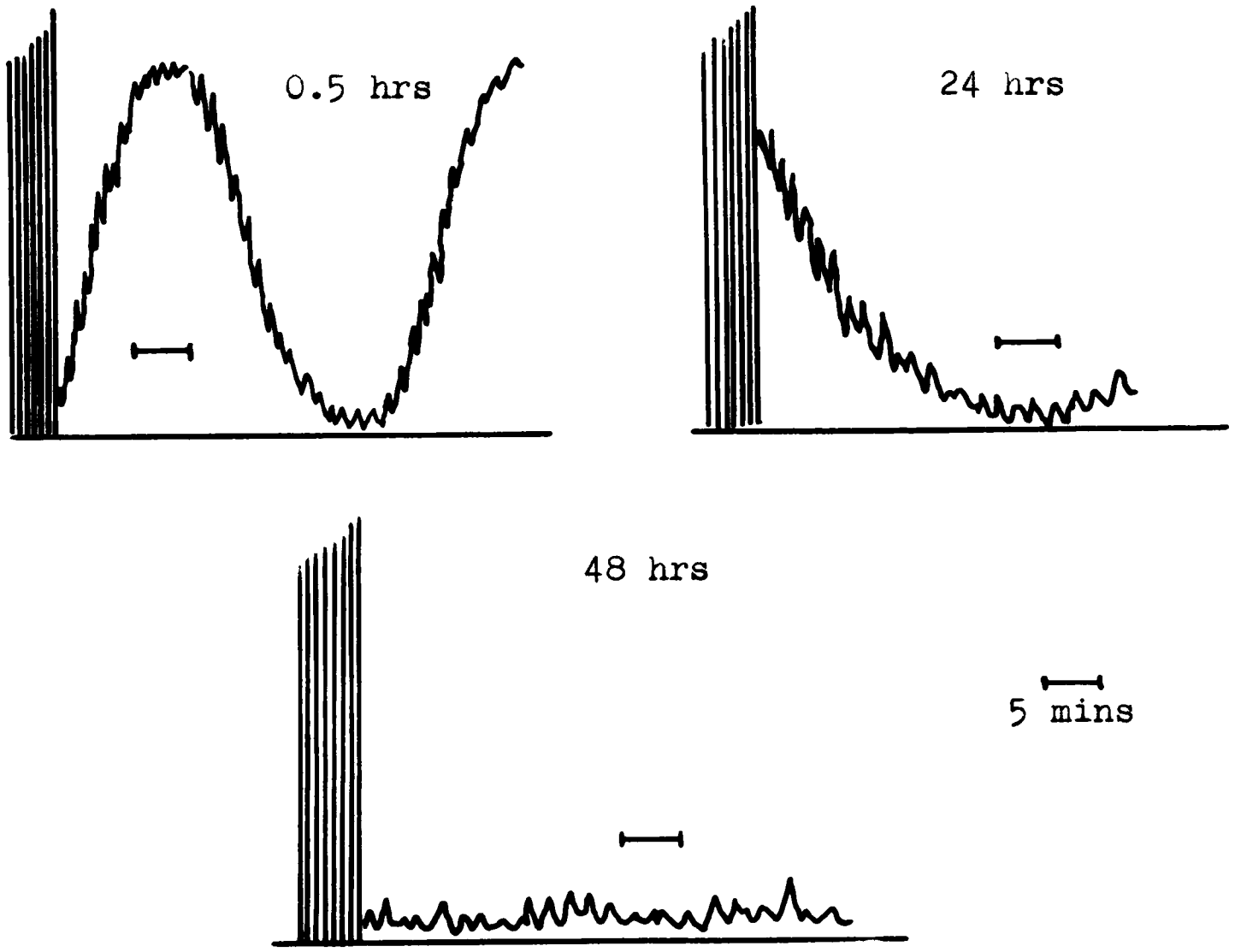


Fig. 2.5

It is seen that after a rapid fluctuation caused by frequency jumps as the etalon and cavity of the laser reach thermal equilibrium, the pattern shows a slow drift of path difference at a rate which decreases with time. This is assumed to be due to creep and after 48 hours has fallen to an acceptable rate.

2.3 The Fringe Forming System

2.3.1 A comparison of fringe forming systems

The requirements of the fringe forming system may be summarized as follows:

i) The recording wavefronts must be identical and as nearly plane as possible.

ii) The path lengths of the two beams must be similar.

iii) The system should be stable over periods of several minutes.

iv) The intensity of the wavefronts should be uniform over the area of the blank. Periodic intensity variations caused by spurious interference fringes are particularly undesirable as they cause a period modulation of the groove depth of the grating, giving rise to ghosts of spectral lines, (These are similar to the Rowland ghosts of a ruled grating).

v) The intensity of the beam should be as high as possible to keep the exposure time short.

The simplest method of achieving identical wavefronts is to produce a single wavefront and by means of a semi-reflecting beam splitter divide this into two wavefronts which differ only by the error introduced by the beam splitter.

The disadvantages of this method are:

i) A metal coated beam splitter may absorb a large fraction of the incident energy;

ii) The reflections from the uncoated surface of the beam splitter may interfere with the main beam and give rise to spurious fringes;

iii) Beam splitters of high quality may not be available with the required dimensions.

A system which avoids the disadvantages of a beam splitter is shown in figure 2.6.

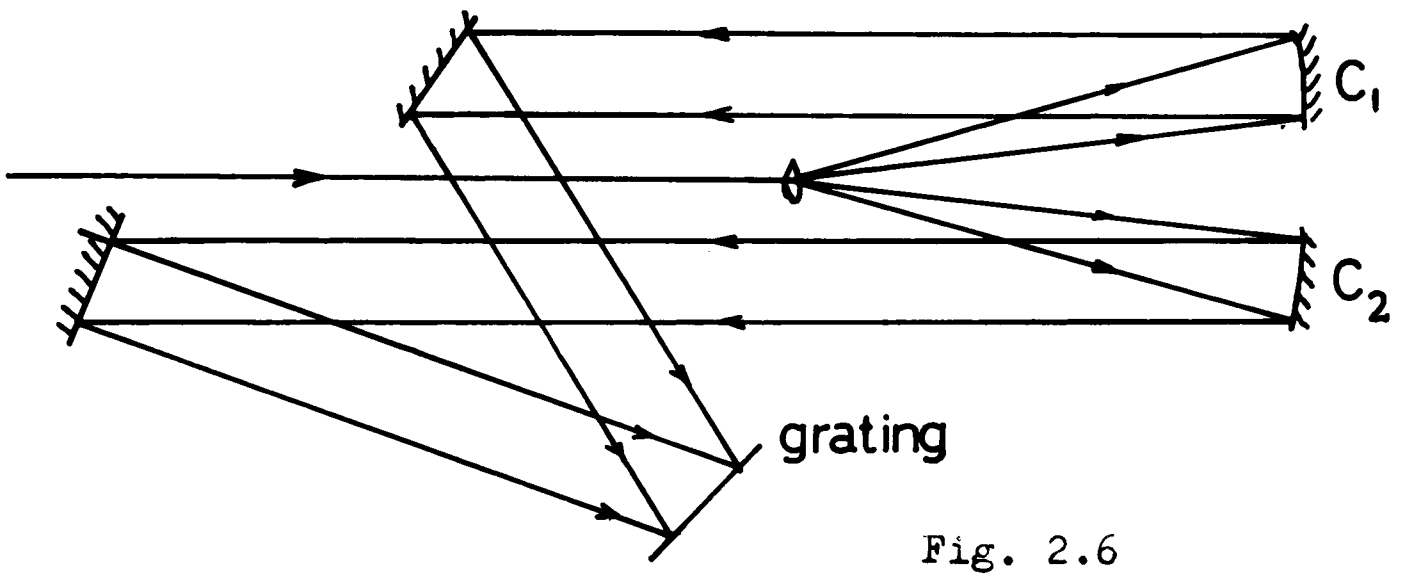


Fig. 2.6

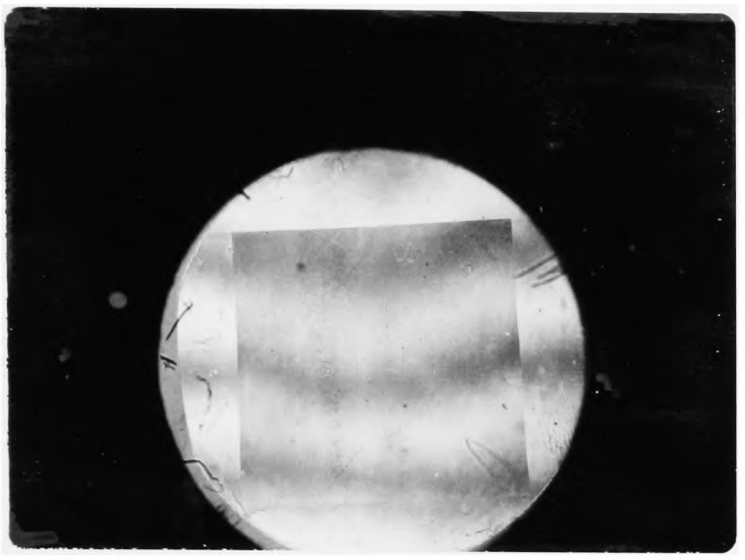
This system has a single beam expander and two collimating mirrors (C1 and C2) to form the recording wavefronts. The advantage of this system are:

i) There is no beam splitter, so that no spurious fringes are formed;

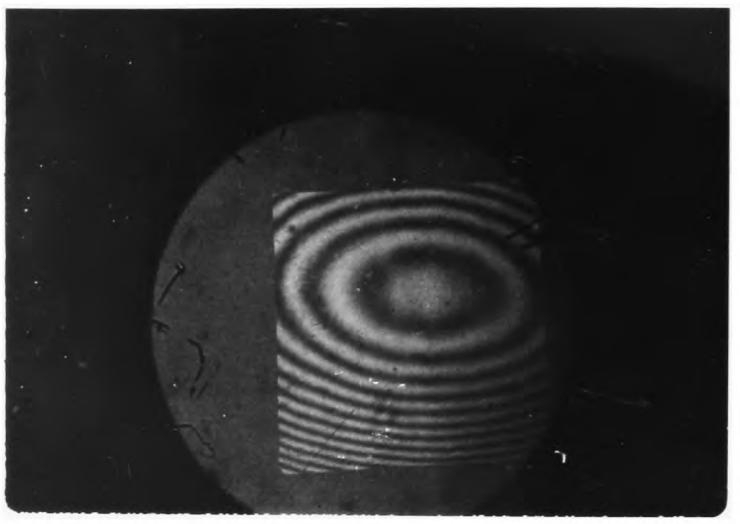
ii) The mean intensity at the blank is approximately double that from a single collimator system and can be made more uniform.

The main objection to this system is that the alignment of the collimators is extremely critical if identical wavefronts are to be formed. This is shown in figure 2.7, which shows interferograms in zero and first order of a grating made in this system. It can be seen

that even after careful adjustment, serious errors remain.



zero



first

2.7

For large aperture systems, where beam splitters are unobtainable, this disadvantage of the two collimator method may be avoided by holographically reconstructing identical wavefronts (17), but this can only be done at the expense of a large loss of intensity and a consequent increase in exposure time and stability problems.

For the relatively small aperture system described here, it was decided to use the beam splitter arrangement, as this requires a minimum of accurate alignment, yet still gives acceptable results over an aperture of several centimetres.

2.3.2 Description of the system used

The design of the system eventually used is shown figure 2.8. Figure 2.9 shows a photograph of the system. The optical components will be described separately.

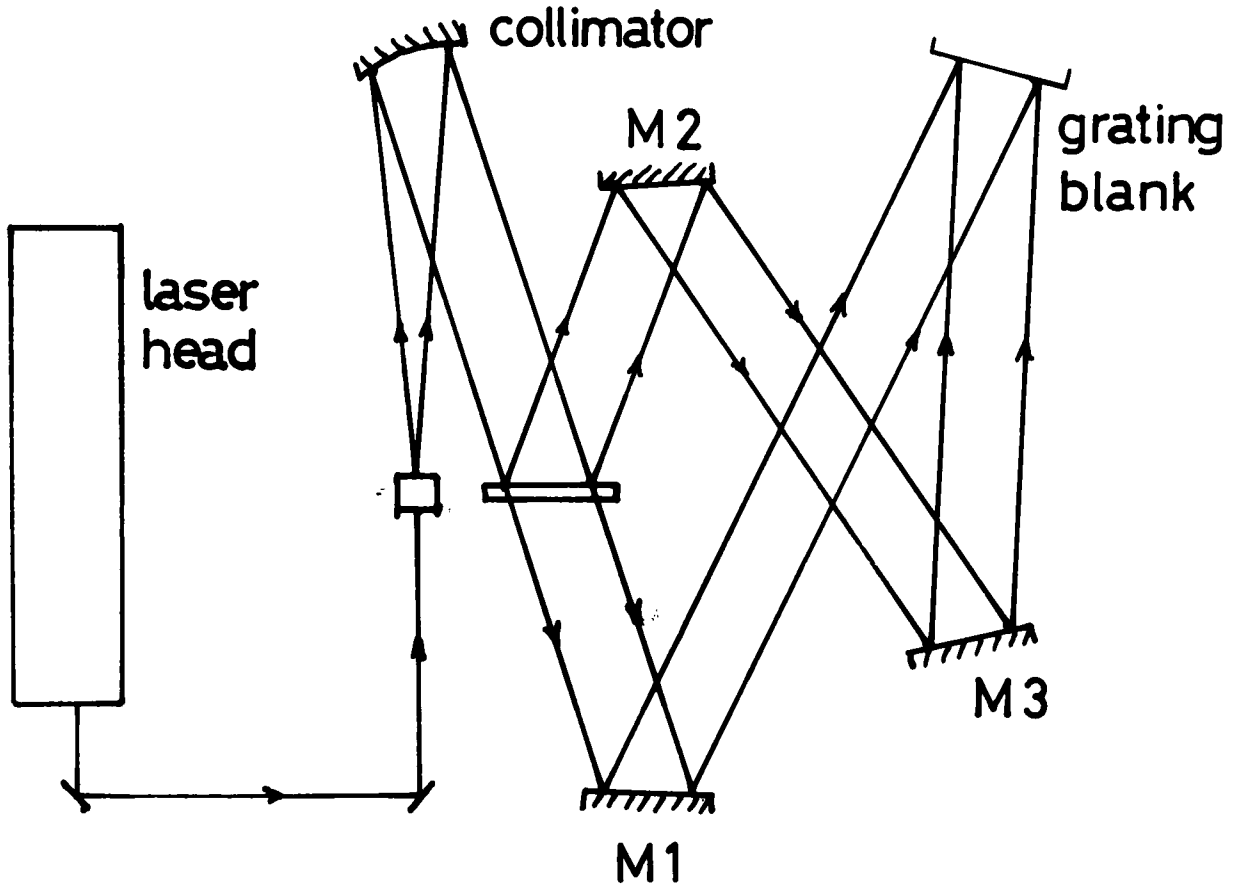


Fig. 2.8

2.3.2(a) Beam Expander

The beam expander consists of a microscope objective whose focal length must be chosen so that the full aperture of the collimator is illuminated uniformly. Since the intensity distribution in the laser beam is Gaussian, a compromise must be reached between loss of intensity caused by overexpanding the beam and insufficient expansion which allows the non-uniform outer regions of the beam to enter the aperture.

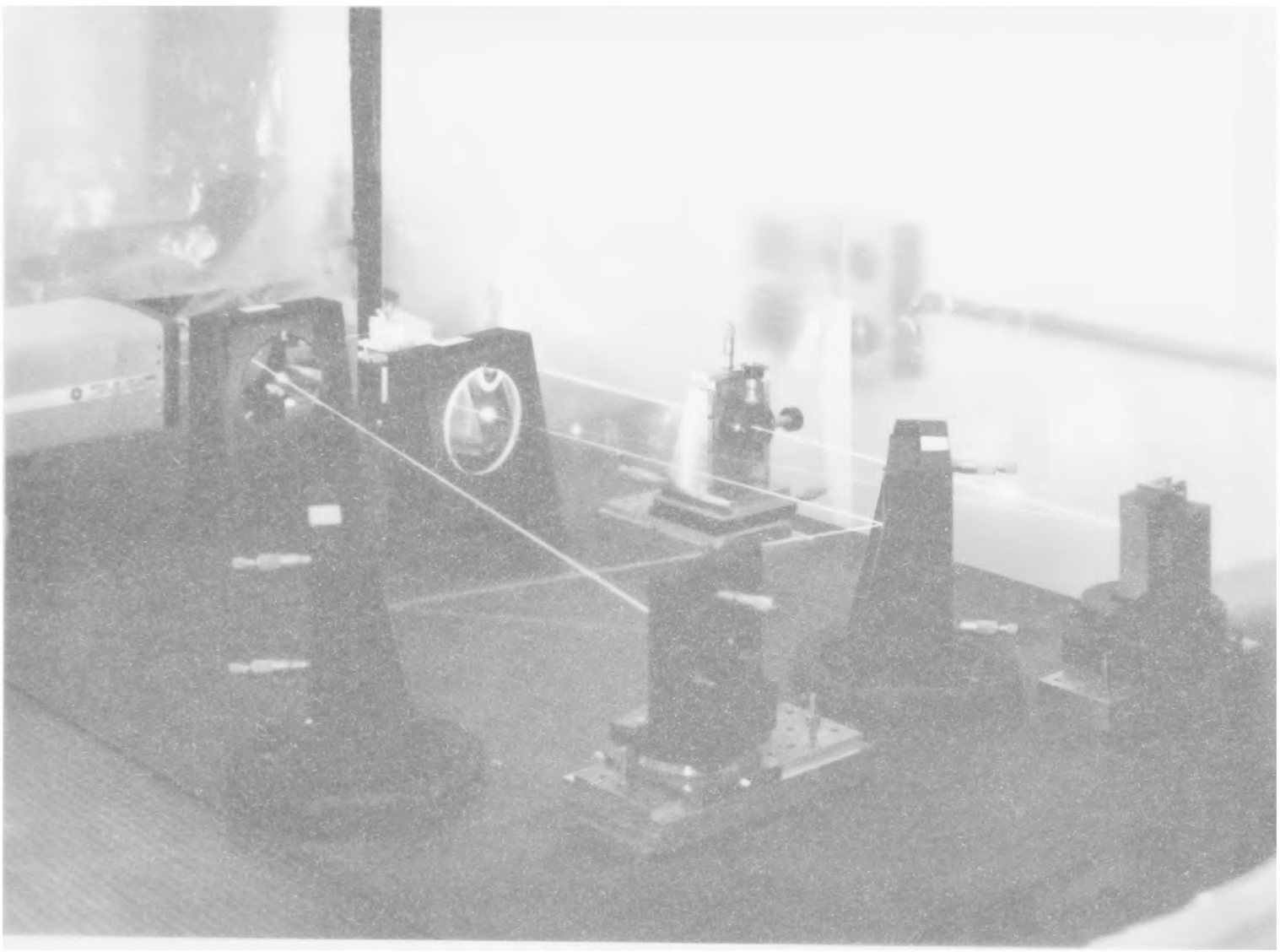


Fig. .9

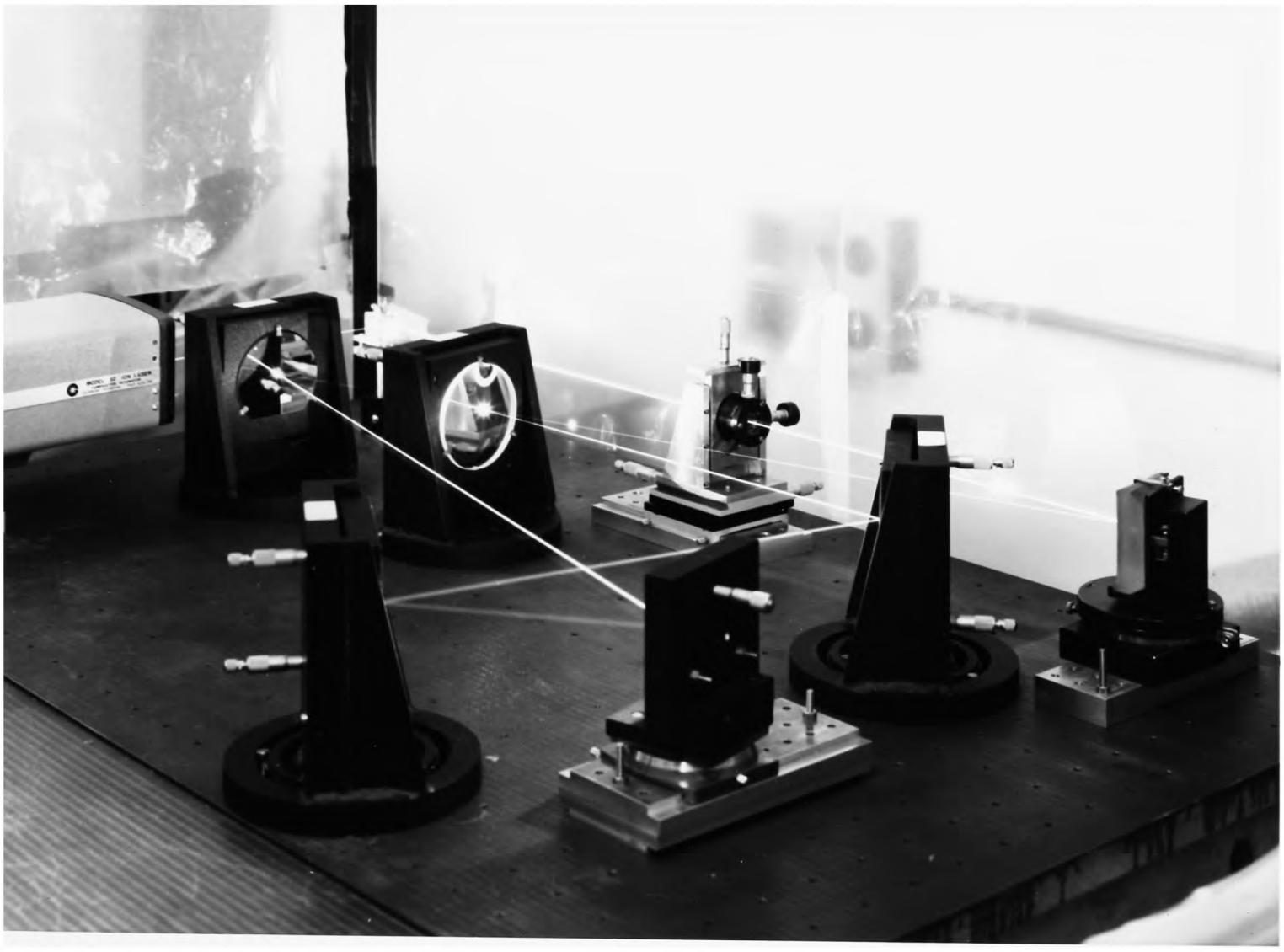


Fig. 2.9

A reasonable compromise was found with a x20 objective for blanks up to 3cm diameter and a x40 objective for blanks up to 6cm diameter. Intensity variations in the beam caused for instance by diffraction from dust particles on the expander lens are removed by a pinhole spatial filter (15) placed at the focus of the microscope objective. A 10μm diameter pinhole is used.

The holder for the beam expander/spatial filter unit must permit several precise adjustments to be made; the pinhole must be placed at the focus of the lens to an accuracy of a few microns, and the whole unit must be positioned at the focus of the collimator to an accuracy of 0.1mm.

The mount for the lens and pinhole is a commercially produced unit (Oriel Optics type B-32-20) capable of providing the necessary relative adjust ments.

It is mounted on a translation stage with micrometer adjust ments in three directions. The complete unit is well shown in figure 2.9.

2.3.2 (b) Collimator

The collimator is an off-axis paraboloidal mirror of focal length 50 cm. It is held in a mount which is capable of fine adjustments of tilt about vertical and horizontal axes.

2.3.2 (c) Beam Handling Components

The beam splitter and plane mirrors used to recombine the two beam must be of a high optical quality so that no additional aberrations are introduced into the recording wavefronts.

The mirrors must be arranged so that each beam undergoes the same number of reflections. This is to reduce the effects of assymmetrical aberrations in the wavefront produced by the collimator.

For this reason, three plane mirrors are required in addition to the beam splitter. The components are 4.5" in diameter with aberrations less than $\lambda/10$ over an aperture of 3".

The mounts for these components must be designed to prevent motion of the mirror surface relative to the base plate. This has been done as follows. The component is held in a cell which is kinematically located relative to the vertical section of the mount. Micrometers provide fine adjustments of vertical and horizontal tilt.

The upright section of the mount is fixed to a circular baseplate and is braced by triangular sections. All the joints are welded for maximum stability. The base plate has circular slots cut into it which enable the mount to be screwed down to the table with any orientation.

2.3.2(d) Blank holder

The blank is held in a cell which is kinematically held relative to an upright member. A micrometer provides adjustment of tilt about a horizontal axis and the unit is mounted on a rotation stage which can be mounted on a dovetail rail.

2.3.2 (e) Reduction of Spurious fringes

If no precautions are taken, light passing through the blank will be reflected from the rear surface and may cause a set of spurious fringes by interference with the incident beams. To avoid this, the reflection must be minimised. Several methods of doing this were tried, including coating the rear surface with various absorbing layers (e.g. matt black paint and "Liquid Opaque", an absorber used in photography).

In most cases, however, the coating had to be carefully removed before development to avoid contaminating the solutions, and a more convenient method was finally adopted.

The absorber is the ground surface of a piece of PMMA ("Perspex") which is painted matt black. The polished face of the PMMA is index matched to the rear face of the blank by a thin film of liquid. The efficiency of this method depends on the accuracy of matching the refractive indices, but even when water (n = 1.33) is used as the matching liquid, satisfactory results are obtained.

2.3.3 Alignment of the system

Once the components have been placed roughly in position, alignment of the system must achieve two conditions:

i) the expanded beam must be uniform in intensity and as nearly plane as possible;

ii) the divided beams must overlap on the blank at an angle which gives the desired groove spacing.

These adjustments will be dealt with separately.

2.3.3(a) Adjust ment of the beam-forming components

The first stage of alignment is to set the pinhole spatial filter at the focus of the beam expander. This is conveniently carried out using the micrometer adjustments on the unit. (The intensity of the laser beam must be reduced to avoid damaging the edge of the pinhole with the intense focussed spot). The position of correct alignment is obvious as that which gives a uniformly illuminated beam free from any diffraction fringes due to the pinhole.

In order to collocate the beam expander and the collimator, some means of observing the flatness of the collimated wavefronts is required. In this arrangement, this is provided by using the beam splitter as a

shearing interferometer (20). The weak reflection from the uncoated face of the beam splitter is sheared relative to the main reflection, and so unless the incident wavefront is perfectly plane, weak interference fringes will be observed where the beams overlap.

This property is utilised by observing the reflected beam while translating the beam expander in a direction parallel to the axis of the collimater. As the position of correct focus is approached, the spacing of the fringes increases reaching a maximum at the best position. The expander mount is locked in this position and then by fine adjustments of collimator tilt and beam expander translation, the alignment can be improved to the point where the expanded beam is of uniform intensity. This is the position of optimum alignment.

The accuracy of this method of alignment depends on the quality of the beam splitter. In this case, the uncoated surface is specified as being flat and parallel to the coated surface to $\lambda/5$. Hence the expanded wavefront will be flat to this order of accuracy.

2.3.3(b) Adjustment of the fringe spacing

If the beams are incident symmetrically onto the blank, the fringe spacing is given by

$$d = \frac{\lambda}{2\sin\theta}$$

where θ is the angle of incidence of the beams. Hence for a particular value of λ, a table of values of θ corresponding to values of d may be drawn up (Table 2.1)

TABLE 1.

INCLUDED HALF-ANGLE BETWEEN BEAMS OF WAVELENGTH LA WHICH INCIDENT ON A BLANK AT THIS ANGLE GIVE A GRATING OF M LINES/MM

LA:	4579	4765	4880	4965	5017	5145	6328
M							
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
100	1.31	1.37	1.40	1.42	1.44	1.47	1.81
200	2.62	2.73	2.80	2.85	2.88	2.95	3.63
300	3.94	4.10	4.20	4.27	4.32	4.43	5.45
400	5.25	5.47	5.60	5.70	5.76	5.91	7.27
500	6.57	6.84	7.01	7.13	7.21	7.39	9.10
600	7.90	8.22	8.42	8.57	8.66	8.88	10.9
700	9.22	9.60	9.83	10.0	10.1	10.4	12.8
800	10.6	11.0	11.3	11.5	11.6	11.9	14.7
900	11.9	12.4	12.7	12.9	13.0	13.4	16.5
1000	13.2	13.8	14.1	14.4	14.5	14.9	18.4
1100	14.6	15.2	15.6	15.8	16.0	16.4	20.4
1200	15.9	16.6	17.0	17.3	17.5	18.0	22.3
1300	17.3	18.0	18.5	18.8	19.0	19.5	24.3
1400	18.7	19.5	20.0	20.3	20.6	21.1	26.3
1500	20.1	20.9	21.5	21.9	22.1	22.7	28.3
1600	21.5	22.4	23.0	23.4	23.7	24.3	30.4
1700	22.9	23.9	24.5	25.0	25.2	25.9	32.5
1800	24.3	25.4	26.1	26.5	26.8	27.6	34.7
1900	25.8	26.9	27.6	28.1	28.5	29.3	37.0
2000	27.3	28.5	29.2	29.8	30.1	31.0	39.3
2100	28.7	30.0	30.8	31.4	31.8	32.7	41.6
2200	30.2	31.6	32.5	33.1	33.5	34.5	44.1
2300	31.8	33.2	34.1	34.8	35.2	36.3	46.7
2400	33.3	34.9	35.8	36.6	37.0	38.1	49.4
2500	34.9	36.6	37.6	38.4	38.8	40.0	52.3
2600	36.5	38.3	39.4	40.2	40.7	42.0	55.3
2700	38.2	40.0	41.2	42.1	42.6	44.0	58.7
2800	39.9	41.8	43.1	44.0	44.6	46.1	62.4
2900	41.6	43.7	45.0	46.0	46.7	48.2	66.6
3000	43.4	45.6	47.1	48.1	48.8	50.5	71.7
3100	45.2	47.6	49.1	50.3	51.0	52.9	78.8
3200	47.1	49.7	51.3	52.6	53.4	55.4	-
3300	49.1	51.8	53.6	55.0	55.9	58.1	-
3400	51.1	54.1	56.1	57.6	58.5	61.0	-
3500	53.3	56.5	58.6	60.3	61.4	64.2	-
3600	55.5	59.1	61.4	63.3	64.6	67.8	-
3700	57.9	61.8	64.5	66.7	68.1	72.1	-
3800	60.5	64.9	68.0	70.6	72.4	77.8	-
3900	63.2	68.3	72.1	75.5	78.0	-	-
4000	66.3	72.4	77.4	83.2	-	-	-

The value of θ taken from this table is transferred to the system as follows:

i) The blank is replaced by an aluminised flat of the same dimensions and this is masked by a piece of card with a 1cm hole at the centre.

ii) An iris diaphragm placed between the expander and the collimator is used to reduce the beam to a thin axial pencil.

iii) M1 (see figure 2.8) is adjusted so that the beam strikes the centre of the flat.

iv) The blank holder is rotated until this beam is reflected back along its path.

v) The blank holder is rotated through an angle θ (measured on the scale of the rotary table) so that the reflected beam travels in the direction of M3.

vi) M3 is placed so that the beam strikes it centrally and is reflected towards M2.

vii) M2 is rotated so that the beam strikes the reflecting surface of the beam splitter at the same point as the incident beam.

viii) The beam splitter is rotated until its reflected beam travels in the reverse direction round the system and strikes the blank centrally. (If a large rotation of the beam splitter is required the process must be repeated from step (iii) to allow for the displacement of the transmitted beam).

ix) The iris is open up to illuminate the full aperture of the blank.

2.3.3(c) <u>Setting the fringe spacing accurately</u>

The above sequence of adjustments enables the fringe spacing to be set to an accuracy of around 1 per cent.

In most cases this accuracy is sufficient, but in some applications (e.g. preparing gratings for calibrated monochromators) it is necessary to set the fringe spacing to a much higher accuracy (typically 0.05%).

This accuracy may be achieved by the use of one of the following methods.

i) If a grating of the correct spacing is available, this is mounted in place of the blank after the system has been set up as described above. The angle between the beams is then varied slightly by means of the micrometer adjustments. When the spatial frequency of the pattern approaches that of the reference grating, a Moiré pattern will be formed. Further adjustments of the mirrors are made until the pattern spreads out and no fringes are visible. The pattern now has the same spacing as the grating. The accuracy of this method can be high. The frequency of the Moiré pattern is the difference between the frequencies of the pattern and the grating and for a grating of diameter 5cm. the minimum detectable spatial frequency is around 0.02 lines/mm. Thus for a grating of 1200 lines/mm an accuracy of around 0.001% is attainable.

ii) When a grating of the correct spacing is not available, the system is set up as before to give a spacing slightly less than that required.

A trial exposure is made and the spatial frequency is measured (see App.3). The turntable is then rotated through an angle α given by

$$\cos\alpha = \frac{d}{d_o}$$

where d_o is the required spacing

d is the measured spacing

The fringe spacing on the blank should now be equal to that required. The accuracy of this method depends on the value of α and accuracy of measuring d.

From the above expression for α it may be shown that

$$\frac{\Delta d_o}{d_o} = \frac{\Delta d}{d} + \tan\alpha \, \Delta\alpha$$

For a typical case where $\alpha = 3^\circ$, $\Delta\alpha = 0.1^\circ$ and $\Delta d/d = 5 \times 10^{-4}$ this shows that $\Delta d/d_o \simeq 6 \times 10^{-4}$.

Chapter 3

RECORDING THE FRINGES

3.1 Photoresist

3.2 Photoresist Techniques

- 3.2.1 Coating the blanks
- 3.2.2 Problems encountered in coating the blanks
- 3.2.3 Post bake
- 3.2.4 Exposure
- 3.2.5 Development

3.3 Metallising

3.4 The influence of exposure and development on groove profile

- 3.4.1 The response of the photoresist
- 3.4.2 Groove profile

In order to form a grating which matches ruled gratings in terms of optical quality and efficiency, it is necessary to record the interference fringes in such a way as to produce grooves in a highly reflecting surface. In addition the recording medium must have sufficient resolution to record accurately spatial frequencies up to 4000 lines/mm.

These requirements rule out any process involving silver halide emulsions and the recording medium must be chosen from those which operate on a molecular scale.

The systems of interest in this application may be divided into two general categories: sensitised gelatine systems and photoresists. Gelatine systems, such as dichromated gelatine (21) are cheaper and in some respects more convenient to use than photoresists, but because the gelatine is hygroscopic, it absorbs water during processing and distorts on drying. Also the film may soften under humid conditions.

This dimensional instability makes gelatine an unsuitable system for recording high quality gratings, and photoresists which are stable and hydrophobic, must be used.

In this chapter, a discussion of photoresists is followed by a description of the techniques used in coating the gratings. Finally, the vacuum evaporation techniques used to overcoat the grating with metal are briefly considered.

3.1 Photoresist

Commercially available photoresists are of two types, positive or negative, which involve different photochemical processes.

Negative resists respond by the hardening produced when polymer chains (usually poly-vinyl cinnamate) cross link on exposure to light and positive resists respond by becoming soluble under the action of the light.

When photoresists are used in industry (for example as etching masks in integrated circuit production) they are given sufficient exposure to render the film completely soluble or insoluble depending on the type of resist. For recording the interference fringes, however, it is desirable to use the resist in the region where the final thickness of the film varies with the total exposure received, for in this way a smooth groove profile may be obtained.

Negative resists are unsuitable for use in this manner, for if the resist is incompletely exposed, a soluble layer is left next to the substrate (22). This can create problems with adhesion. For this reason, positive resist is used exclusively in this application.

The positive photoresist used here is AZ1350, manufactured by Shipley Chemicals Ltd., Coventry.

The use of AZ1350 as a holographic recording medium has been studied by several authors (23,24) and it may be appropriate to summarise here what is known about its properties.

The resist consists of an alkali soluble resin containing a photosensitive inhibitor, which in the unexposed condition prevents the resist from dissolving. When the resist is exposed, the inhibitor is broken down, and the developer is able to etch the resist film at a rate dependent on the total exposure received.

AZ1350 is claimed to have good long term stability and has been shown to be capable of recording spatial frequencies up to 10,000 lines/mm (26).

The spectral sensitivity depends on the absorbtion spectrum of the inhibitor and in common with most other photoresists, the peak sensitivity is in the near u.v. region of the spectrum. The manufacturer's sensitivity curve shows some response up to about 500 nm, but the sensitivity increases rapidly from 450 nm to the peak at 320 nm.

At the wavelength used here, 457.9 nm, the resist is relatively slow.

3.2 Photoresist Techniques

The various stages of preparing and processing the blanks are as follows:

i) Coating	Clean blanks Whirler coat Bake
ii) Exposure	
iii) Development	Develop Rinse Dry

These will be considered in detail.

3.2.1 Coating the blanks

The resist is supplied dissolved in an organic solvent and in order to prepare the thin films required, the substrate must be coated with the solution and the solvent allowed to evaporate. There are several methods of applying the solution to the blanks, but the only one which produces a uniform film of controllable thickness

is spin coating.

In this process the substrate is held horizontally in a chuck and a small amount of resist solution is dispensed onto it. The chuck is then rotated at a fixed velocity (typically 1500-3000 r.p.m.) and the solution is thrown off apart from a thin layer which is retained by the effects of surface tension and viscosity.

When the solvent evaporates, a film of resist is left whose thickness depends only on the angular velocity of the chuck for a given set of conditions.

3.2.1 (a) Controlling the resist thickness

Many factors affect the process and so an accurate prediction of the final film thickness is difficult.

An empirical expression has been suggested, however, which accounts for the major factors (27).

The results quoted in this reference show that for a given set of conditions the thickness of the dried film is given by the expression:

$$d = k \frac{p^2}{\sqrt{\Omega}}$$

where k is a constant depending on the exact details of the process

p is the percentage solids content of the resist solution

Ω is the angular velocity of the spinner .

In addition, it is shown that for maximum uniformity, the angular acceleration of the blank should be as high as possible.

The spinner used in this work is a commercial unit (Headway Research Inc. type EC101) which uses a vacuum chuck to hold the substrates. The angular velocity is variable between 500 and

10,000 rp.m. and the acceleration is high (20,000 r.p.m./sec for the unloaded chuck). A timer is provided to stop the chuck after a predetermined time interval.

Figure 3.1 shows the dried film thickness (measured by interferometric techniques) plotted as a function of $1/\sqrt{\Omega}$ for undiluted resist. The straight line indicates that the empirical variation with Ω applies.

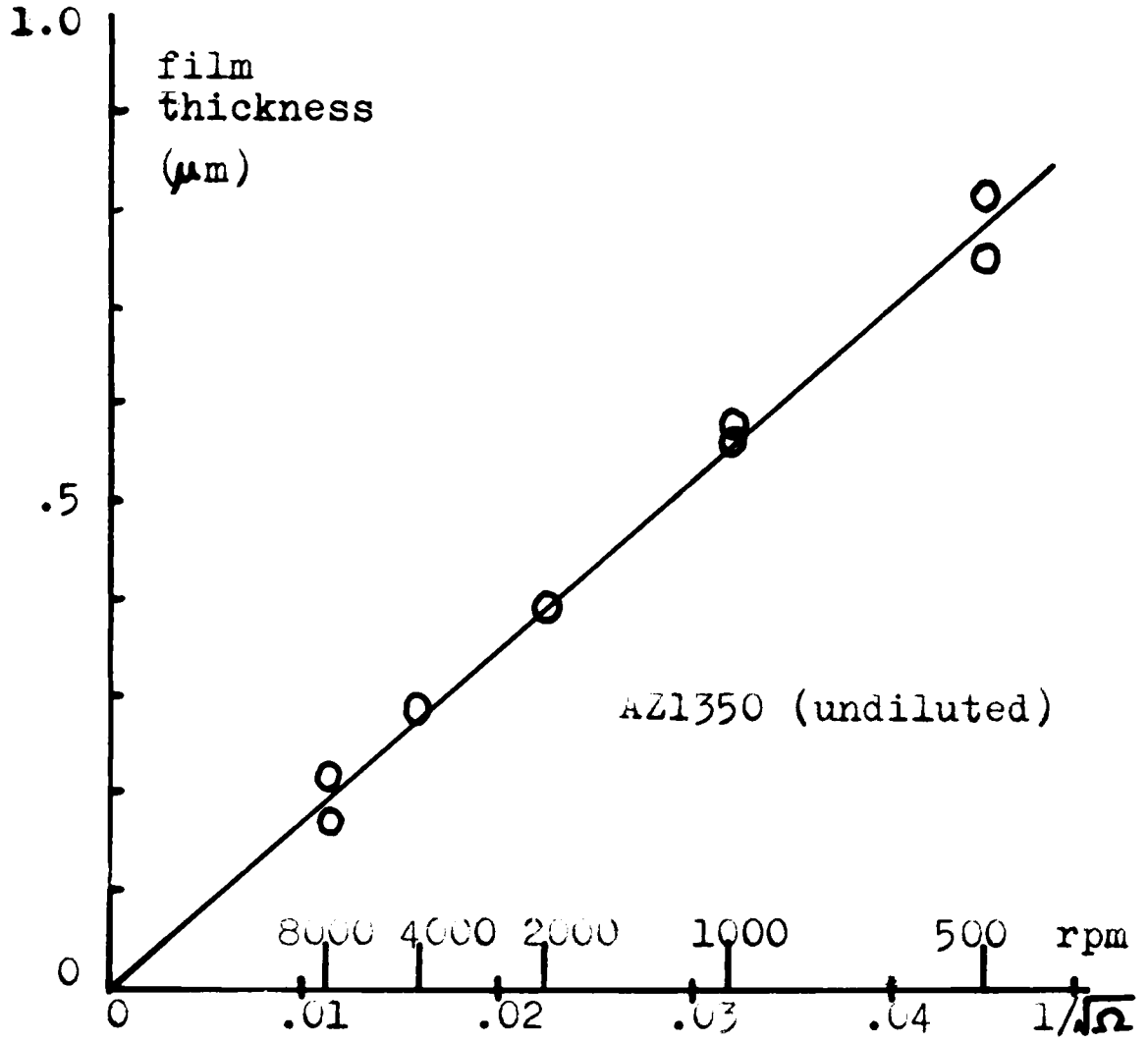


Fig 3.1

3.2.2 Problems encountered in coating the blanks

Several difficulties were encountered in producing uniform films of good quality, These will be considered seperately.

3.2.2 (a) Adhesion

In the initial stages of the project it was found that the adhesion between the resist and the substrate was poor, so that the film lifted or distorted during processing.

This was assumed to be due to the presence of grease and moisture on the substrate. (AZ1350 is reported to be sensitive even to minute traces of moisture (23,24)).

To avoid these problems, the following cleaning process was adopted:

i) Wash in hot detergent solution using an ultrasonic cleaner to assist the detergent action.

ii) Rinse in filtered tap water.

iii) Rinse in ethanol to remove most of the water.

iv) As a final cleaning process, a lens tissue is laid across the blank and a small amount of iso-propyl alcohol is dropped on it. Surface tension draws this out over the surface of the blank and holds the tissue in close contact with the surface. The tissue is slowly pulled off parallel to the blank surface and the isopropanol evaporates, leaving a clean, dry surface.

v) The blank is baked at 120°C for thirty minutes to remove the last traces of moisture.

vi) The blank is allowed to cool in a dessicator prior to coating.

3.2.2 (b) Dust particles and air bubbles

It has been found that bubbles in the resist solution leave a pinhole in the film and dust particles cause the formation of a comet-like feature due to the streaming of the resist round the obstacle during the whirling process. To eliminate the dust, the resist is applied to the blank from a syringe with a 1μ "Millipore" filter attachment, and any dust on the substrate is blown off with

dried and filtered compressed air prior to resist application.

As a further precaution, all the preparation and processing is carried out in a laminar flow clean air cabinet.

The air bubbles in the resist are introduced when the syringe is filled; they are avoided by allowing the syringe to stand vertically for some time after filling.

3.2.2 (c) Striations

Many of the coatings made by spinning were found to have a periodic thickness variation with angle, leading to radi al "striations". These are particularly troublesome in gratings, as their periodic nature causes spurious images to be formed and leads to a high level of stray light.

The striations were eventually found to be due to rapid evaporation of the solvent, resulting in the resist drying before surface tension was able to draw the film flat. In this way the flow lines were "frozen" into the film.

The problem was avoided by ensuring that the blanks had cooled before coating and by covering the chuck during the whirl. This increases the concentration of solvent vapour over the blank and so increases the drying time of the resist.

3.2.3 Post bake

As a final stage in the coating process, the blank is baked at 85°C for thirty minutes to remove all traces of solvent from the film. It has been found that the presence of solvent causes erratic variations in sensitivity across the blank.

3.24 Exposures

The standard exposure used to form the gratings corresponds to a mean energy of 2.8 mJ cm^{-2}.

The intensity in the expanded beam was measured and the following expression was used to calculate the exposure time for various laser powers and angles of incidence

$$t = \frac{222}{P \cos\theta} \quad \text{minutes}$$

where θ is the angle of incidence of the beams

P is the laser output power in milliwatts

This expression is calculated for the X40 beam expander. When the x20 expander is used, the exposure times must be divided by four.

Exposure times are controlled by a synchronous motor timer which applies a stable voltage to the modu lation input of the laser for a preset period. The laser power may be controlled by means of the applied voltage.

3.2.5 Development

Development takes place in undiluted AZ1350 developer. This was found to give more consistent results than the dilution recommended by the supplier.

The sequence of processes is as follows:

i) Development

The blank is held in a clip and is placed in the developer for the required developing time. Agitation was found necessary for consistent results and this is provided by slowly moving the blank up and down through the solution.

ii) First Rinse

The blank is placed in a bath of running filtered tap water for one minute. The slight acidity of the water terminates the etching process and the developer is rinsed away.

iii) Second Rinse

The blank is allowed to stand in distilled water for one minute.

iv) Dry and bake

The blank is blown dry with a jet of dried and filtered compressed air. The developed grating is then baked at 85°C for thirty minutes to remove the last traces of moisture.

When the grating is removed from the oven, it is stored in a dessicator until it is metallised.

3.3 Metallising

To give the gratings a high efficiency when used in reflection, they are overcoated with metal. The metals most commonly used in this work are aluminium, silver and gold.

The metals were deposited by evaporation in a vacuum. The apparatus, which was designed and operated by Mr.C.W.Goodwin, consists of a vacuum tank containing the vapour source (a heated tungsten wire which carries the metal to be evaporated) and the gratings to be coated.

Provision is made for ionic bombardment of the grating to give a clean surface for deposition.

The thickness of the metal layer is measured indirectly by observing the optical transmission of a monitor plate. In all cases an opaque layer was deposited on the monitor plate, but in some

cases, particularly for gratings with deep grooves, the films deposited on the gratings were transparent.

3.4 The influence of exposure and development time on groove depth

3.4.1 The response of the photoresist

The dynamics of the exposure-development process are complicated by absorption of the incident light through the resist layer, but if this is neglected, the rate of etching may be shown to be an exponential function of exposure (23). It is shown in this reference that the error incurred by neglecting absorption in films of 1μm thickness at a wavelength of 458 nm is around 15%. Again if absorption is neglected the etch rate is constant throughout the resist layer, and so the final thickness depends linearly on development time.

The combined effect of the two parameters, exposure and development time may thus be expressed as

$$d = d_o - t\alpha \exp(\beta E)$$

where d = film thickness after development (Å)

d_o = initial film thickness

t = development time (min.)

α = etch rate at zero exposure (Å/min)

β = wavelength dependant constant

E = total exposure ($mJ\ cm^{-2}$)

The accuracy of this expression was verified by exposing films of photoresist to a known constant exposure and developing for a known time. The thickness of a step cut in the film was then measured by observing interference fringes between it and a flat.

The results are presented in figures 3.2 and 3.3.

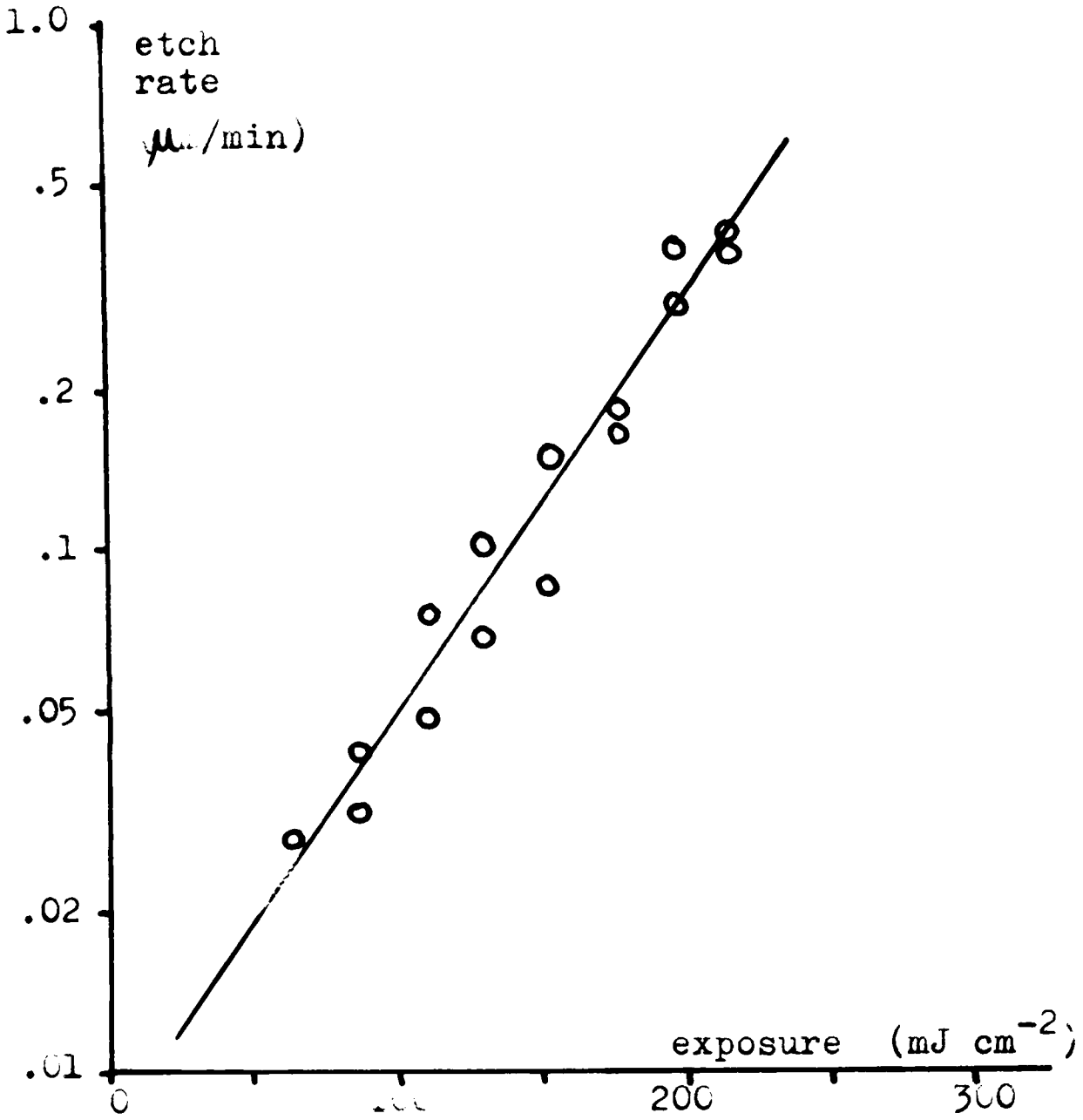

1.0
etch
rate
(μm/min)
.5
.2
.1
.05
.02
.01
0
200
300
exposure (mJ cm^{-2})

Fig. 3.2

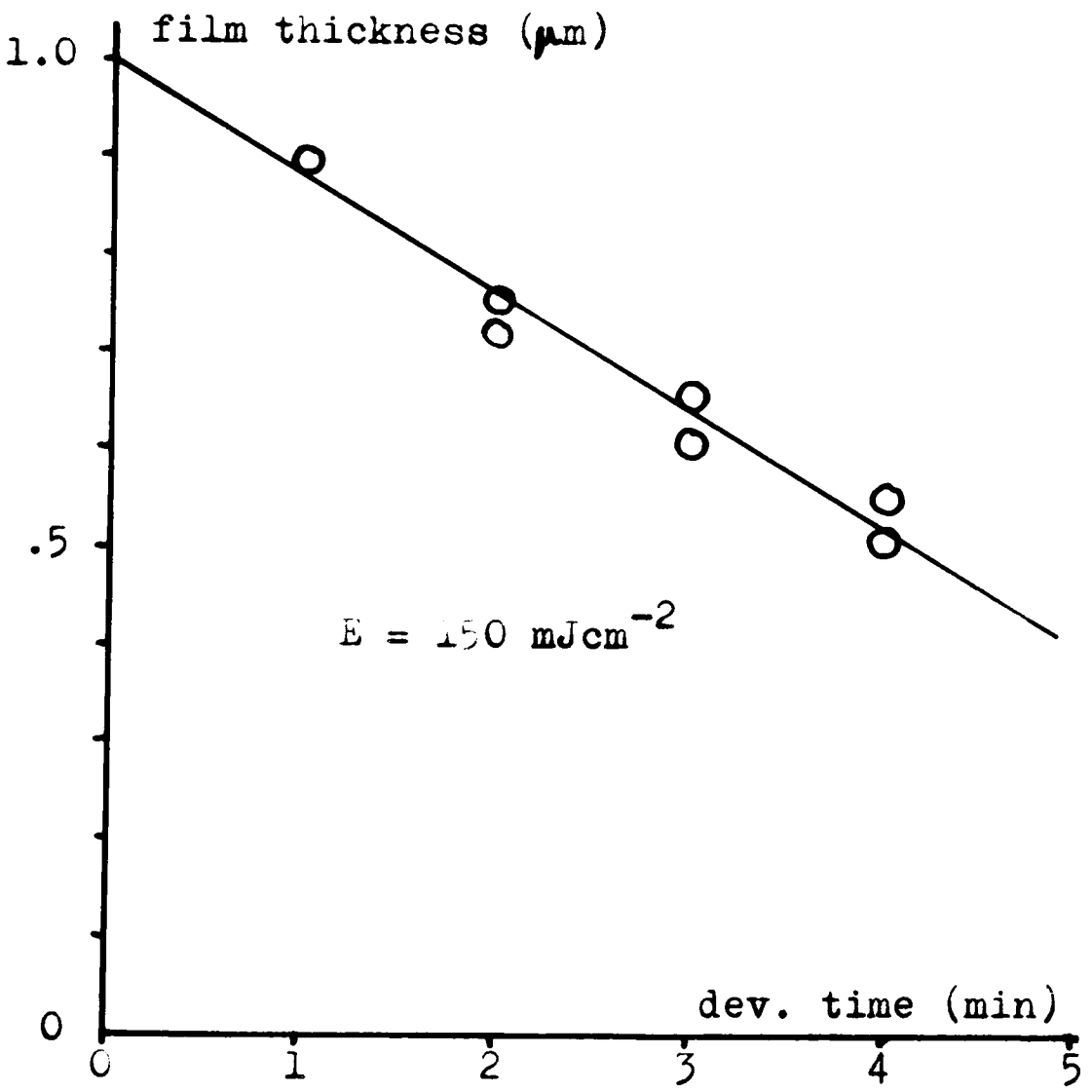

film thickness (μm)
1.0
.5
0
E = 150 mJcm^{-2}
dev. time (min)
0
1
2
3
4
5

Fig 3.3

Figure 3.2 shows the etch rate plotted on a logarithmic scale as a function of exposure. It may be seen that the points approximate to a straight line indicating that the exponential dependance is maintained at least over the range of exposures considered. Figure 3.3 indicates that the resist is etched linearly.

3.4.2 Groove Profile

It is seen that the recording process is inherently non-linear. In this work, however, it is desirable to record the sinusoidal intensity distribution accurately as sinusoidal grooves so that the groove profile may be expressed simply in terms of two parameters.

It is therefore necessary to establish which combinations of E and t will give the minimum departure from a sinusoidal profile over a wide range of groove depths.

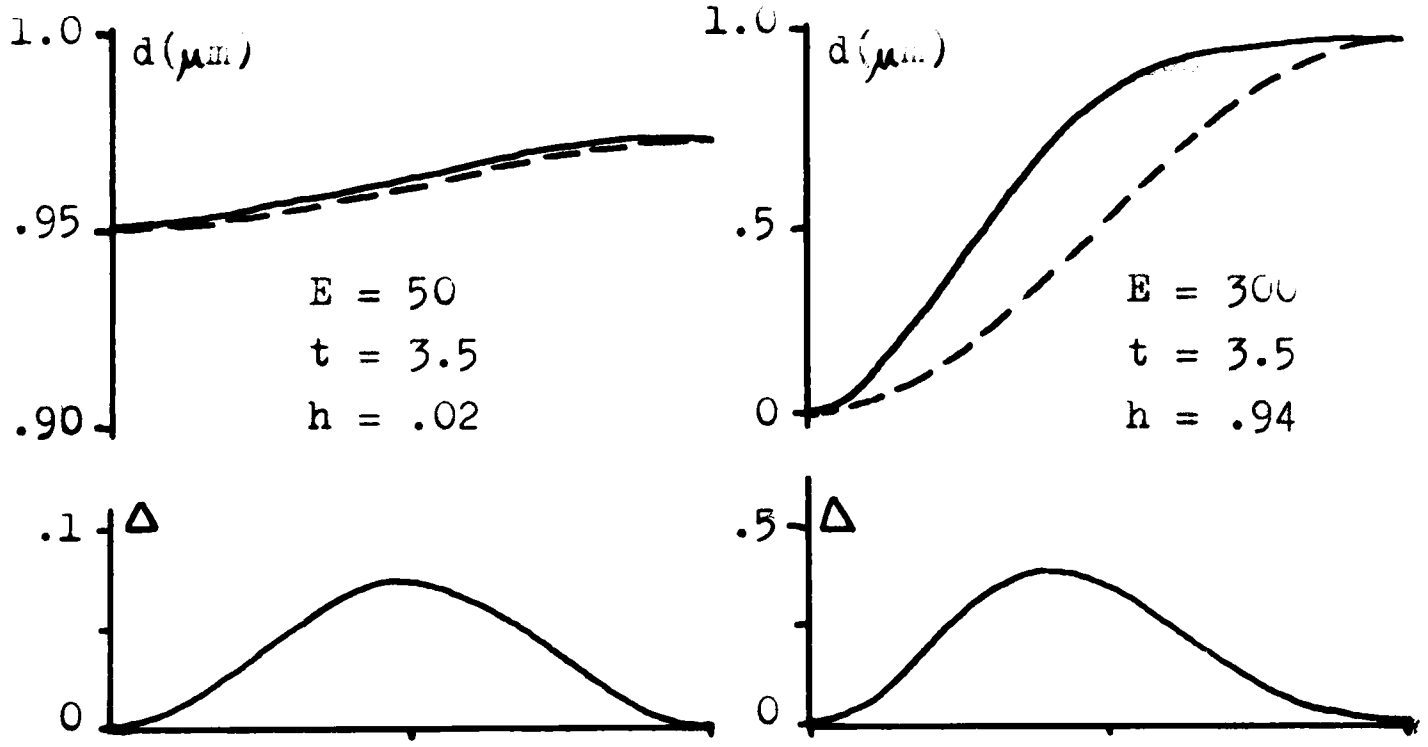


Figure 3.4 shows the effect of using a constant development time and varying the groove depth by means of the exposure. (The diagrams show the computed response and the ideal sinusoidal profile, together with the fractional distortion - the difference

between the two profiles divided by the groove depth).

It is seen that as is to be expected from the exponential characteristic, the distortion increases dramatically at large exposures (note the different scales). The effect of the distortion is to flatten the peak of the groove. The effect of the exponential characteristic may be avoided by using a pre-exposure of uniform intensity:

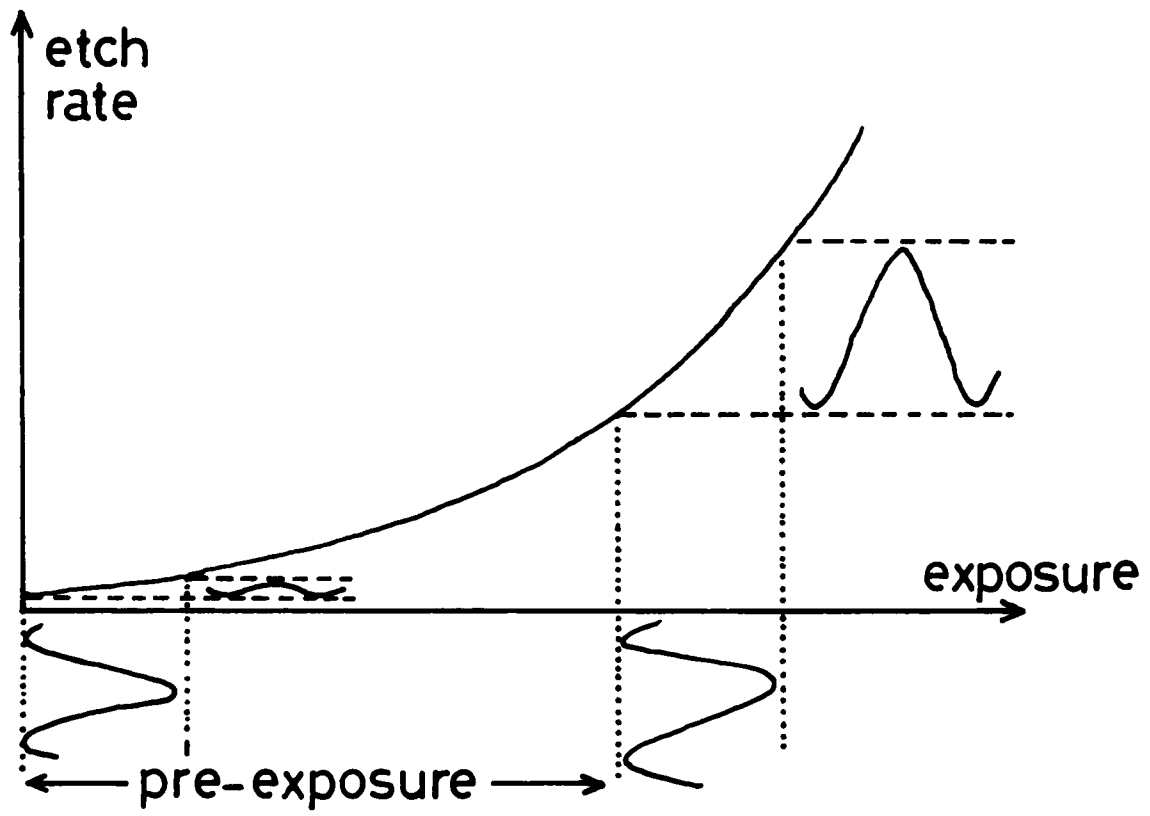


Figure 3.5 shows how this technique increases the sensitivity of the system. In addition, it may be shown that if the exposure to the fringes remains constant, the profile distortion is independent of pre-exposure. Therefore it is possible to change the groove depth simply by varying the pre-exposure.

The disadvantage of this method is that an extra variable, the pre-exposure energy, is introduced into the system, which because

of the exponential characteristic, has a large effect on the system at large values. Because of this, it was found difficult to obtain consistent results.

The mode of operation finally adopted was to keep the exposure constant, but to change the groove depth by varying the development time.

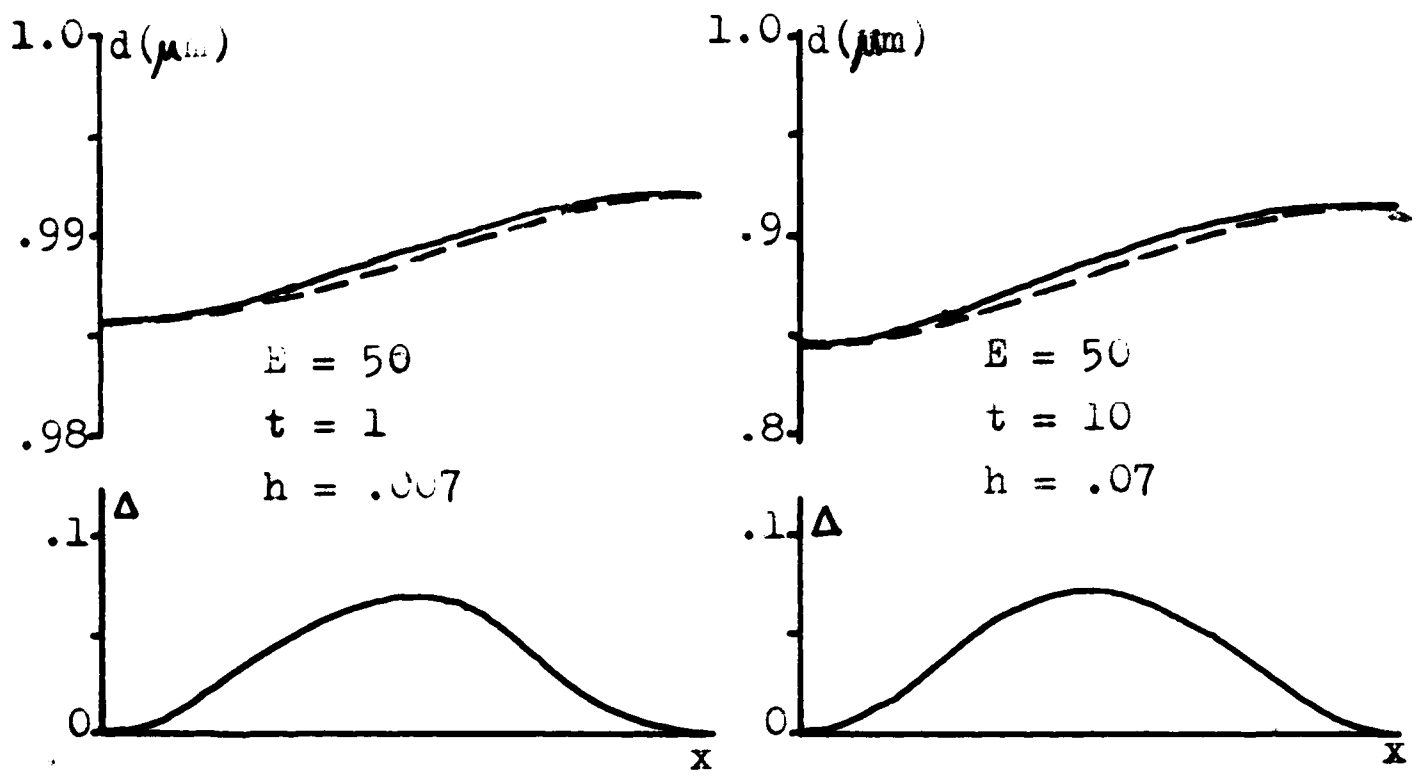


Figure 3.6 shows the effect of this on the groove profile. The groove distortion is independent of groove depth and is determined only by the exposure.

In figure 3.7, the groove depth of several 300 line/mm gratings is plotted against development time.

The groove depth was estimated by comparing measured efficiencies with the predictions of scalar diffraction theory.

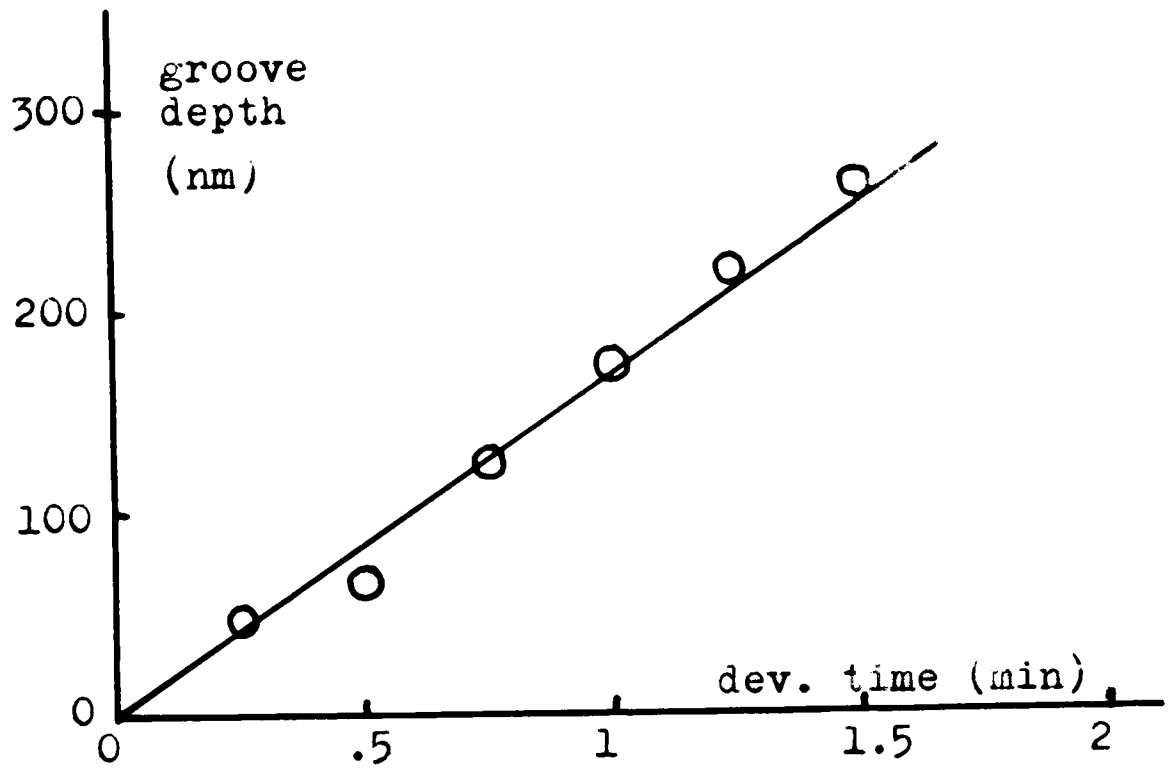


Fig. 3.7

The linear dependence on development time is well shown, while the close agreement of the measured efficiencies with scalar theory shows that the departure from a sinusoidal groove profile is small.

Chapter 4

THE EXPERIMENTAL DETERMINATION OF GRATING PROPERTIES - APPARATUS AND TECHNIQUES

4.1 The measurement of efficiency

4.1.1 Efficiency at constant wavelength

4.1.2 Efficiency in constant deviation mounts

4.1.3 Efficiency in zero order

4.2 The measurement of scattered light

4.2.1 The total flux of scattered light as a function of incidence

4.2.2 The angular distribution of scattered light

The properties of diffraction gratings which are of interest in this study are efficiency, defined as the ratio of the intensity in a diffracted order to the incident intensity, and the intensity and angular distribution of scattered light.

The efficiency of a diffraction grating is a function of the angle of incidence, wavelength and polarisation of the incident light and also depends on the groove profile and spatial frequency of the grating and the optical properties of the metal in which it is formed.

Because of this, the conditions under which the efficiency is measured must be accurately defined. The techniques described here allow the following measurements to be made:

i) efficiency as a function of angle of incidence at constant wavelength;

ii) efficiency as a function of wavelength with a constant deviation between incident and diffracted beams;

iii) efficiency in zero order as a function of wavelength at constant angle of incidence.

The properties of scattered light which are of interest are:

i) the total flux of scattered light as a function of angle incidence at constant wavelength;

ii) the angular distribution of scattered light.

All the measurements described were carried out in the spectral region

350-800 nm.

4.1. Efficiency Measurements

4.1.1 Efficiency at constant wavelength

4.1.1 (a) Apparatus

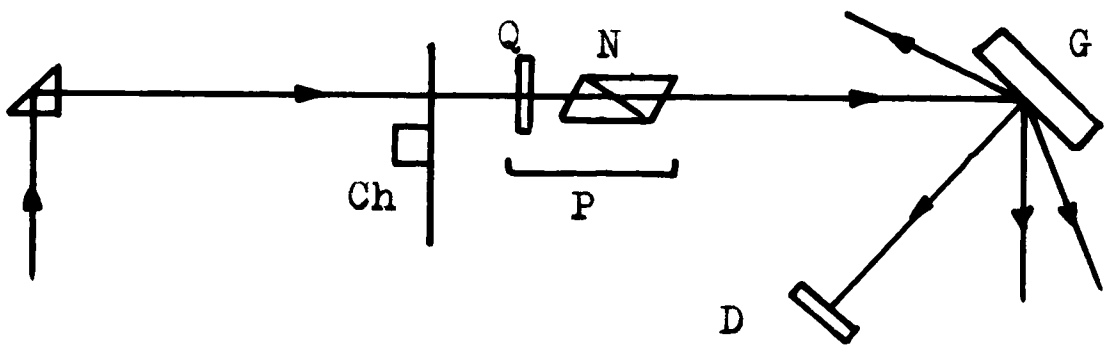


Fig. 4.1

The optical system is shown in figure 4.1. A beam of light from a laser is directed to the optical bench by a series of prisms and passes through a chopper wheel driven by a synchronous motor.

The direction of linear polarisation of the beam can be changed by the polariser P. This consists of a quarter wave plate, Q, which converts the linear polarisation of the beam to elliptical polarisation, followed by a Nicol prism, N, which can be rotated to select any angle of polarisation.

Once N has been adjusted, Q may be rotated to maximise the intensity. The beam passes to the grating, G, mounted on a turntable, and after diffraction the orders are intercepted by the detector D.

The detector consists of a PbS photocell which is mounted on an arm which pivots about the axis of the turntable. This ensures that the detector is always normal to the beams.

The output from the detector is amplified and passes to a phase sensitive detector, which uses a reference signal derived from the chopper wheel. The output voltage of the detector is displayed on a digital voltmeter.

The light source for this measurement is an argon ion laser which can be stabilised in intensity by means of an internal feedback circuit. Measurements were made at the following wavelengths: 454.5; 457.9; 488.0 and 514.5 nm.

4.1.1 (b) Operation

To take a set of measurements, the following sequence of operations was used.

i) The polariser was adjusted to give the required direction of polarisation.

ii) The detector was reversed in its mount and moved to intercept the incident beam. The gain of the amplifier and the intensity of the laser beam were adjusted until the DVM displayed 10.0 V.

iii) The detector was returned to the correct position. The grating was mounted and aligned so that it was normal to the incident beam with its grooves parallel to the axis of rotation of the turntable.

iv) The detector was rotated to intercept each order inturn, the efficiency being read directly from the DVM.

v) The turntable was rotated to give other angles of incidence and step (iv) was repeated until sufficient points have been taken to build up a curve of efficiency against angle of incidence for each order.

Because of obstruction of the incident beam by the detector, it is impossible to measure the efficiency of orders which are diffracted at angles close to the angle of incidence. In most cases, this is not serious, as the obstruction is small and the missing points may be interpolated with sufficient accuracy. Figure 4.2 shows a typical record of the variation of efficiency in the first order of a HDG as a function of angle of incidence.

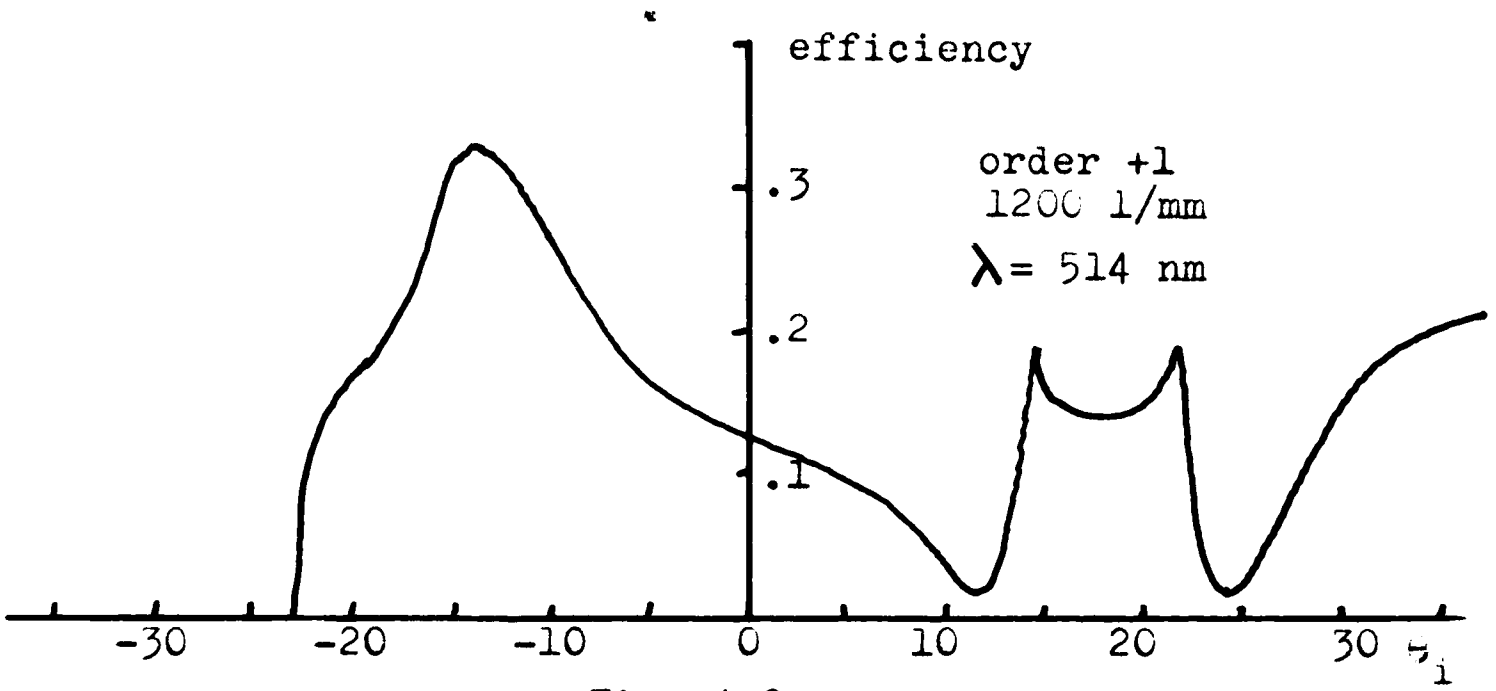


Fig. 4.2

4.1.2 Efficiency in constant deviation mounts

4.1.2 (a) Discussion

The variation of efficiency with wavelength is difficult to measure, as now one must take into account the spectrum of the source and the spectral response of the detector and other optical components.

To make absolute efficiency measurements, two monochromators are required. The first provides a monochromatic beam of known intensity. This then enters the second, which carries the grating under test. Unless the motions of these monochromators can be synchronised, this method is prohibitively time consuming if measurements are to be made with enough resolution to observe narrow Wood's anomalies.

A system was investigated which avoids the use of synchronised monochromators. This is shown in figure 4.3.

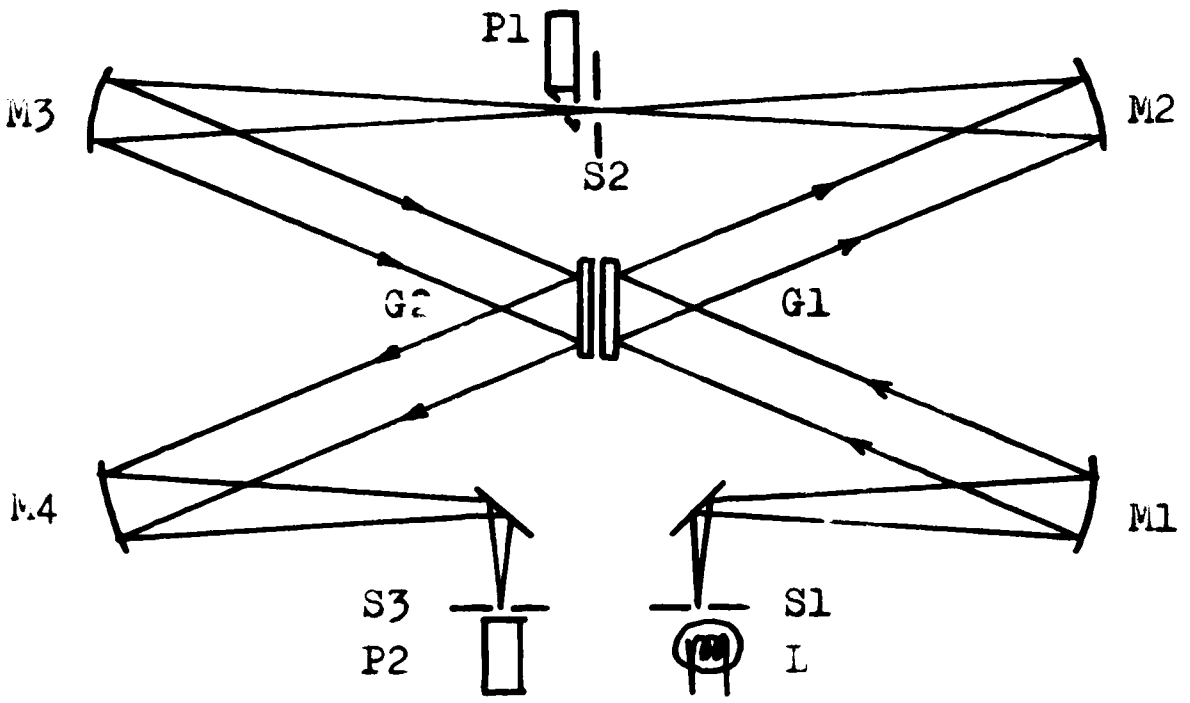


Fig. 4.3

The system consists of a double Czerny-Turner monochromator in which the gratings are mounted back-to-back on the same turntable.

The first monochromator, S1, M1, G1, M2, S2 forms a spectrum of the source, L, in the plane of S2 and light of a particular wavelength, λ , passes into the second monochromator, S2, M3, G2, M4 S3. Some of this light is sampled by a small prism behind S2 and the intensity measured by the photomultiplier P1. Now provided that the two monochromators have identical geometries and that G2 has the same groove spacing as G1, light of wavelength λ will propagate through the second monochromator where its intensity is measured by P2 (matched with P1).

Thus by rotating the turntable, the incident and diffracted intensities in the second monochromator may be recorded simultaneously as a function of wavelength and so by electronic or numerical division, the efficiency of G2 may be found.

This system has the advantage that absolute efficiency measurements may be made using a single turntable, but was eventually abandoned for the following reasons:

i) The complexity of the system meant that alignment was difficult and the system drifted easily;

ii) The system was unreliable when the efficiency of the gratings was low. This was particularly damaging because anomalies occurred simultaneously in the two monochromators. A strong anomaly in G1 could mask that in G2 by reducing the incident intensity to a low level;

iii) The system lacked versatility. It was found to be desirable to measure efficiencies in Littrow, or near Littrow configurations, which was impossible using this system.

In view of these disadvantages, it was decided to use a single monochromator containing the grating under test and to calibrate the system to allow for the spectral response of the components.

The features of interest in the efficiency curves are the Wood's anomalies. This means that the calibration need not be very accurate for the following reasons:

i) The anomalous regions occur over a restricted wavelength range, so the variation of spectral response may not be significant;

ii) Anomalous effects are polarisation dependent, so if necessary, the ratio of the response in the two orthogonal

polarisations may be measured, to give a quantity which is independant of the spectral response of the instrument.

4.1.2 (b) Optical system

The system finally adopted is shown in figure 4.4

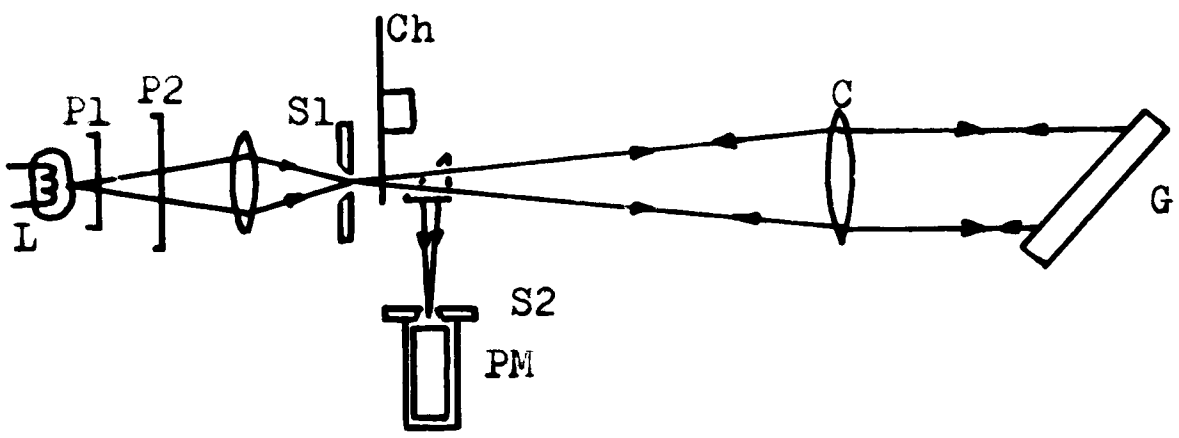


Fig. 4.4

The arrangement is basically that of a Littrow monochromator.

The light source is a quartziodine projector lamp (24V, 150W), which is supplied from a constant voltage transformer to reduce the effects of mains fluctuations.

The light from the lamp is condensed onto the entrance slit, S1, and after passing through the chopper, Ch, is collimated by the lens C.

After diffraction by the grating, G, mounted on a turntable, the light passes back through C, but is reflected by an auxiliary prism below the plane of the incident beam onto the exit slit. S2, behind which is a photomultiplier.

Polarisation

The incident light is polarised with the electric vector parallel or perpendicular to the grooves by means of the polariser P2. This is a solenoid operated pendulum which carries two pieces of polarising sheet (Polaroid HN38) with their axes in the required directions. It is driven by a Schmitt trigger so the state of polarisation depends on the magnitude of the voltage applied to the input.

The lamp was found to have a significant polarisation in its output, so an additional polariser, P1, was placed in the system. This is rotated until the incident beam has equal intensity in both polarisations.

Turntable

The turntable is driven from a stepping motor (Impex type PD18) via a 360: 1 gear train. There are 48 steps per revolution of the motor, so one step rotates the turntable through 1.25 minutes of arc.

In a Littrow configuration at an angle of incidence θ , this rotation corresponds to a wavelength step

$$\Delta\lambda = 7.272 \times 10^{-4} \frac{d \cos\theta}{n}$$

where d is the grating period

n is the order

For a 1200 line/mm grating in first order at 500 nm, this represents a wavelength step of 0.6 nm.

Photomultiplier

The photomultiplier is an eleven stage tube with an S-20 photocathode (EMI type D125B). The extended spectral response enables measurements to be made between 350 and 800 nm.

4.1.2 (c) Detection system

For convenience in collecting and processing the data from the system, the signal is converted into digital form and stored in a multichannel scaler (MCS). Each step of the motor is associated with two memory channels in the MCS, one for each polarisation.

A block diagram of the system is shown in figure 4.5.

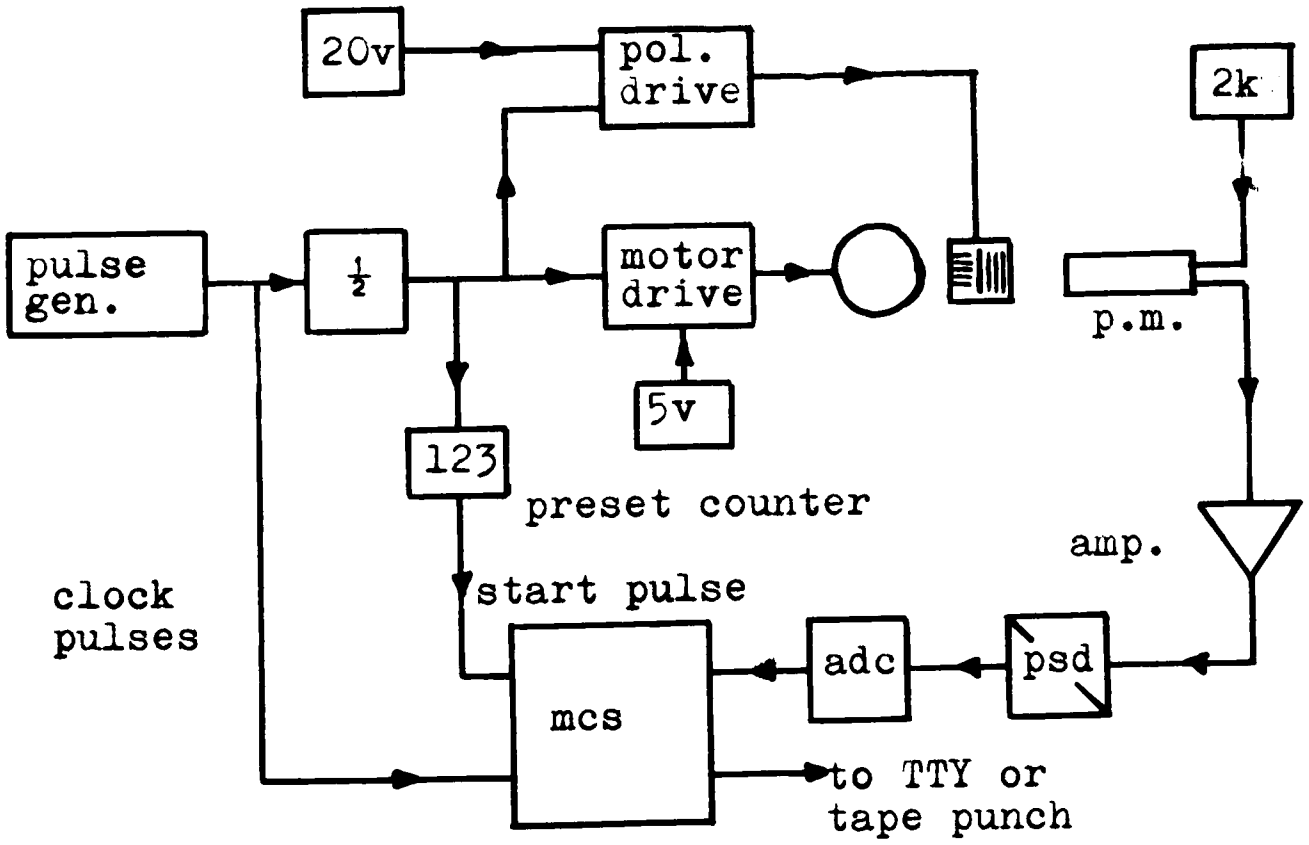


Fig. 4.5

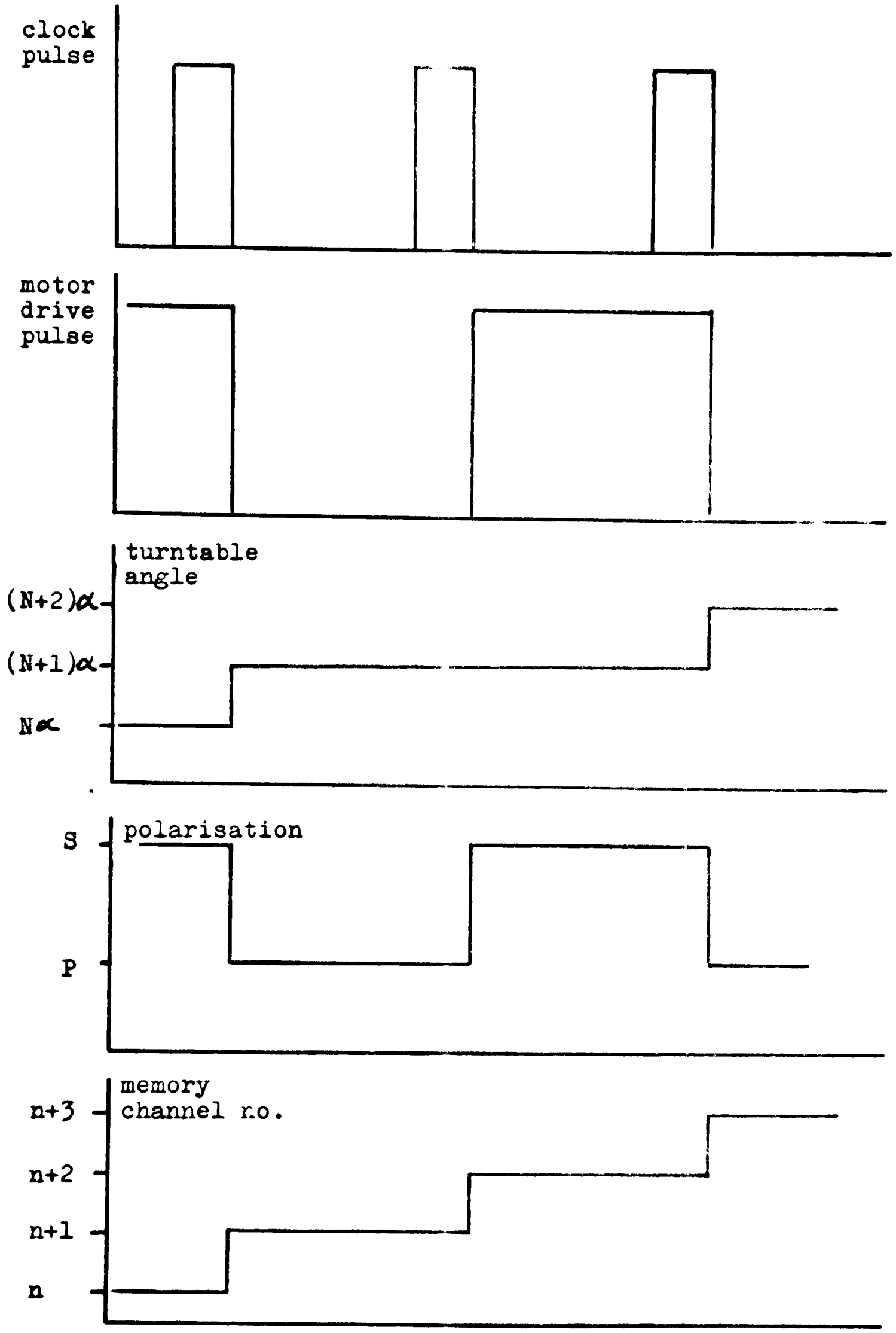

clock
pulse
motor
drive
pulse
turntable
angle
(N+2)α
(N+1)α
Nα
S
P
polarisation
memory
channel no.
n+3
n+2
n+1
n

Fig. 4.6

The system operates as follows. (This should be read in conjunction with figure 4.6, which shows the switching sequence).

The pulse generator provides a train of clock pulses which are applied to the input of a flip-flop. The square wave output from this drives the stepping motor and the polariser; thus the polariser spends half the step period in each state.

The output of the flip-flop also drives an electronic counter, which can be preset to give a pulse after any number of steps. This pulse is used to start the MCS which is clocked by the pulse generator each time the polariser changes state.

The result is that the response of the grating in each polarisation at successive wavelengths is stored in alternate channels of the MCS. A small error is incurred in this system because the polariser takes longer to change state than the MCS.

The MCS channel number is incremented in 1μS, whereas the polariser changes from S to P in 5-10 mS, and during this period a mixture of polarisations is transmitted.

The effect of this error is reduced by ensuring that the total counting period is much longer than the transition period of the polariser. Clock rates of typically 0.5 Hz were used.

4.1.2 (d) Alignment and Operation

Alignment of the system was carried out as follows:

(i) The condenser and collimator were adjusted to give a uniformly illuminated parallel beam and the grating was mounted so that the grooves were parallel to the axis of rotation.

ii) The zero order image was set coincident with the entrance slit and the grating was tilted about a horizontal axis until the reflected light passes through the auxiliary prism. The exit slit was positioned so that the image of the entrance slit is sharply focussed on the slit jaws.

iii) The grating was rotated through an angle equal to half that required between the incident and diffracted beams and the exit slit placed so that the light passes through it (If a large angle between the beams was required a system using two lenses was used. The principles of alignment remain the same).

When the system was aligned, measurements were taken as follows:

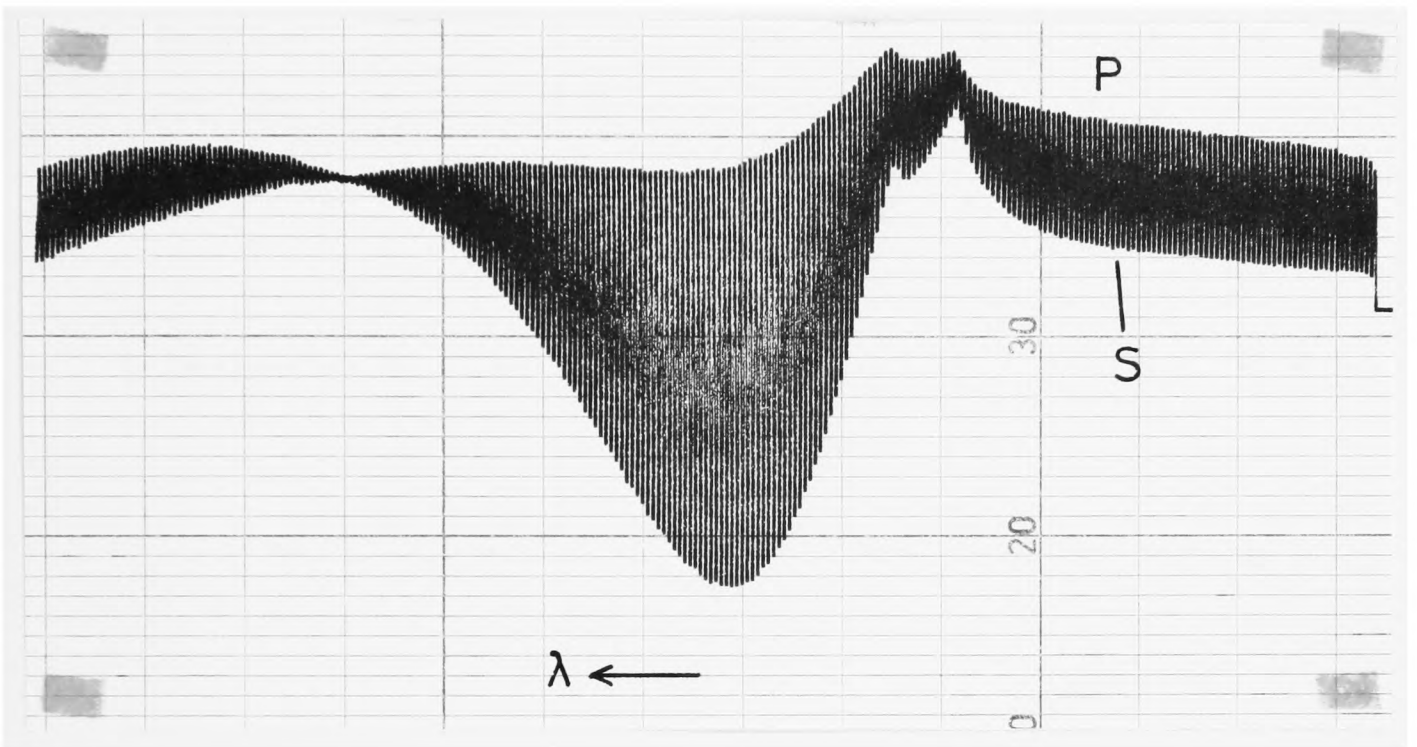
i) The region of interest was scanned rapidly and the gain of the detector was set so that no overloading occurred in either polarisation.

ii) The grating was rotated until zero order fell on the exit slit (which was closed by a shutter to avoid overloading) and the preset counter was set for the desired initial step number.

iii) The turntable was started at a fairly high step rate. As the preset step number was approached (as observed by a subsidiary electromagnetic counter) the step rate was reduced to that required for the scan (typically 0.5 Hz) When the preset number was reached, the MCS started to count.

iv) When the scan was complete, the contents of the MCS were read out onto a teletype for analysis.

Figure 4.7, which was obtained by connecting the output of the detector to a chart recorder, shows graphically the format of the results.



4.1.2 (e) Wavelength Calibration

Since all the measurements are referred to zero order, the results are automatically calibrated in terms of grating rotation.

The formula connecting channel number with grating position is

$$\theta = \frac{P + N/2}{48} \text{ degrees}$$

where

θ is the rotation from the zero order position

P is the preset initial step number

N is the channel number (N/2 rounded down to the integer value)

Hence the wavelength is

$$\frac{\lambda}{d} = \frac{2}{n} \sin \left(\frac{(P + N/2)}{48}\right) \cos \alpha /2$$

(see appendix 2 for definition of α).

This shows that the wavelength calibration depends only on the measurement of d.

The state of polarisation of a particular reading is determined by whether N is odd or even.

4.1.2 (f) Efficiency Calibration

The response of the system is given by

$$V(\lambda) = A.\eta_n(\lambda) \cdot S(\lambda) \cdot R(\lambda) \cdot P(\lambda)$$

where

V is the detector output voltage

η_n is the nth order efficiency of the grating

S is the powerspectrum of the source

R is the combined response of the optical components

P is the spectral response of the photomultiplier and

A is a calibration constant

Hence a single measurement of $V(\lambda)$ does not give sufficient information to obtain a value of $\eta_n(\lambda)$ and we need to be able to measure the system response $F(\lambda)$ ($F(\lambda) = S(\lambda)R(\lambda)P(\lambda)$) or in same way eliminate its effect.

The simplest way of eliminating $F(\lambda)$ depends on the fact that in the anomalous reginns, the efficiency of the grating is strongly polarisation dependent. Hence if the ratio of the detector voltages is taken for each wavelength, the spectral response $A.F(\lambda)$ is cancelled out, leaving only the ratio of the efficiencies. An example of this treatment is shown in figure 4.8.

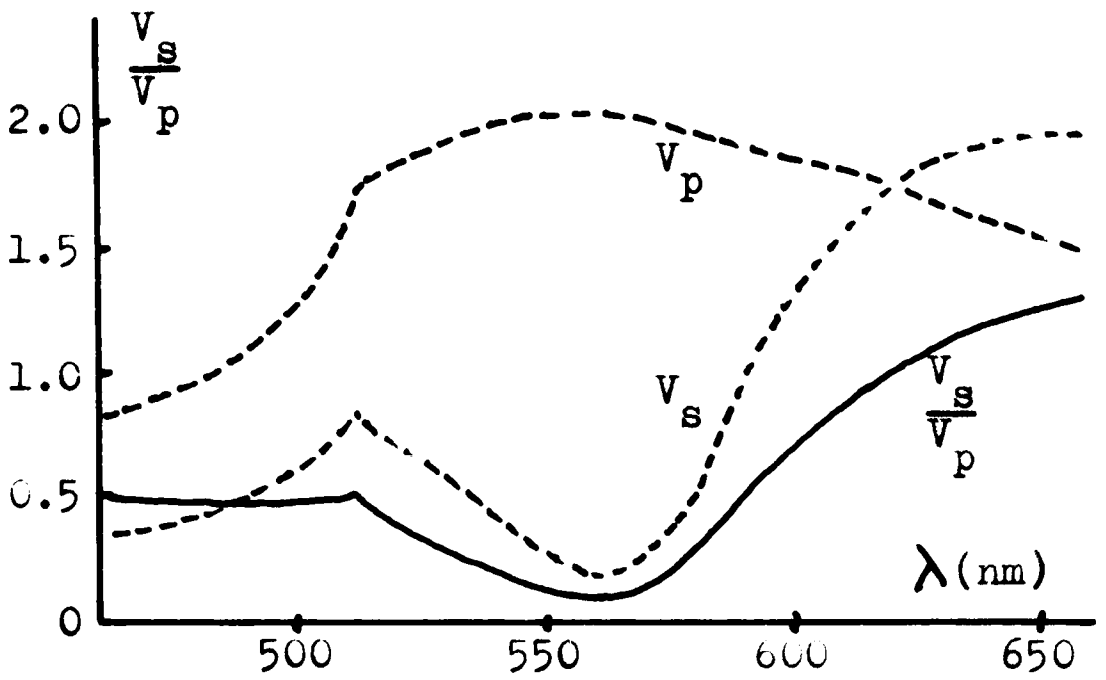


Fig. 4.8

To obtain the actual values of efficiency, however the instrumental response and the calibration constant must be measured. The spectral response is measured as follows:

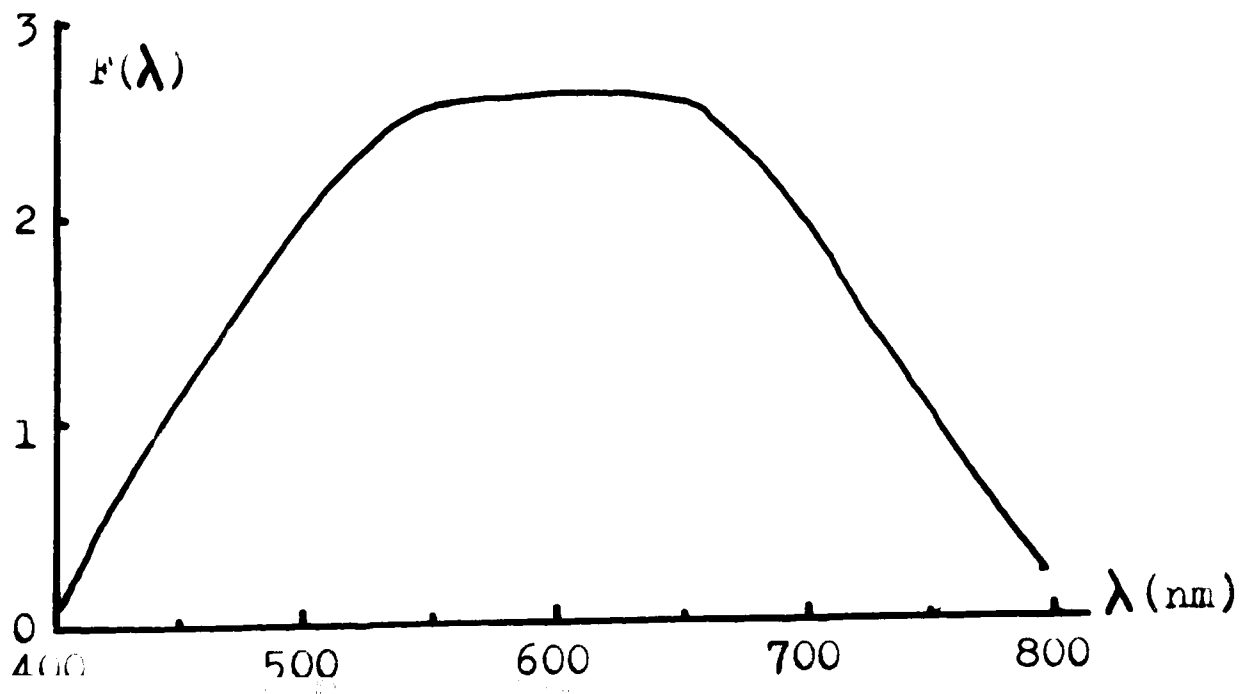
i) The absolute Littrow efficiency $\eta(\lambda)$, of a particular grating is measured at 50 nm intervals using an auxiliary monochromator.

ii) The response, $V(\lambda)$, for this grating is measured.

iii) At each wavelength the response of the system may be calculated from

$$F(\lambda) = \frac{V(\lambda)}{\eta(\lambda)}$$

The response of the system used here is shown in figure 4.9.



To find the absolute efficiency of any grating, a single absolute measurement is made at a wavelength λ_o, using either a laser or an auxiliary monochromator. The efficiency at any other wavelength may now be found from

$$\eta_n = \frac{V(\lambda)}{F(\lambda)} \cdot \frac{F(\lambda_o)}{V(\lambda_o)} \cdot \eta_n(\lambda_o)$$

An example of this is shown in figure 4.10

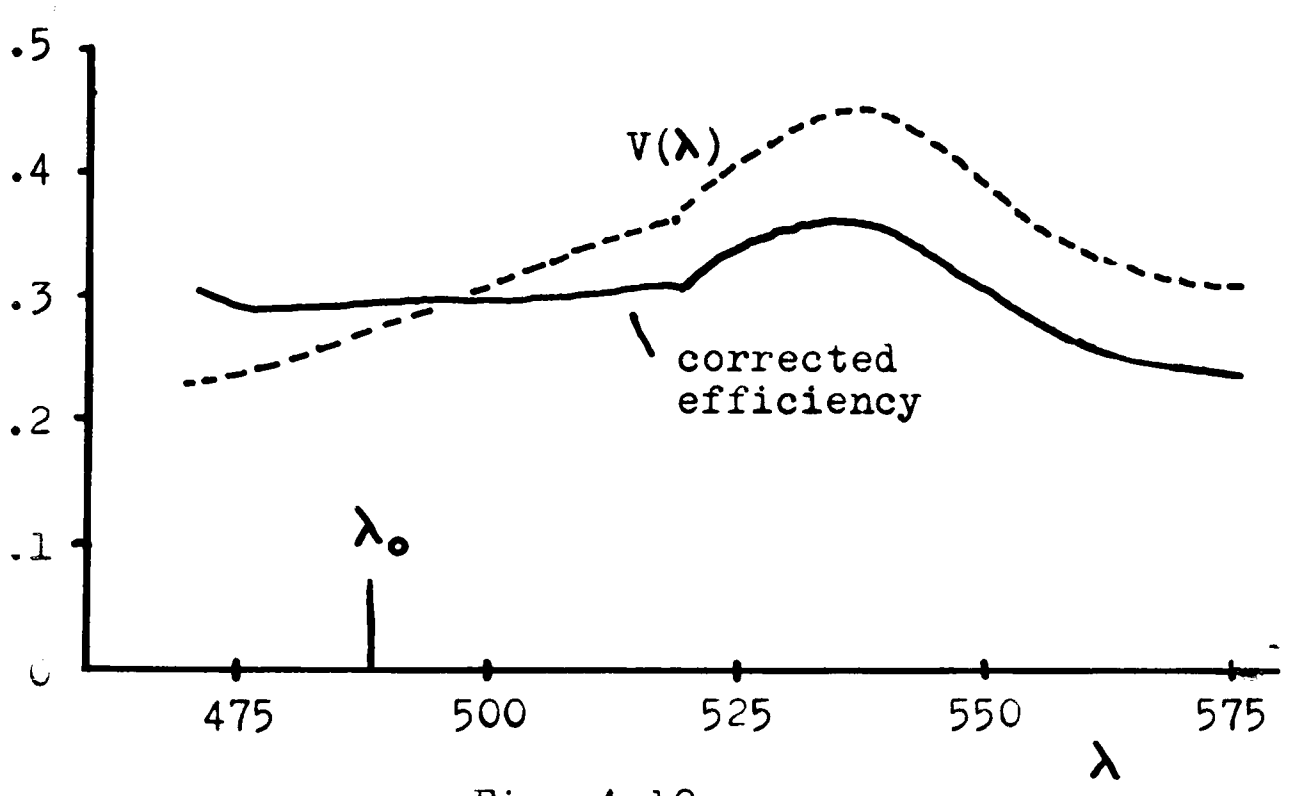


Fig. 4.10

It should be noted that the accuracy of the system is highest when $F(\lambda)$ is high (i.e. between 500 and 700 nm). This is because the signal to noise ratio is highest in this range. In addition, the response is varying slowly with wavelength in this region, which reduces the effect of calibration errors.

In view of this, the configurations were chosen to produce anomalies in this range if possible.

4.1.3 Zero Order Efficiency

4.1.3(a) Apparatus

Measurements of zero order efficiency are particularly convenient, as no motion of the grating is required during the scan.

The apparatus uses the light source described in figure 4.4, but now the zero order is intercepted by a collector mounted on an arm pivoted about the axis of the turntable.

This consists of a short focal length lens with a pinhole at the focus. A fibre optic cable is mounted behind the pinhole and the light passes along this to the entrance slit of a monochromator (Spex "Minimate").

The output of the monochromator is measured by a photomultiplier and p.s.d. and the output voltage is displayed on a chart recorder. The monochromator is scanned by a synchronous motor and gearbox and wavelength calibration is provided by manual operation of a marker pen on the chart recorder every 5 or 10 nm.

4.1.3(b) Operation

i) To measure the incident intensity, the detector was reversed to intercept the incident beam and adjusted for maximum response. The spectral region of interest was scanned.

ii) The grating was mounted and the turntable rotated to give the desired angle of incidence. The detector was moved to intercept the zero order.

iii) The region of interest was scanned once more, using the same detector gain.

The efficiency was calculated by dividing the zero order intensity by the incident intensity at each wavelength of interest.

4.2 The Measurement of Scattered Light

Scattered light in the context of diffraction gratings may be defined as any of the radiation leaving the grating which does not travel in the direction of one of the diffracted orders.

Apart from its deleterious effect in high resolution spectroscopy, it was found during the course of this work that the properties of the scattered light provide useful information concerning the process taking place at the surface of the grating.

The features of interest are:

i) The variation of the total flux of scattered light with angle of incidence at constant wavelength

ii) The angular distribution of scattered light at constant wavelength.

4.2.1 The total flux of scattered light as a function of angle of incidence

The absolute flux of the total scattered light is difficult to measure, but a figure proportional to this may be obtained by using a detector with a large acceptance angle, to reduce the effect of any changes of angular distribution.

The way in which this was done is shown in figure 4.11. The configuration is basically that of figure 4.1, using the argon ion laser as a light source. A fibre optic cable is used as a light collector, since these have a large cone of acceptance (typically 30^{o} half angle). The flux collected by the cable is further increased by the use of a horizontal diffusing screen mounted below the incident beam. Light scattered downward by the grating is scattered back by the diffuser in the direction of the fibre.

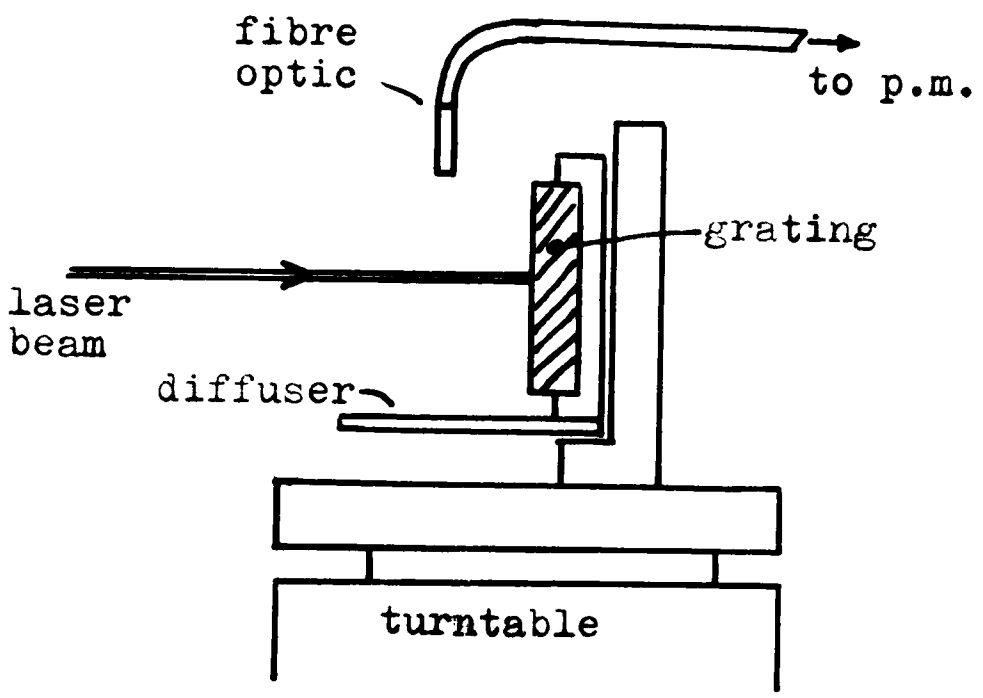


Fig. 4.11

The fibre optic and the diffuser are fixed relative to the grating and rotate with it. The light passes down the fibre to a phobomultiplier and p.s.d. system. The level of scattered light from a HDG is low, even using an intense laser, and so to reduce the effects of noise, a multichannel scaler is used to store the signal at each angle of incidence. Long counting times are used to integrate the signal. The clock pulses are passed through a divider before reaching the MCS, so that each channel can correspond to 1, 2, 4, 8 or 16 steps. Since there are 512 channels and each step is 1/48 degree, the 16 step/channel setting enables a range of 170° to be scanned. Figure 4.12 shows the variation of scattered light from a HDG, with angle of incidence for two orthogonal polarisations.

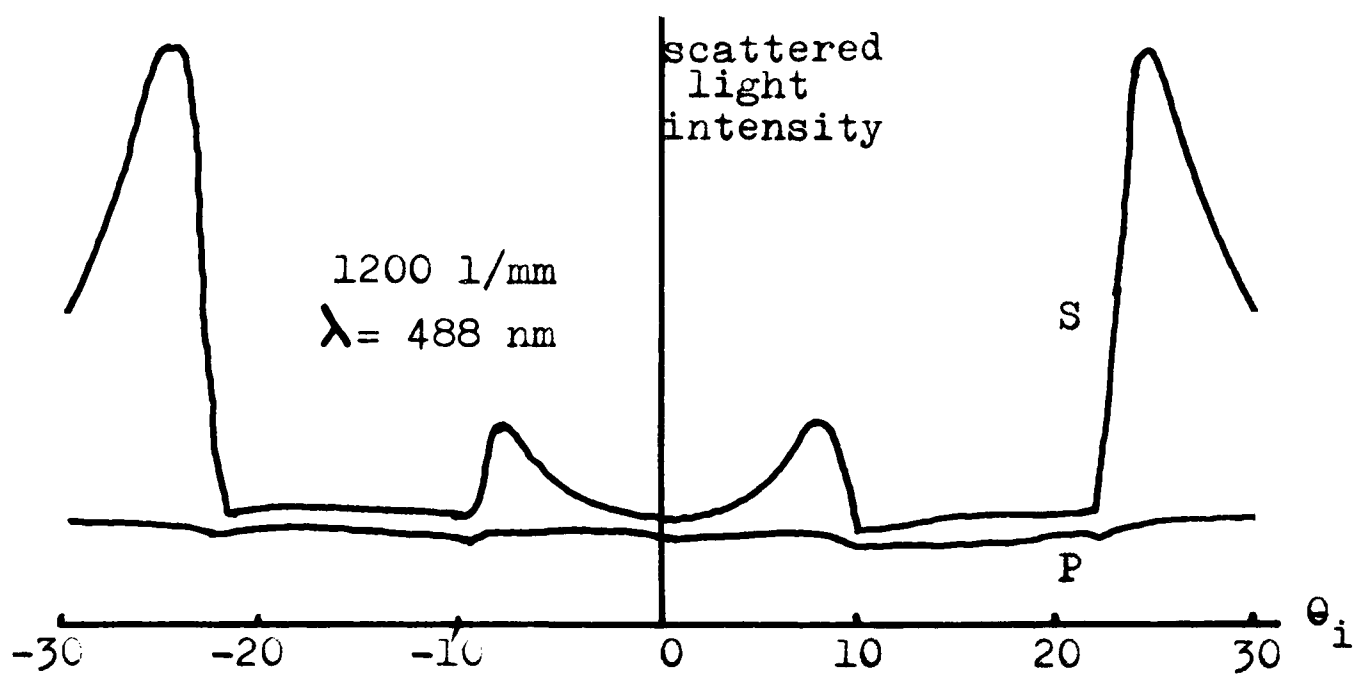


Fig. 4.12

4.2.2 Angular distribution of scattered light

The intensity of scattered light is measured as a function of angle from the grating normal (θ) and angle of rotation about the normal (ø). The apparatus is shown in figure 4.13.

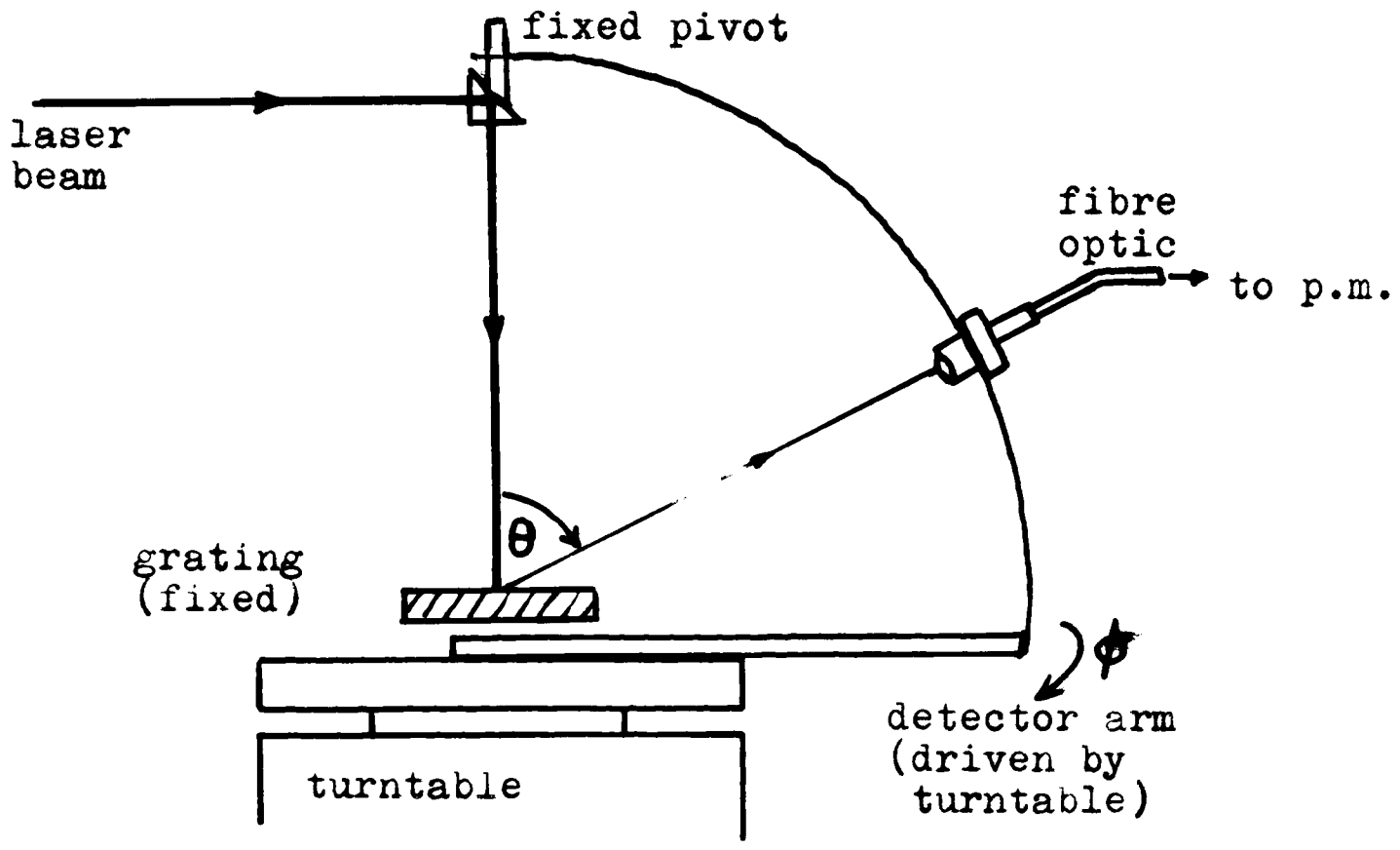


Fig. 4.13

The grating is held horizontally and is fixed relative to the incident laser beam.

The detector is that described in chapter 4.1.3(a), but now the lens can be moved along a circular track which is centred on the point where the laser beam strikes the grating. Thus the angle that the scattered light makes with the grating normal (θ) can be fixed.

The complete detector unit can be rotated about an axis through the grating normal by means of the turntable driven by the stepping motor. In this way, the variation of scattered light as a function of ϕ at constant θ may be determined.

Because the scattered light intensity is very weak, a multichannel scalar and long integration times are used, as described in 4.2.1.

A typical result is shown in figure 4.14.

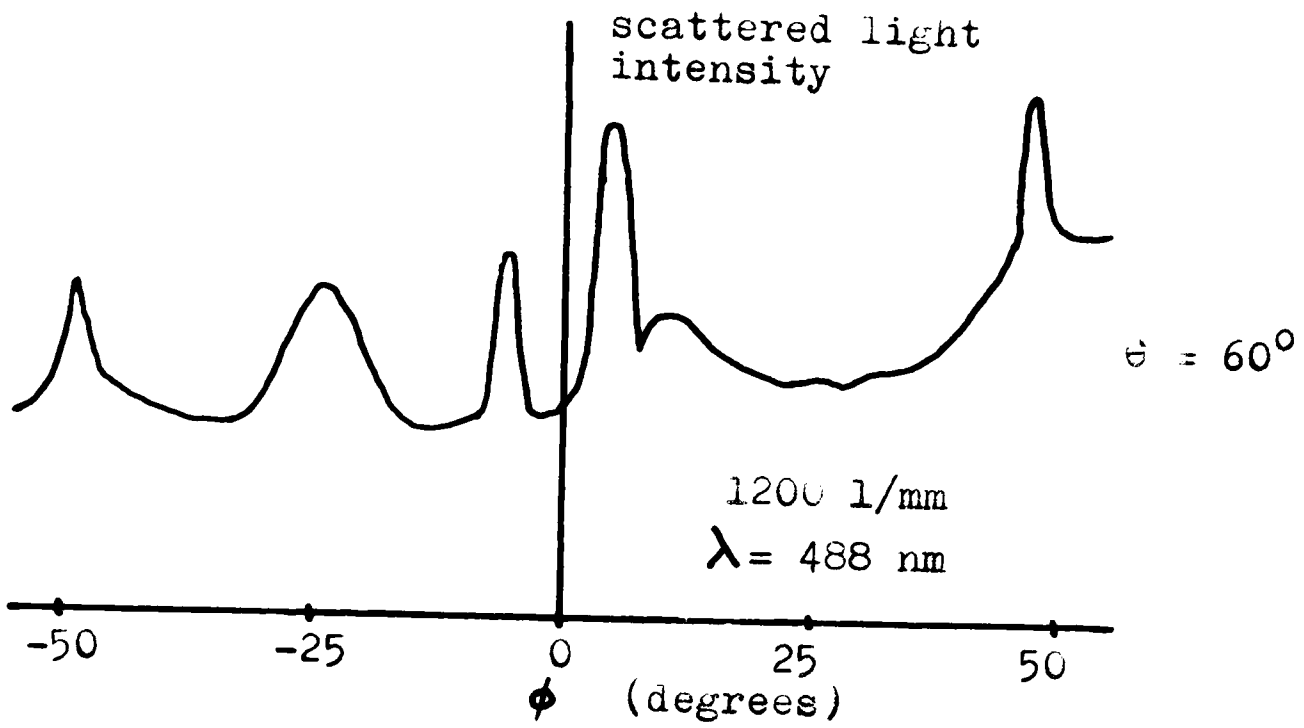


Fig. 4.14

Chapter 5

THE EFFICIENCY OF HOLOGRAPHIC DIFFRACTION GRATINGS

5.1 Wood's Anomalies

5.2 Scalar theory

5.3 Electromagnetic theory

5.3.1 Infinite conductivity theory

5.3.2 Finite conductivity theory

The efficiency of a diffraction grating when used in a particular spectroscopic configuration is determined largely by the groove profile and to a smaller extent by the metal in which it is formed.

In recent years, much theoretical work has been directed towards a full explanation of the dependence of efficiency on groove profile, and this has now been brought to the stage where the observed variation of efficiency may be predicted with some accuracy.

In this chapter, measurement of the efficiency of HDG's under various conditions are compared with theoretical predictions taken from various sources in the literature.

We first discuss the formation and properties of Wood's anomalies, which form a sensitive test of grating theory. We then proceed to examine the validity of the simplest theory of grating efficiency.

This is followed by a consideration of rigo rous electromagnetic theory. We shall find that for a complete description of the performance of diffraction gratings. the optical properties of the metal must be taken into consideration.

In this study, HDG's are particularly convenient for the following reasons:

i) They are easily made with any groove spacing and groove depth;

ii) The sinusoidal profile is described by only two parameters;

iii) Anomalous effects are much stronger than in ruled gratings, which facilitates their observation. This is because the blaze of a ruled grating masks the effect of the anomaly.

5.1 Wood's Anomalies

Anomalies were first observed by Wood (25), who noticed that the continuum spectrum produced by a grating spectrograph was crossed by bright and dark bands at certain wavelengths.

These were soon explained qualitatively by Rayleigh (28) who showed that the rapid variation of efficiency was caused by the redistribution of energy following the disappearance of one of the orders other than the order of observation. On the basis of this, it was shown that an anomaly would occur when

$$\frac{\lambda_R}{d} = \frac{1}{m} \quad (\sin\theta_i \pm 1)$$

where λ_R is the wavelength of the anomaly (the Rayleigh wavelength)

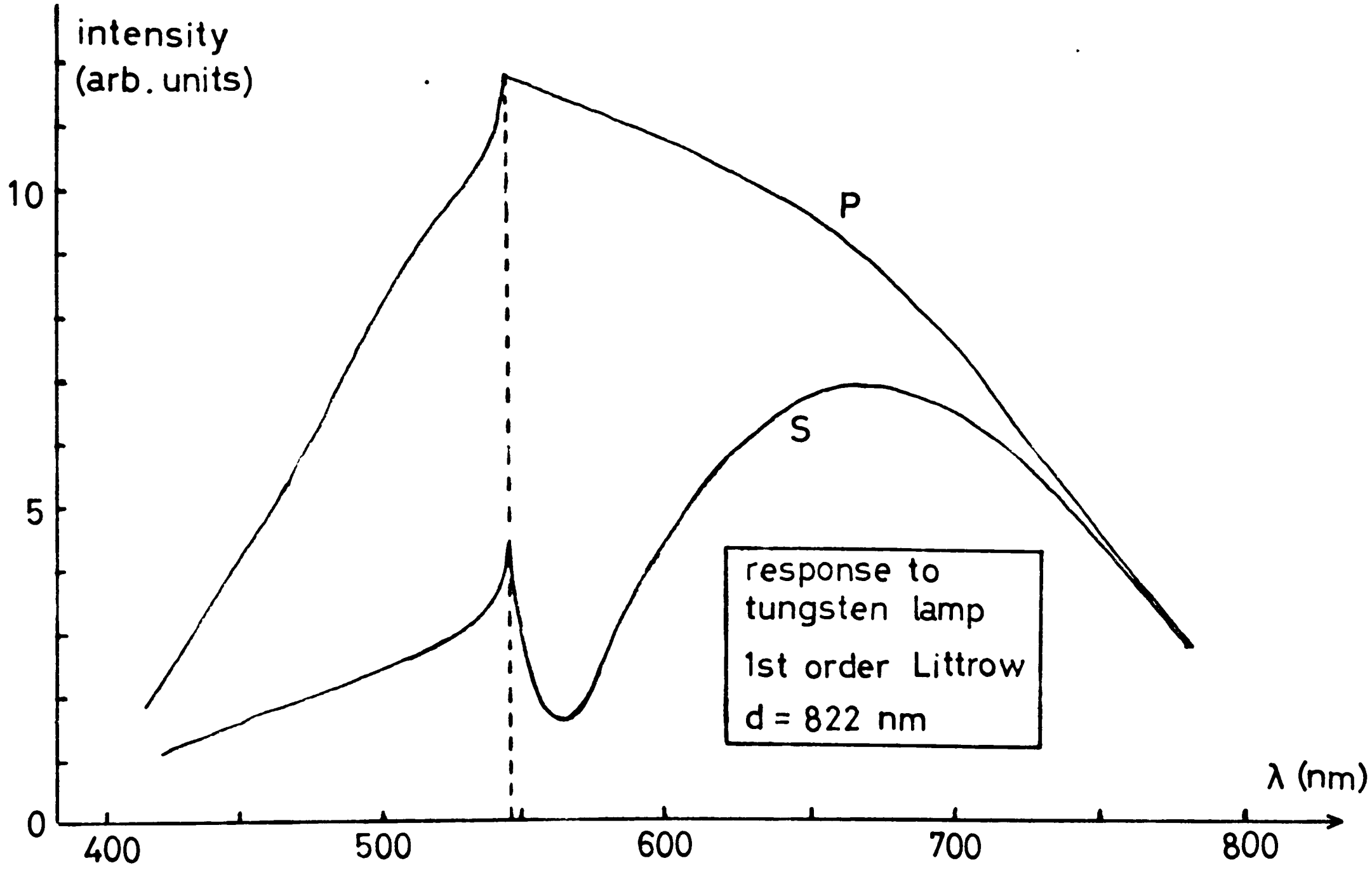
d is the groove spacing

θ_i is the angle of incidence

m is the number of the order passing off.

Detailed observation of the shape of the anomaly showed that this equation predicted accurately the positions of the bright bands in the spectrum, but no explanation was found for the associated dark anomalies.

Anomalies were also found to be strongly polarisation dependent. The largest variation was seen to occur in the S polarisation (E perpendicular to the grooves). In the orthogonal P polarisation, a much smaller effect was observed. In particular, the dark anomaly was absent


intensity
(arb. units)
10
5
0
P
S
response to
tungsten lamp
1st order Littrow
d = 822 nm
λ (nm)
400
500
600
700
800

in the P polarisation. . (Figure 5.1)

The detailed structure of the S anomaly is a sensitive test of the accuracy of grating theories.

5.2 Scalar theory

The simplest approach to any problem involving diffraction is to assume that the amplitude of the radiation field may be represented by a scalar function of position and time. Then Kirchhoff's integral approach may be used, and under certain conditions the amplitude of the diffracted wave may be shown to be proportional to the Fourier transform of the amplitude transmission function of the object (29).

The conditions under which this approach is valid are:

i) the path difference between the incident and diffracted wavefronts should vary linearly with the position co-ordinates on the grating;

ii) the wavelength of the light should be small compared with the dimensions of the object. (In this case the groove spacing).

In this section, we investigate the effect of departing from the second condition.

In appendix 1, it is shown that for a perfectly reflecting sinusoidal grating, the efficiency in nth order is given by

$$\eta_n = J_n^2 \left(\frac{2\pi h}{\lambda}\right)$$

where h is the peak to peak groove depth. In Figure 5.2,

the efficiencies of HDG's are plotted for various values of λ/d. The solid bar represents the measured values and the open bar is the Scalar theory prediction.

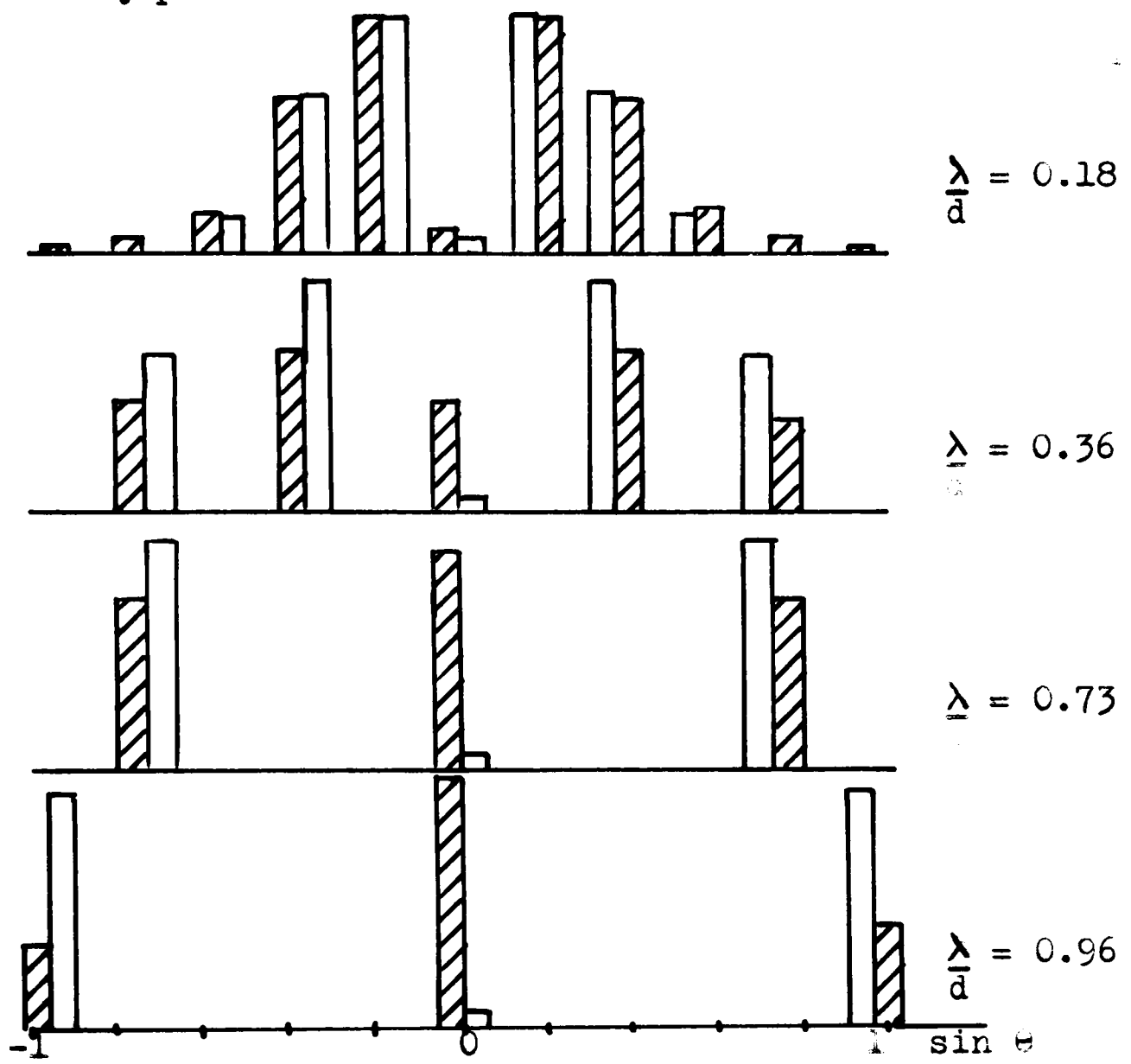


Fig. 5.2

It is seen that Scalar theory breaks down as the number of orders decreases, with a corresponding increase in the intensity of the highest orders.

Following Rayleigh's explanation of anomalies, we would expect that for low values of λ/d, where the intensity of the orders passing off is very small, anomalous effects will be absent. On the other hand, when the intensity of the grazing orders reaches a significant level, anomalous effects will become increasingly important.

In figure 5.3, the zero order efficiency of gratings with the same groove depth is plotted as a function of wavelength for groove spacings ranging from 2.5μ to 0.4μ.

It is seen that although the agreement for the 400 line/mm grating is good, the remainder of the series show anomalous effects not predicted by Scalar theory. (Scalar theory predicts that the gratings would all show the same efficiency).

Thus we may summarise the conditions for scalar theory to be applicable as follows:

i) The ratio of λ/d must be $\lesssim 0.25$ and the high orders must be of negligible intensity.

ii) The metal must be a perfect conductor at the wavelength of interest. (Thus in the u.v., for example, where condition (i) may be satisfied, care must be taken, since the metal may absorb strongly in this region.)

As an example of the use of scalar theory in its region of applicability, figure 5.4 shows the zero order effiency of a series of 300 line/mm gratings of increasing groove depth.

The points represent the measured values and the curves were obtained by fitting the scalar theory expression to them. The value of h found from these curves was used to confirm the linearity of the development process.

The good agreement of the fitted curves indicates that the profile of the HDG's is well described by a sinusoidal function. (The variation of h with development time has been plotted in figure 3.7).

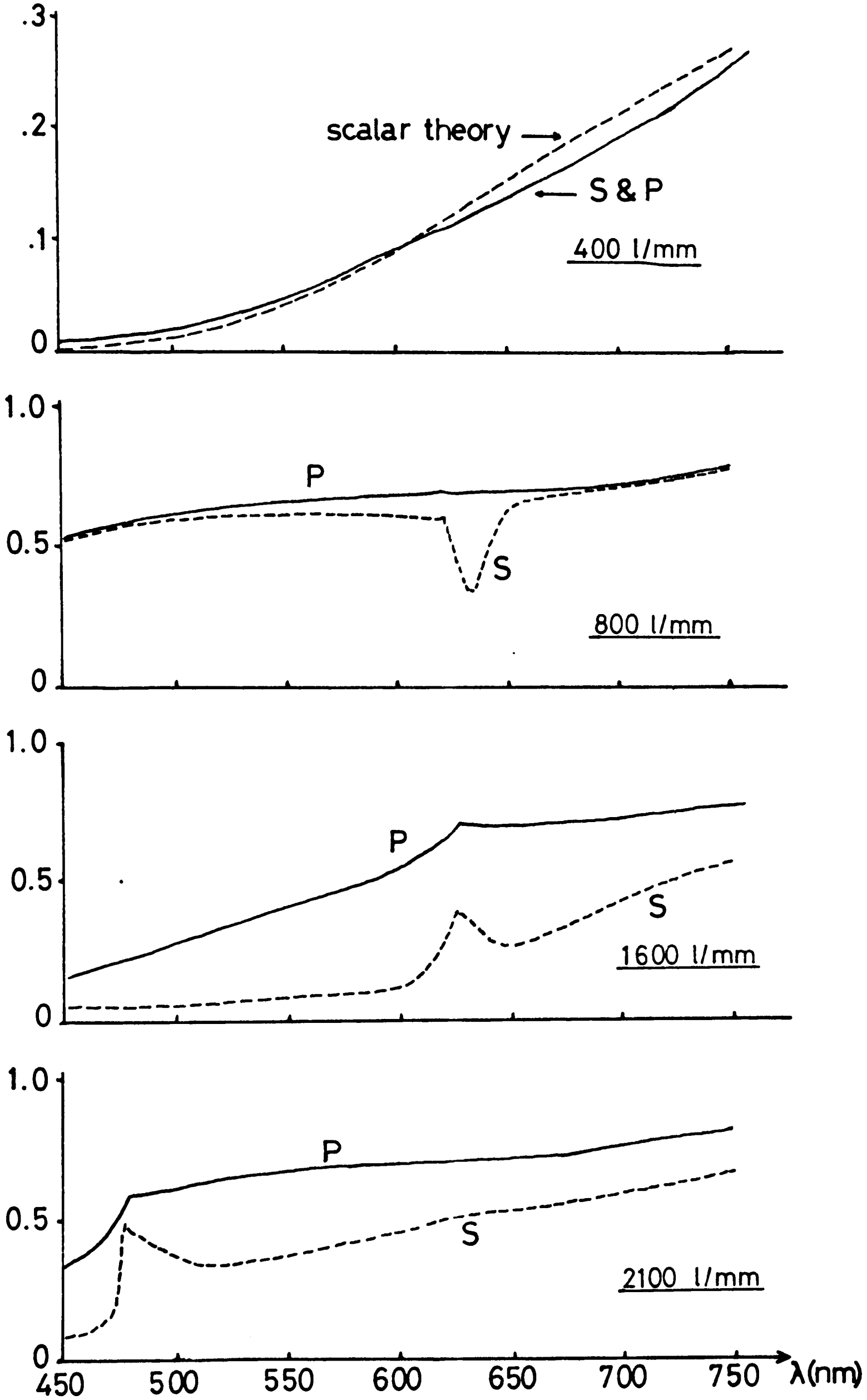

.3
.2
.1
0
scalar theory
S & P
400 l/mm
1.0
0.5
0
P
S
800 l/mm
1.0
0.5
0
P
S
1600 l/mm
1.0
0.5
0
P
S
2100 l/mm
450
500
550
600
650
700
750
λ(nm)

Fig. 5.3

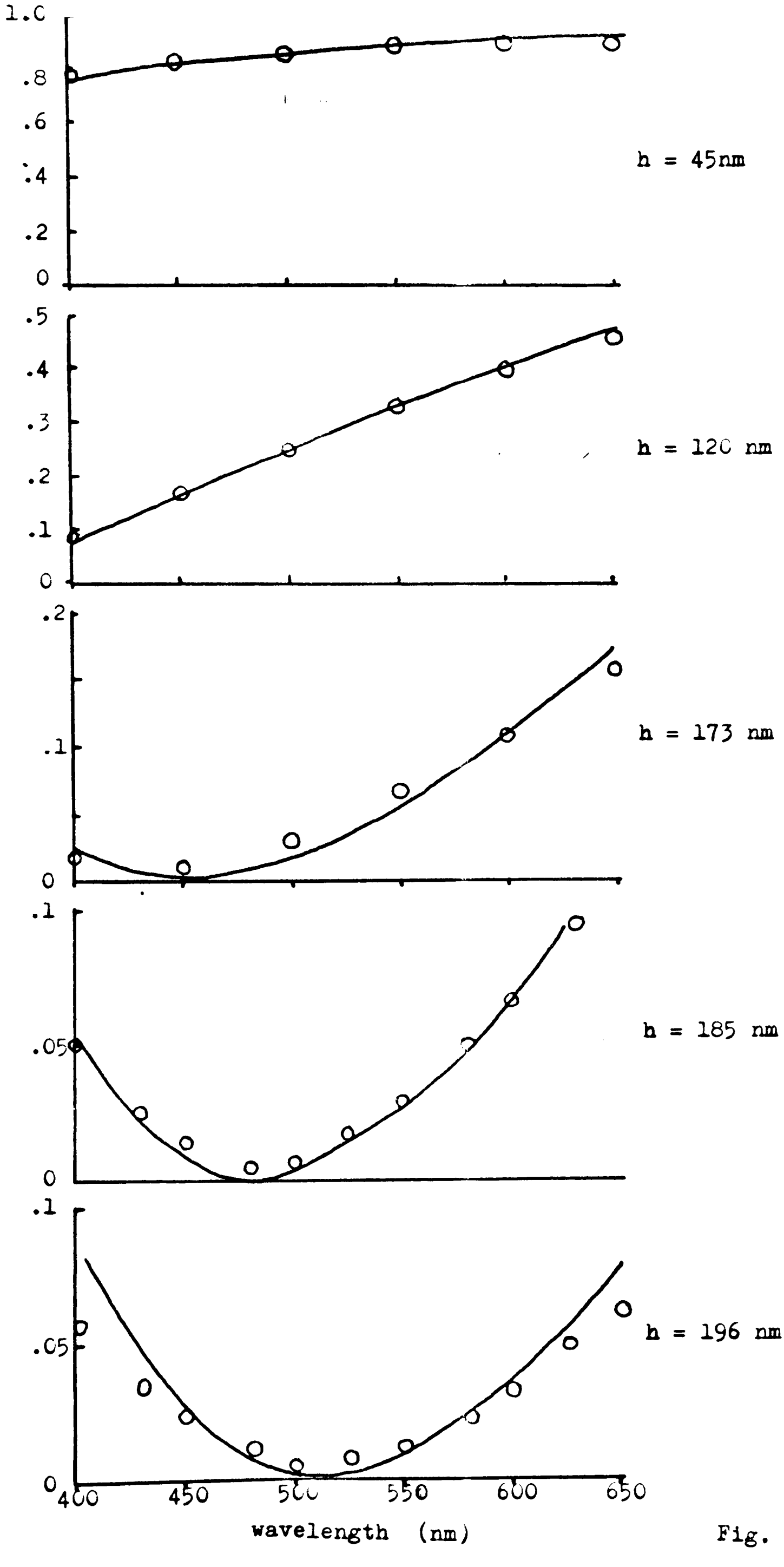

1.0
.8
.6
.4
.2
0
h = 45nm
.5
.4
.3
.2
.1
0
h = 120 nm
.2
.1
0
h = 173 nm
.1
.05
0
h = 185 nm
.1
.05
0
400
450
500
550
600
650
wavelength (nm)

Fig. 5.4

5.3 Electromagnetic Theory

The attempts to obtain a rigorous solution to Maxwell's equations with periodic boundary conditions and thereby solve the grating problem may be divided into two groups: infinite conductivity and finite conductivity theories.

In the former, it is assumed, that the metal is a perfect conductor ($\sigma = \infty$), and then various mathematical techniques are used to reduce Maxwell's equations to a soluble form. For a review of these techniques up to 1960 see (30).

In finite conductivity theories, the optical properties of the metal are taken into account and numerical methods are used to solve the equations. This work has only become possible with the development of digital computers.

The techniques and potentials of this theory are outlined in (31).

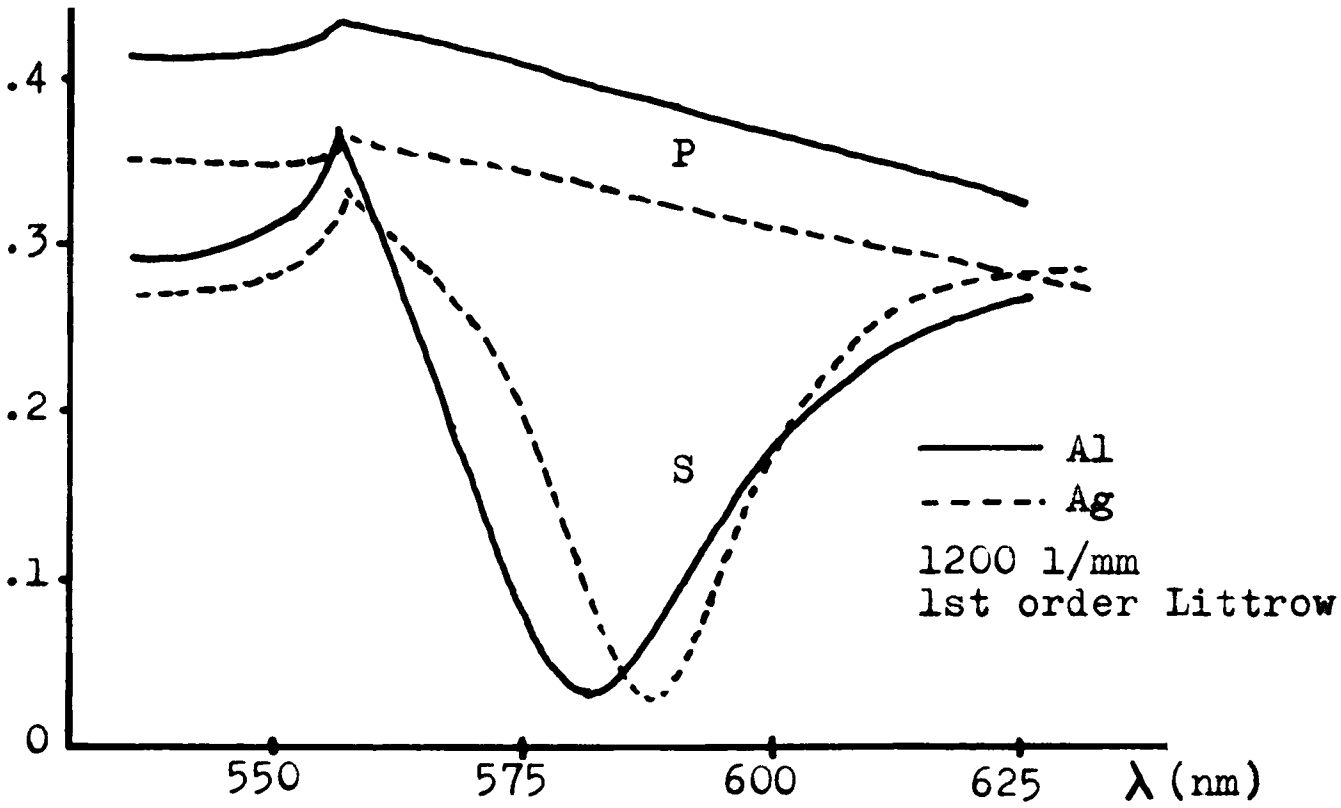
5.3.1 Infinite conductivity theory

The majority of infinite conductivity methods were developed before HDG's had become feasible and sinusoidal profiles were experimentally unobtainable. As a consequence there are very few detailed calculations published of the kind which were carried out for echellettes.

In this section, therefore, rather than make detailed comparisons between theory and experiment, we shall examine some of the general faults of infinite conductivity theory.

The ways in which infinite conductivity theory breaks down are discussed below.

i) The efficiency is predicted to be independent of the metal in which the grating is ruled. This is obvious, since the theory takes no account of metal properties, but it can be an important failing as seen in figure 5.5, which shows the measured 1st order efficiency of a 1200 l/mm grating coated successively with Al and Ag.



It is seen that in the region of the S anomaly, there is a large discrepancy between the curves which cannot be accounted for simply by a difference of reflectivity.

ii) Infinite conductivity theory predicts that the efficiency depends only on the normalised wavelength, λ/d, and not on the actual wavelength. The error involved in this is shown in figure 5.6, in which the Littrow efficiencies of 1360 and 1025 line/mm gratings with the same groove depth are plotted on a normalised scale.

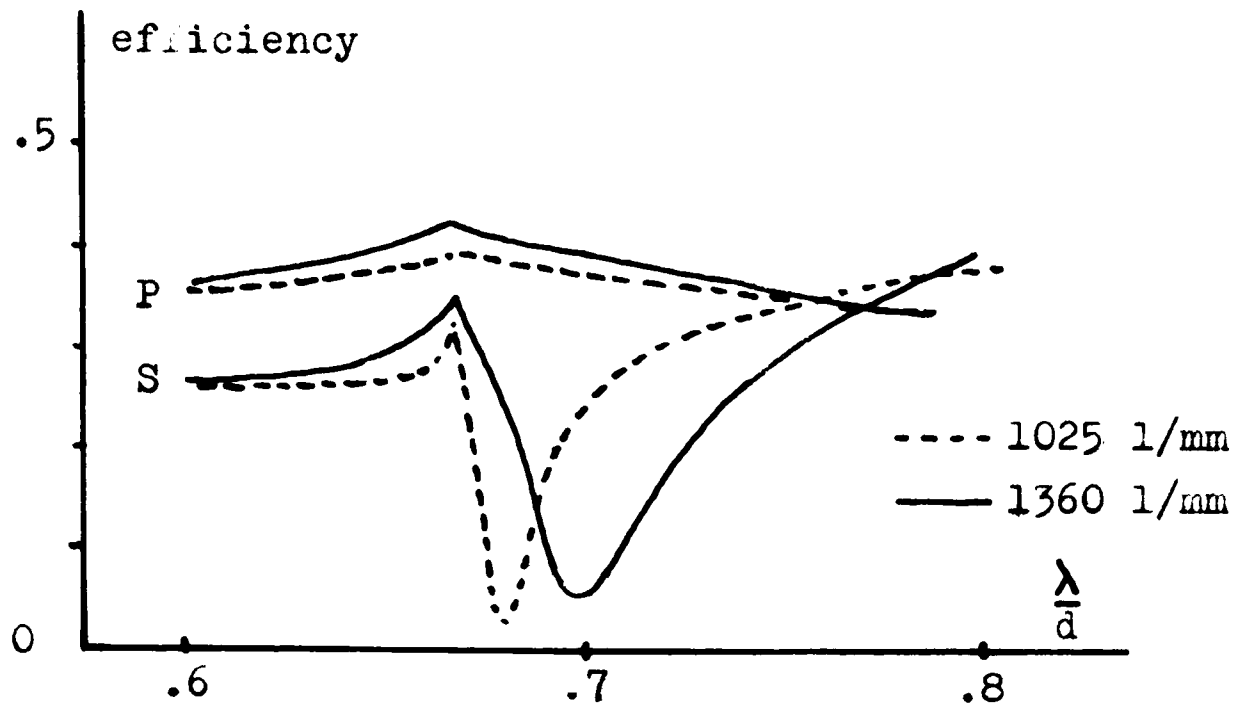


Fig. 5.6

The curves show that the efficiency variation in the region of the S anomaly differs. The P curves are similar.

iii) The efficiencies predicted by infinite conductivity theory are generally higher than the observed efficiencies and the discrepancy is greater than can be accounted for by the lower reflectivity of the surface.

This is shown in figure 5.7, in which a theoretical curve from (32) is plotted together with the measured efficiency in the equivalent situation.

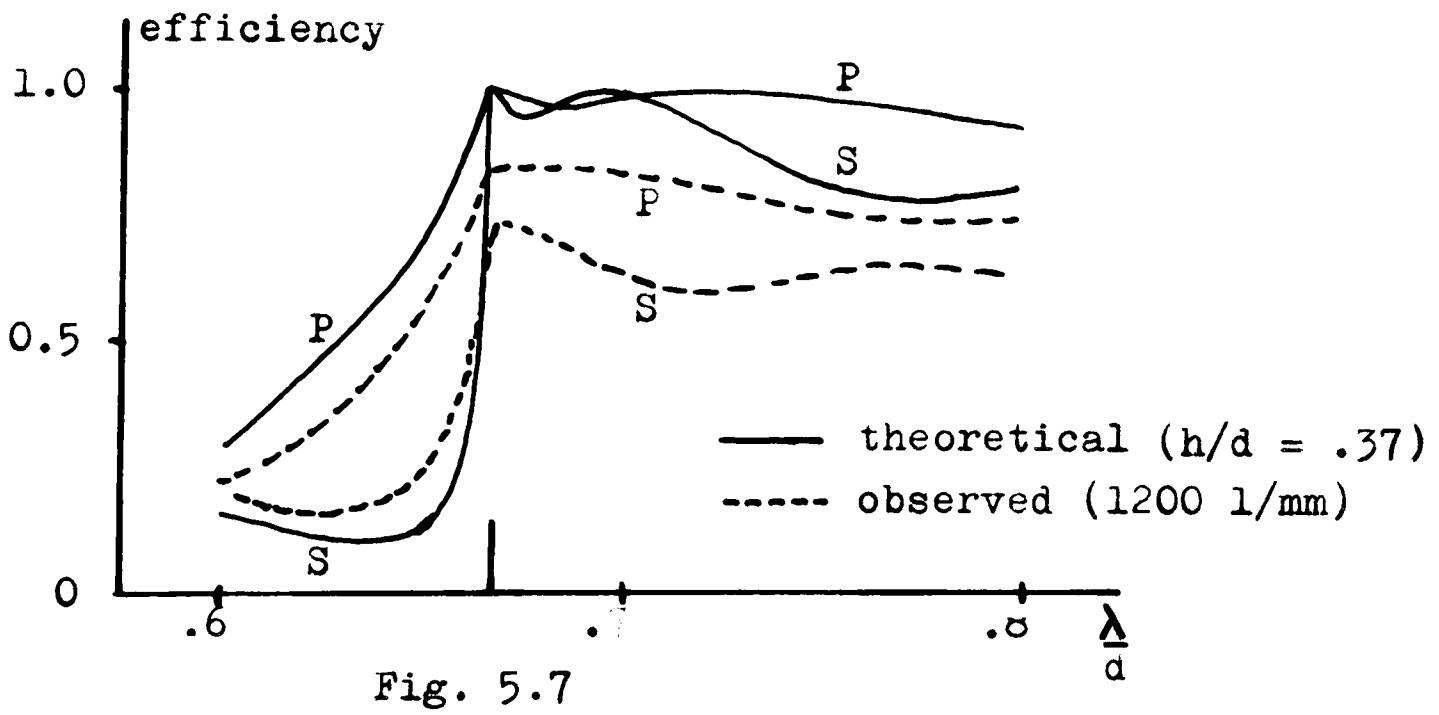


Fig. 5.7

The theoretical curve reaches 100% at the Rayleigh wavelength, but the maximum efficiency recorded by the grating is only of the order of 70%. This difference cannot be accounted for by the finite reflectivity of aluminium, which at this wavelength is 0.95.

Thus we see that infinite conductivity theory is unable to account for some of the observed behaviour of real gratings and that the discrepancy is greatest in the region of the S polarised Wood's anomalies.

5.3.2 Finite conductivity theory

In recent years, the problem of diffraction by a grating in a metal of finite conductivity has been approached, largely by the Marseilles group under Professor R.Petit. The results of this work have been so successful that all the observed behaviour of the efficiency of diffraction gratings in the visible region may be explained.

It is even suggested (33) that the theory may be used in reverse, to estimate the groove depths of HDG's from the observed efficiency curves.

In this section we present experimentally determined efficiency curves in confirmation of published theoretical predictions for HDG's. It is found that agreement is close, even in the region of the S anomalies. It is likely that any discrepancies are due to uncertainties in the knowledge of the groove profile, since it has been found theoretically that the efficiency curves are very sensitive to small deviations from a true sinusoidal profile (33).

The fact that finite conductivity theories are necessary to explain the behaviour of diffraction gratings means that the metal cannot be considered perfect even in the visible region.

In the next chapter we examine the physical basis of the interaction between the metal and the radiation field.

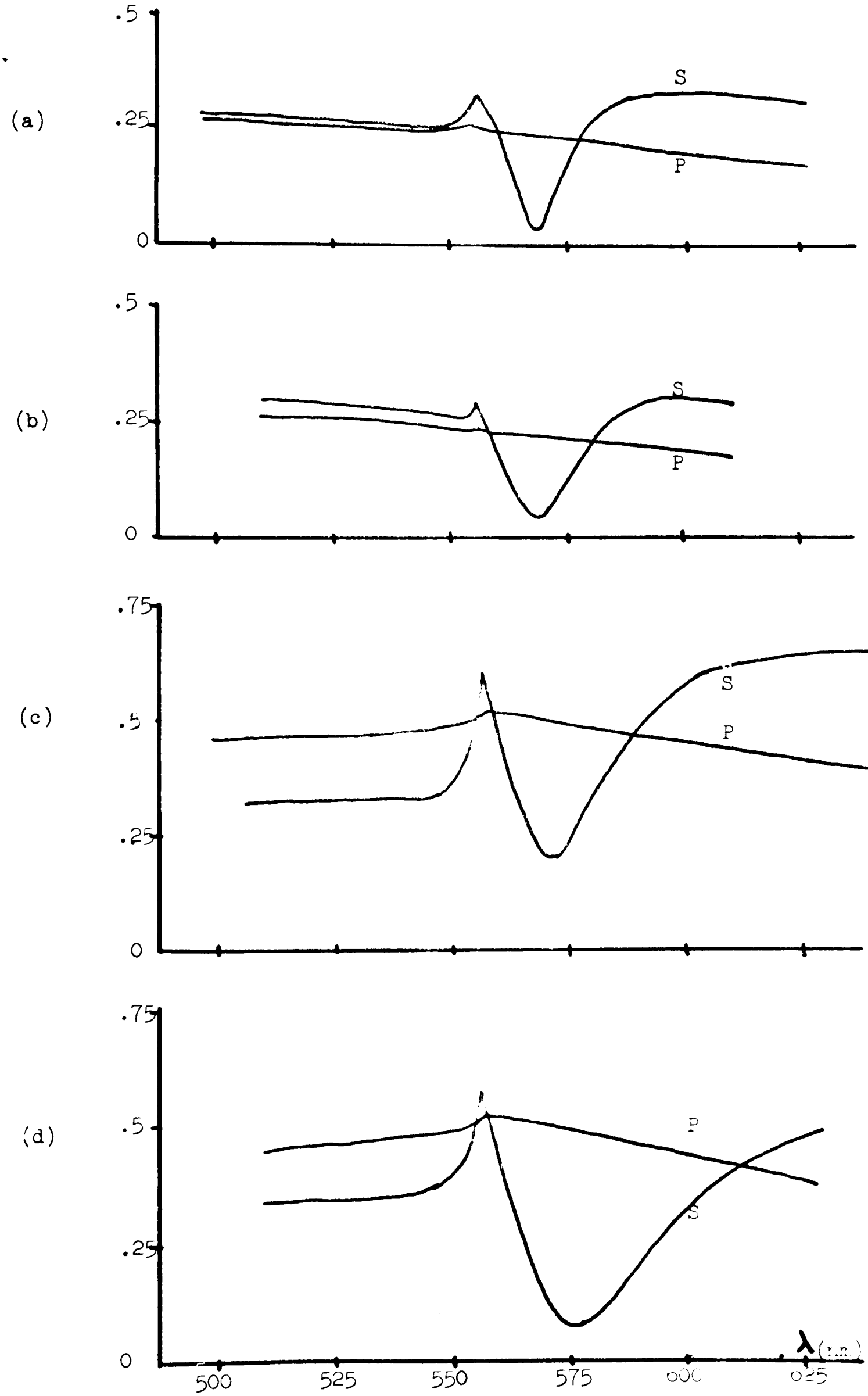


Fig. 5.8 (cont'd overleaf)

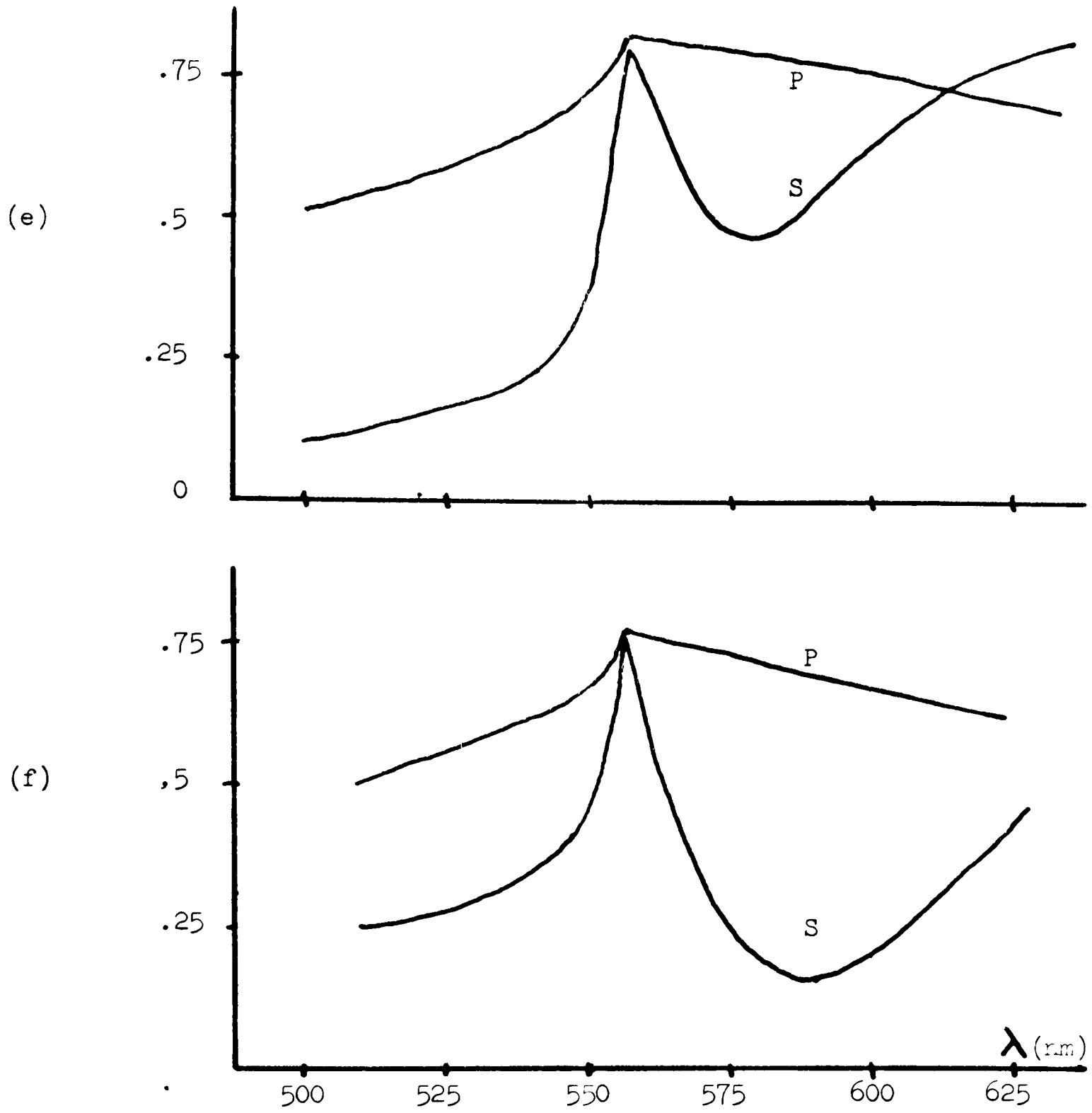


Fig. 5.8: Comparison between theory and experiment 1200 l/mm, Al coated. (theoretical curves taken from (30))

(a) h = 100 nm theory; (b) h = 100 nm expt.
(c) h = 156 nm " ; (d) h = 150 nm "

(e) h = 220 nm " ; (f) h = 200 nm "

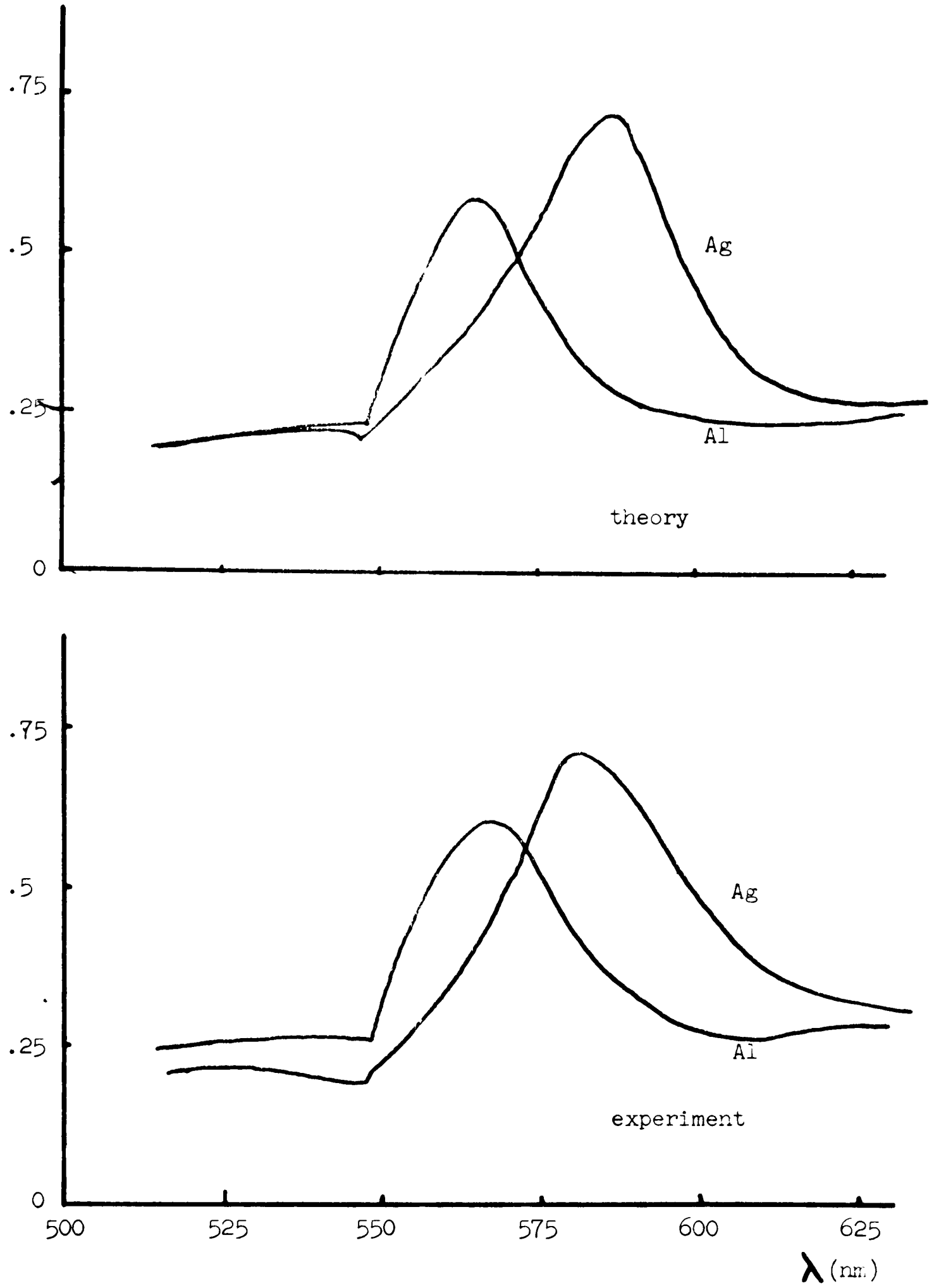


Fig. 5.9 :Comparison between theory and experiment.
770 l/mm; S polarisation; Al & Ag coated. h = 220nm
(Theoretical curves from (36))

Chapter 6

SURFACE PLASMON EFFECTS IN DIFFRACTION GRATINGS

6.1 The Surface Plasmon Resonance

6.1.1 Surface plasmons

6.1.2 Surface plasmons on periodic surfaces

6.2 Surface plasmons and efficiency

6.2.1 Efficiency at constant wavelength

6.2.2 Efficiency in Littrow mountings

6.3 Surface plasmons and scattered light

6.3.1 The total intensity of scattered light

6.3.2 The angular distribution of scattered light

The results of the previous chapter have shown that a treatment of the properties of diffraction gratings cannot be complete unless the optical properties of the metal coating are taken into account.

To do this with any degree of rigour is a complicated task requiring advanced mathematical techniques, but a simpler understanding of the process taking place may be obtained using the concept of surface plasmons (SP's).

This chapter will describe the properties of SP's on periodic surfaces and show how they influence the efficiency and scattered light of diffraction gratings. The final section shows how the properties of SP's, and hence the dielectric properties of the metal may be inferred from measurements on diffraction gratings.

6.1 The Surface Plasmon Resonance

6.1.1 Surface Plasmons

It is well known that the conduction electrons of a metal may be treated as a plasma and that this plasma will support bulk oscillations at a single frequency known as the plasma frequency ω_p. If the electrons are assumed to move in a constant potential (i.e. they are completely free), the value of ω_p may be shown (34) to be

$$\omega_p = \left(\frac{4\pi ne^2}{m}\right)^{\frac{1}{2}}$$

where n = the number density of the conduction electrons

and e and m are the electronic charge and mass respectively.

For most metals, this frequency corresponds to light in the u.v. region of the spectrum. It is possible to observe effects due to bulk plasma oscillations in the reflection spectra of metals of these wavelengths (35).

In addition to bulk oscillations, the plasma is also capable of supporting longitudinal waves localised near the surface of the metal; these are known as surface plasmons (SP's). A review of SP's and their place in solid state physics may be found in (35).

Unlike the bulk plasmon, which has a single frequency, the dispersion relation of an SP is given by

$$k = \Gamma(\omega)\frac{\omega}{c}$$

where $\hbar k$ is the momentum of the SP

ω is the frequency

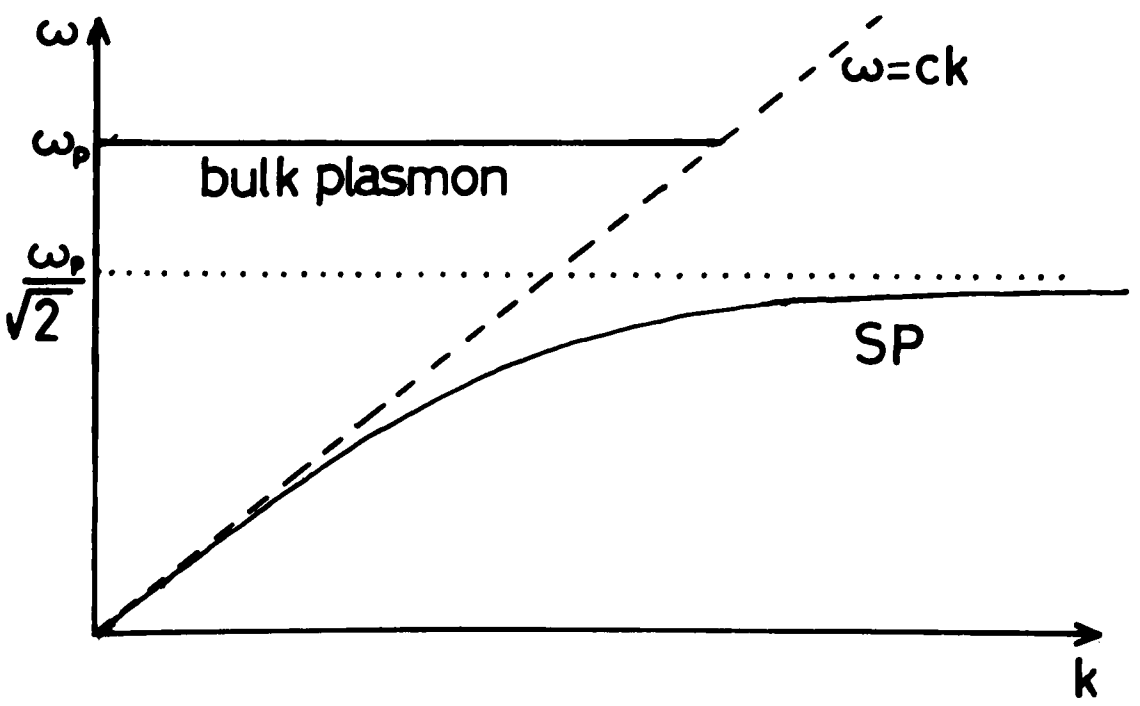
and $\Gamma(\omega)$ is a factor slightly greater than unity

For a free electron metal with a plane surface, this factor is given by (37)

$$\Gamma(\omega) = \left(\frac{\epsilon(\omega)}{1+\epsilon(\omega)}\right)^{\frac{1}{2}}$$

where $\epsilon(\omega)$ is the real part of the dielectric constant of the metal.

The dispersion relations for the two modes of oscillation are shown in figure 6.1 together with the photon dispersion curve, $\omega = ck$



It is seen that the frequency of the SP tends to $\omega_p/\sqrt{2}$ at large values of k and is always less than that of a photon with the same wavenumber. Thus in the absence of any other effects, an interaction between a photon and an SP cannot take place because energy and momentum cannot be conserved simultaneously.

6.1.2 Surface plasmons on periodic surfaces

In 1967 (38), it was shown that if the surface of the metal is not perfectly plane, the fluctuations of mean electron density are sufficient to cause a coupling between the SP and the incident radiation.

This may be understood as follows: If the electron density is periodic as in a reflection grating, then phase modulation of the incident radiation field will introduce momentum components parallel to the surface given by

$$k = k_i + n.\frac{2\pi}{d}$$

where k_i is the parallel component of the incident wavevector

n is an integer

and d is the period of the surface

When the momentum of the photon, modified by the modulation effect, matches that of the SP of the same frequency, then an interaction will take place between the two fields.

Thus the conditions to be satisfied are:

$$\omega_i = \omega_{sp} \quad \ldots\ldots \quad 6(i)$$

$$\vec{k_i} = -\delta_p \vec{k_{sp}} + \frac{\vec{2\pi}.p}{d} \quad \ldots\ldots\ldots 6(ii)$$

where ω_i is the frequency of the photon

ω_{sp} is the frequency of the SP

k_i is the parallel component of the photon wavevector

k_{sp} is the SP wavevector

p is a positive or negative integer

and $\delta_p = \begin{cases} +1, & p > 0 \\ -1, & p < 0 \end{cases}$

In the vector equation 6(ii), $2\pi\vec{p}/d$ is always normal to the grating grooves, and so if the photon is incident in a plane normal to the grooves the equation may be written in scalar form:

$$\frac{2\pi \sin\theta}{\lambda} = -\delta_p \, k_{sp} + \frac{2\pi p}{d} \qquad \ldots\ldots \quad 6(iii)$$

where λ is the wavelength of the incident photon

θ is the angle of incidence

Now in section 6.1.1, the value of k_{sp} was stated to be

$$k_{sp} = \Gamma(\omega) . \frac{\omega}{c}$$

thus since ω is the frequency of the incident photon, we may write

$$k_{sp} = \Gamma(\lambda) . \frac{2\pi}{\lambda}$$

and 6(iii) becomes

$$\sin\theta = -\delta_p \Gamma(\lambda) + p . \frac{\lambda}{d} \qquad \ldots. \quad 6(iv)$$

When this condition is satisfied, and when the incident electric field has a component perpendicular to the grooves, a surface plasmon resonance occurs.

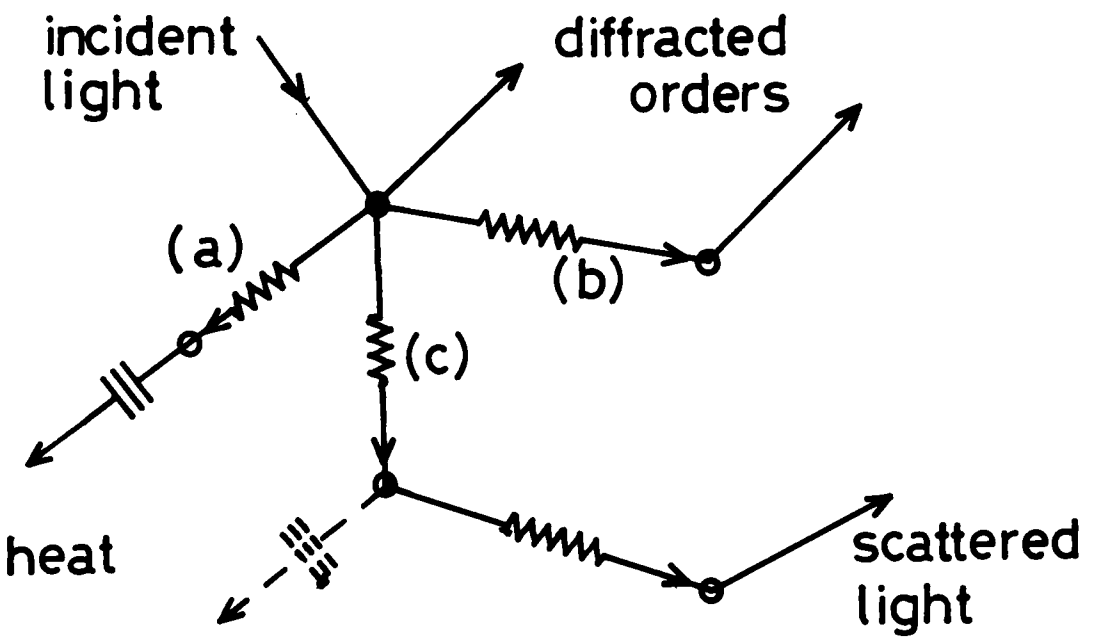
The effect of this is that SP's are created by the incident light and energy is absorbed from the radiation field into the SP field.

The similarity between 6(iv) and the expression in 5.1 for the Rayleigh wavelength shows that the SP resonance is closely associated with anomalous effects.

Since $\Gamma(\lambda)$ is slightly greater than 1, the resonance occurs just after the corresponding order has passed off the grating. This suggests that SP's are closely related to evanescent waves on the grating (39).

6.1.3 Surface plasmon processes in diffraction gratings

When the SP resonance occurs, large numbers of SP's are created. The ways in which this affects the performance of the grating are shown schematically in figure 6.2.



This diagram shows the ways in which the energy in the incident light is dissipated when a SP resonance occurs. A fraction of the incident energy is diffracted in the usual way. This includes light scattered by irregularities in the grating. The rest of the energy is absorbed in three ways:

a) The SP decays by electronic collisions etc. This leads eventually to Joule heating of the metal and the energy is lost from the radiation field.

b) The SP may decay to give a photon. This is the reverse of the process of creation, and it may be shown (see section 6.2.1) that if the SP is propagating with its initial direction and momentum the decay photon will propagate in one of the diffracted orders.

c) The SP may be scattered (e.g. from a surface irregularity or a photon) before it decays. In this case, the decay photon will contribute to the scattered light from the grating. If the SP scattering is inelastic, then a fraction of the energy will be dissipated as Joule heating.

On the basis of the foregoing description of SP processes, the following effects may be predicted in association with SP resonances.

i) The total efficiency of the grating (defined as the ratio of the total flux in the diffracted orders to the flux in the incident beam) will be reduced. This is due to the absorption of energy by the grating (process (a)) and to a smaller extent to the scattering of light out of the orders (process (c)).

ii) The incident energy will be redistributed among the orders.

When the plasmons decay via process (b), the intensity of each order depends on the relative phase and amplitude of the SP contribution to that order. It is found that this tends to concentrate the energy into a single order.

iii) The scattered light from the grating will increase.

As a result of process (c), the large increase in the number of SP's which occurs at resonance will be reflected in a corresponding increase in the scattered light level. The decaying SP's must also satisfy the momentum conservation condition, so it will be expected that the scattered light will have a well-defined angular distribution.

iv) These effects will be polarisation dependant.

In order for the radiation field to couple with the longitudinal electron waves, it is necessary for the radiation to have a component of its electric vector normal to the grooves of the grating. Thus SP effects are absent in P polarisation.

In the rest of this chapter, these effects will be examined in detail.

6.2 Surface plasmon effects and efficiency

6.2.1 Efficiency at constant wavelength

6.2.1(a) Total efficiency

The significance of the effects of SP resonances on the performance of a diffraction grating may be judged from the variation of the total efficiency with angle of incidence at constant wavelength. This is shown in figure 6.3 for a 1200 line/mm grating at a wavelength of 457.9 nm.

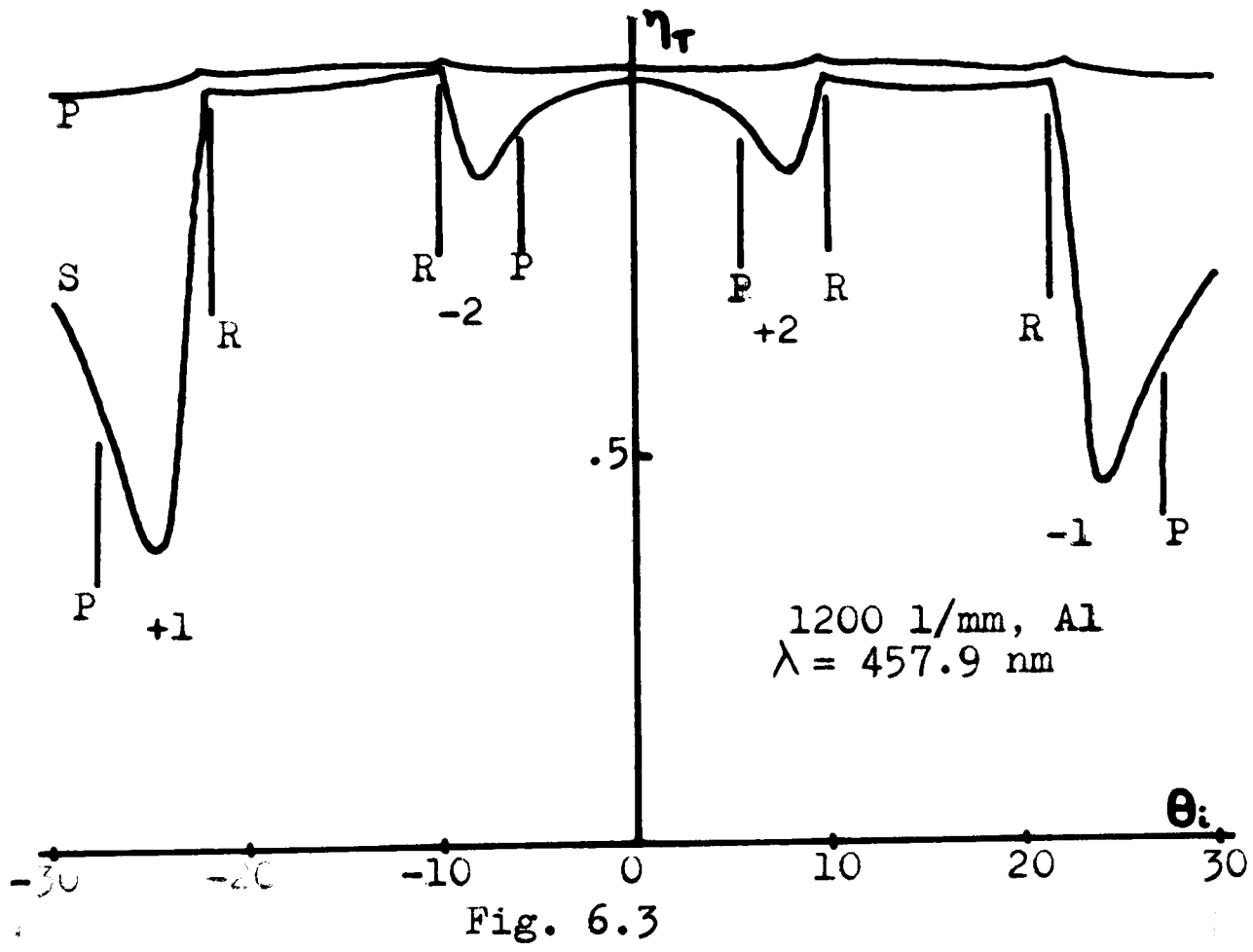


Fig. 6.3

The angles marked "R" are the Rayleigh angles at which an order passes off the grating. The angles marked "P" are the SP resonance angles calculated from equation 6(iv) using a value of $\Gamma(\lambda)$ calculated from free electron theory.

It is seen that at these angles a substantial fraction (as much as 50% in this case) of the energy in the S polarised incident beam is absorbed by the grating. No corresponding effect is seen in P polarisation. This large absorption of energy by the grating at the resonance shows that the metal cannot be considered to be a perfect conductor in this region.

This explains why infinite conductivity theories are unable to explain accurately the shape of the S anomaly.

6.2.1(b) The redistribution of light among the orders

In section 6.1.3 it was stated that a SP decaying with its original momentum produces a photon travelling in one of the diffracted orders. This will now be proved.

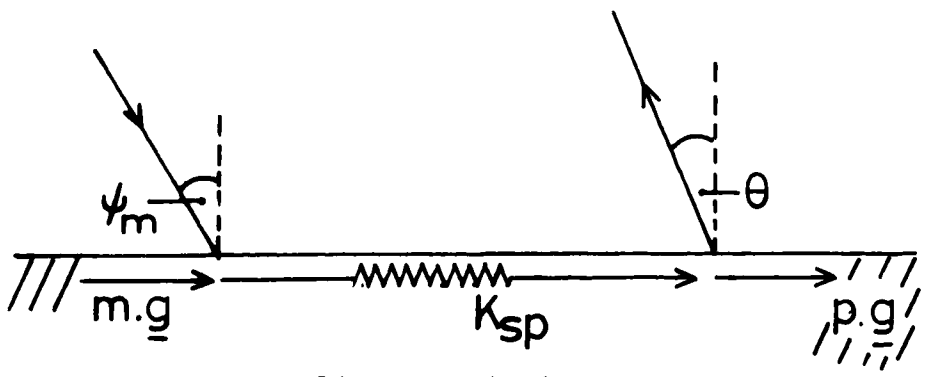


figure 6.4

In figure 6.4, a photon of wavelength λ incident at the resonance angle ψ_m creates a SP with wavevector k_{sp}. This subsequently decays to give a photon propagating at θ to the normal.

Since energy is conserved at each interaction, the frequency of the decay photon is equal to that of the incident photon. Momentum is also conserved, so we may write:

$$\frac{2\pi \sin\psi_m}{\lambda} + m\frac{2\pi}{d} = k_{sp} \quad \ldots\ldots \text{creation}$$

$$-\frac{2\pi \sin\theta}{\lambda} + p.\frac{2\pi}{d} = k_{sp} \quad \ldots\ldots \text{decay}$$

where m and p are the orders of the interactions.

From these equations we obtain

$$\sin\theta + \sin\psi_m = -(m + p).\ \frac{\lambda}{d}$$

This is the grating equation and shows that θ is the direction of the nth diffracted order, where

$$n = -(m + p)$$

It should be noted that directions of the orders and the SP decay only coincide at resonance. As the angle of incidence moves away from ψ_m, the positions of the orders change, but the direction of the SP decay photon, which depends only on the SP wavenumbers, remains fixed.

The effect which the SP decay photons have on the diffracted orders depends on the amplitude and phase of the SP relative to the order and the coupling between the various evanescent orders.

In general an enhancement of one or more orders at the expense of the others is observed, but it is not possible on the basis of this simple model to predict the effects accurately. The effects of resonant decay are shown in figure 6.5, which shows the variation of efficiency with angle of incidence in all orders of a 1200 line/mm grating at 514.5 nm (S polarisation).

The strongest features are the sharp peaks in zero order caused by the $m = \pm 1$ resonances at $\mp 24.5^{\circ}$. Peaks also occur in the first orders corresponding to the $m = \pm 2$ resonances at $\pm 12.3^{\circ}$.

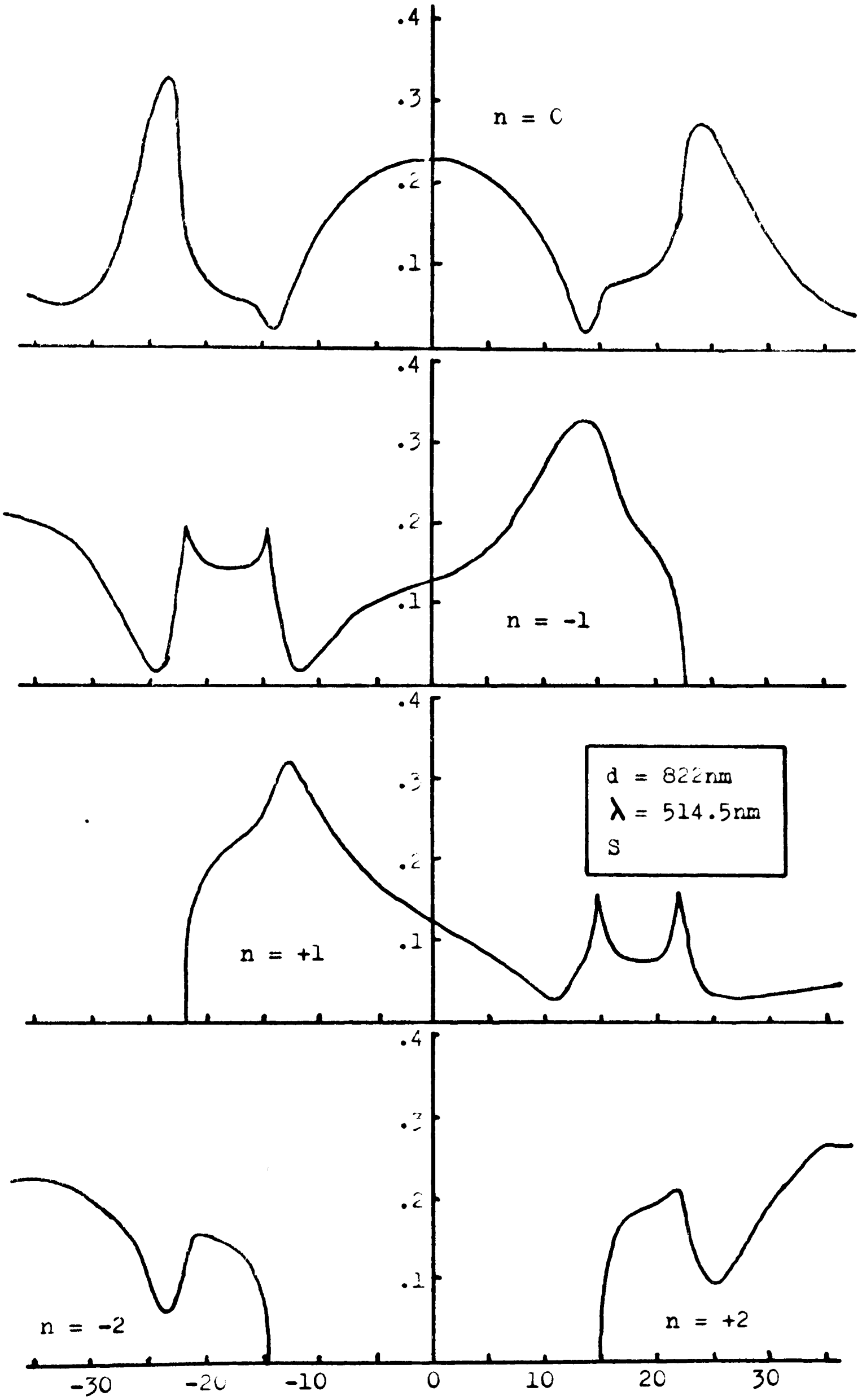

.4
.3
.2
.1
n = 0
n = -1
n = +1
n = -2
n = +2
d = 822nm
λ = 514.5nm
S
-30
-20
-10
0
10
20
30
angle of incidence (degrees)

Fig. 6.5

It is seen that these resonances have the effect of blazing the grating by concentrating the propagating energy into a single order. This effect is unlikely to be of use because of the narrow range of the blaze effect and also because other SP processes reduce the overall efficiency to a low level.

6.2.2 Efficiency in Littrow mountings

6.2.2(a) The wavelength of the SP resonance in Littrow mountings

The resonance wavelengths for a grating used in a Littrow configuration may be calculated as follows. As shown in appendix 2, the wavelength and the angle of rotation of the grating, θ, are related by the expression

$$\lambda = -\frac{2d}{n}\sin\theta$$

where n is the order of observation.

Now θ is also the angle of incidence and so the resonance condition may be written

$$\sin\theta = -\delta_m\Gamma(\lambda) + m.\frac{\lambda}{d}$$

Elimination of $\sin\theta$ from these equations gives

$$\frac{\lambda}{d} = \frac{\delta_m\Gamma(\lambda)}{(\frac{n}{2} + m)}$$

This expression is similar to that given in appendix 2 for the Rayleigh wavelength in Littrow mountings, which again demonstrates the close association between anomalies and SP effects. It is seen that the resonance occurs at a slightly longer wavelength than the corresponding anomaly.

It should be noted that because of the variation of $\Gamma(\lambda)$ with wavelength the above expression cannot be calculated exactly.

Accurate values of the wavelength may be found by an iterative process if $\Gamma(\lambda)$ is well known.

6.2.2(b) Efficiency Variation

Figure 6.6 shows the observed first order Littrow efficiency of a 1200l, mm HDG in the region of the anomaly caused by the simultaneous passing off of the -2 and +1 orders. The calculated wavelength of the resonance and the Rayleigh wavelength are indicated.

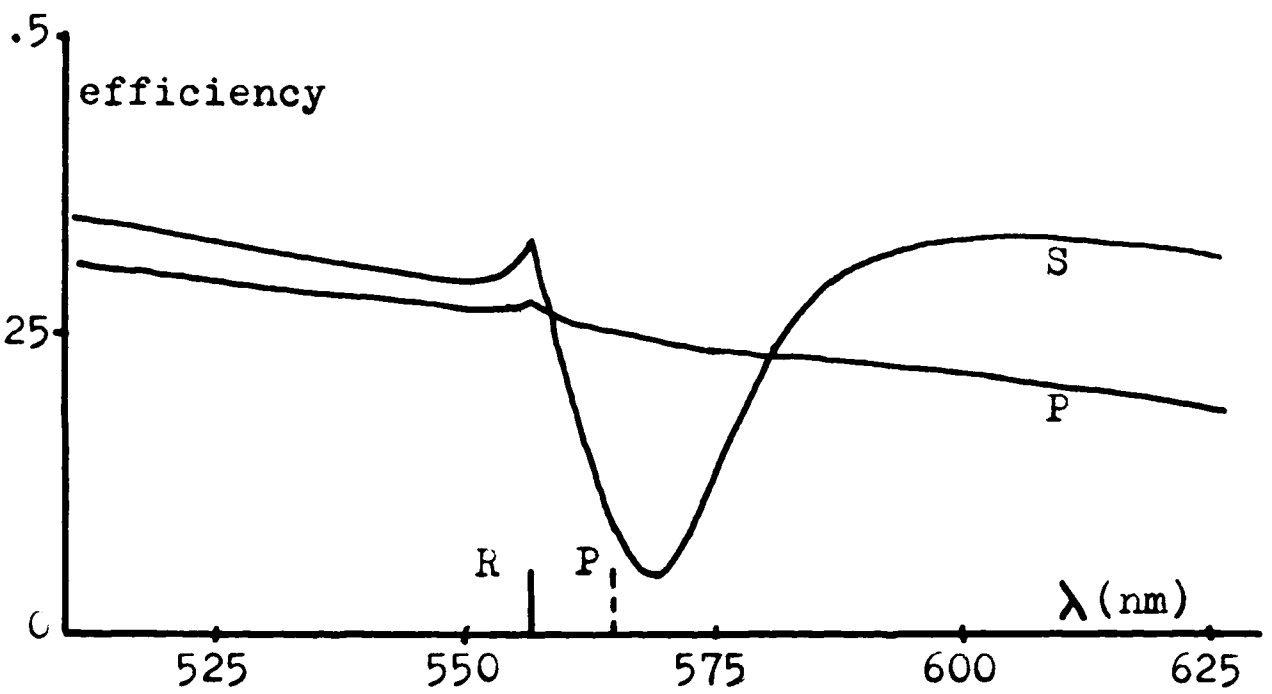


Fig. 6.6

We see that the double SP resonance gives rise to a strong loss of efficiency in S polarisation. This corresponds to the dark anomaly first observed by Wood (25). This loss of efficiency is not always the case as shown in figure 6.7. Here we see the Littrow efficiency of a 770 line/mm grating in the region of the anomaly caused by the passing off of orders +2 and -3.

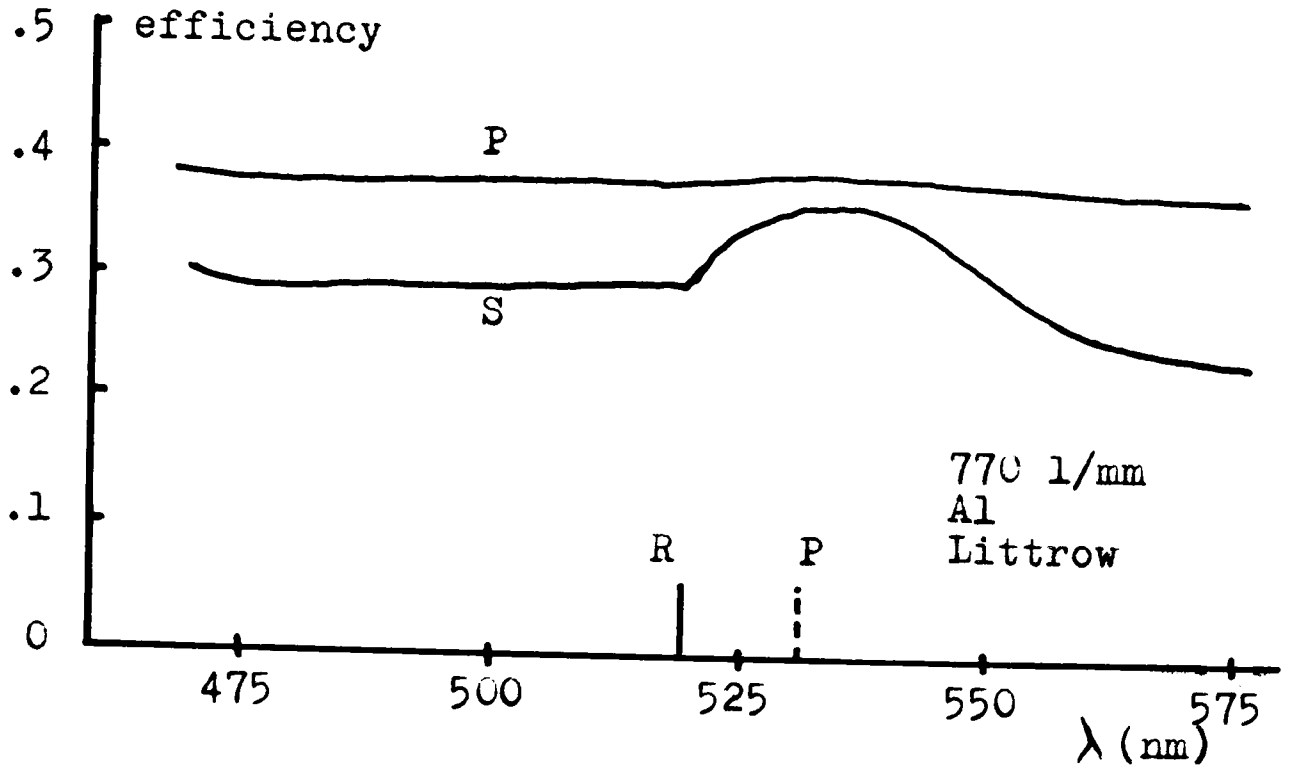


Fig. 6.7

There is now a peak of efficiency in S polarisation corresponding to the SP resonance. The difference between the two anomalies must be explained in terms of the redistribution of energy among the orders due to resonant SP decay (section 6.2.1 (b)). A detailed explanation, however, is beyond the scope of this simple model, and must be left to rigorous diffraction theory.

6.2.2(c) Czerny-Turner Mountings - The merging of resonances

As the angle between the beams in a constant deviation mounting increases from zero, the two anomalies, which are coincident in the Littrow mounting separate (see appendix 2). Thus it is possible to study the behaviour of SP resonances as they merge.

The exact wavelengths at which the resonances occur are difficult to calculate analytically. If necessary they may be found by solving numerically the simultaneous equations:

$$-\sin(\theta + \alpha/2) = -\delta_m \Gamma(\lambda) + m\frac{\lambda}{d}$$

$$\sin\theta \cos\frac{\alpha}{2} = -\frac{n\lambda}{2d}$$

where the symbols have their usual meanings. Here, the upper equation is the resonance condition using the angle of incidence $-(\theta + \alpha/2)$ and the lower equation is the Czerny-Turner equation.

In figure 6.8, we have chosen a configuration in which one anomaly produces a peak and the other a loss of efficiency in the order of observation. The quantity plotted is the ratio I_s/I_p, which shows accurately the shape of the resonances. The Rayleigh wavelengths calculated from appendix 2 are indicated.

We see that the resonances merge and subsequently separate with no interchange of identity.

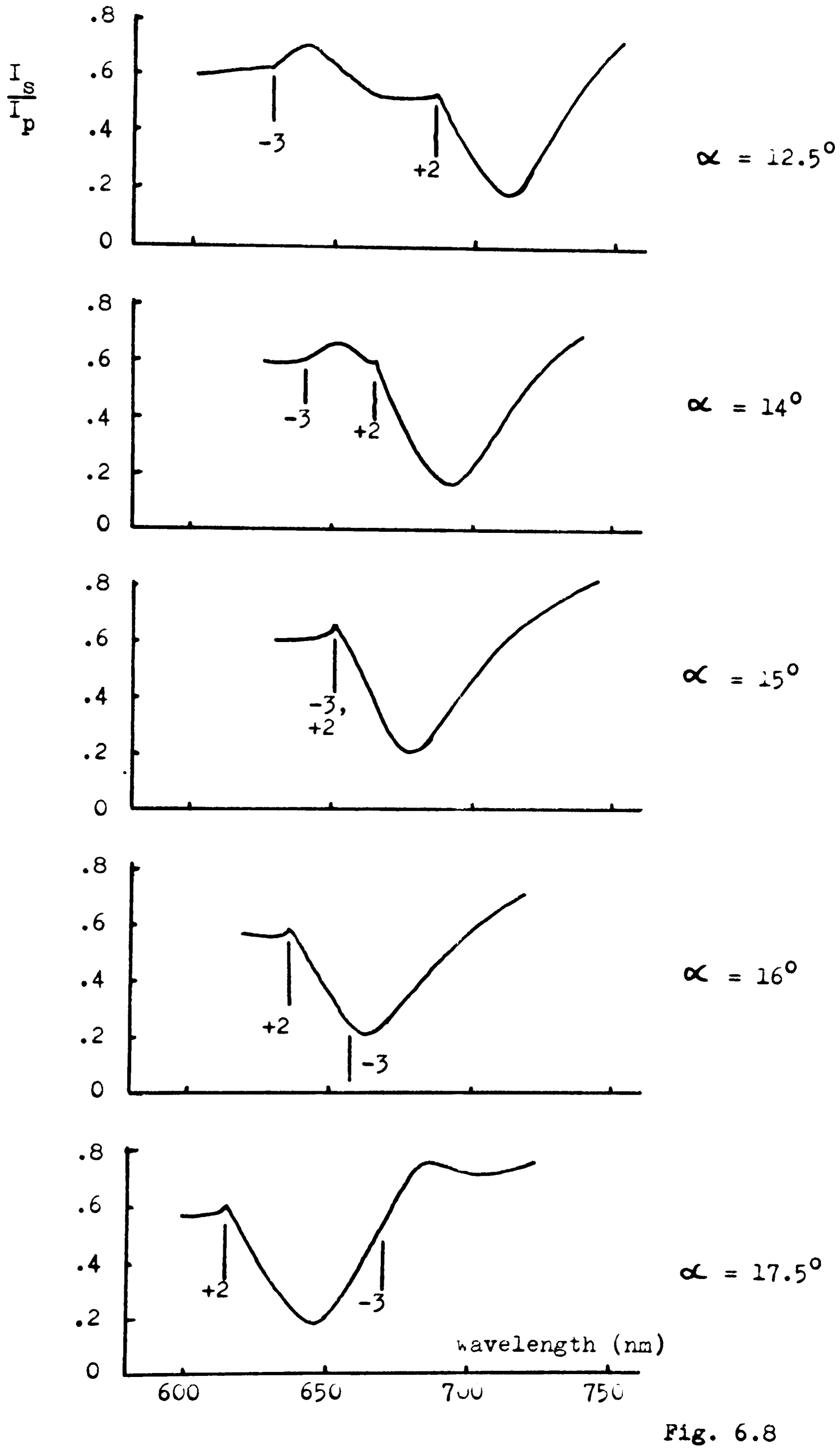

$\frac{I_s}{I_p}$
.8
.6
.4
.2
0
-3
+2
∝ = 12.5°
-3
+2
∝ = 14°
-3,
+2
∝ = 15°
+2
-3
∝ = 16°
+2
-3
∝ = 17.5°
wavelength (nm)
600
650
700
750

Fig. 6.8

6.2.2(d) The effects of groove depth on the shape of the resonance

So far we have considered only the basic properties of the SP resonance (i.e. the conditions under which it occurs and the relative intensities of the orders at resonance) and have ignored the effect of the groove profile of the grating.

The values of $\Gamma(\lambda)$ which we have been using are calculated for a plane surface. On a surface with a profile, this value changes, so that one effect of altering the groove profile will be to shift the resonance. In addition, increasing the groove depth is found to broaden the resonance. In figure 6.9 we see the ratio I_s/I_p for the +1 order of a Littrow mounted HDG for several groove depths.

From this diagram we note the following:

i) The resonance is terminated discontinuously at the Rayleigh wavelength;

ii) As the groove depth increases, the centre wavelength of the resonance moves toward longer wavelengths.

iii) As the groove depth increases, the resonance becomes broader and shallower.

The wavelength shift and the broadening will now be considered in detail.

Wavelength Shift

One theory of the shift of resonance wavelength with groove depth is that it is due to the extra distance which must be travelled by the SP in one period of the grating. (This assumes that the SP follows accurately the profile of the grating). The effect of this will be to slow down the SP and so increase its apparent momentum. This causes the value of $\Gamma(\lambda)$ to increase and so the resonance is shifted to longer wavelengths.

If this is correct, the wavelength of the resonance will be given

$$\lambda_p = \lambda_o \cdot \frac{s}{d}$$

where λ_o is the resonance wavelength extrapolated to zero groove depth

S is the length of surface in one period of the groove

d is the groove period

For a sinusoidal profile, this becomes

$$\lambda_p = \frac{\lambda_o}{d} \cdot 4 \int_o^{d/4} \left[1 + \left(\frac{A}{d} \cos \frac{2\pi x}{d} \right)^2 \right]^{\frac{1}{2}} dx$$

where A is the peak-to-peak groove depth.

This variation is shown in figure 6.10 (broken line) together with the observed values (points). It is seen that the predicted shift is much greater than that observed experimentally.

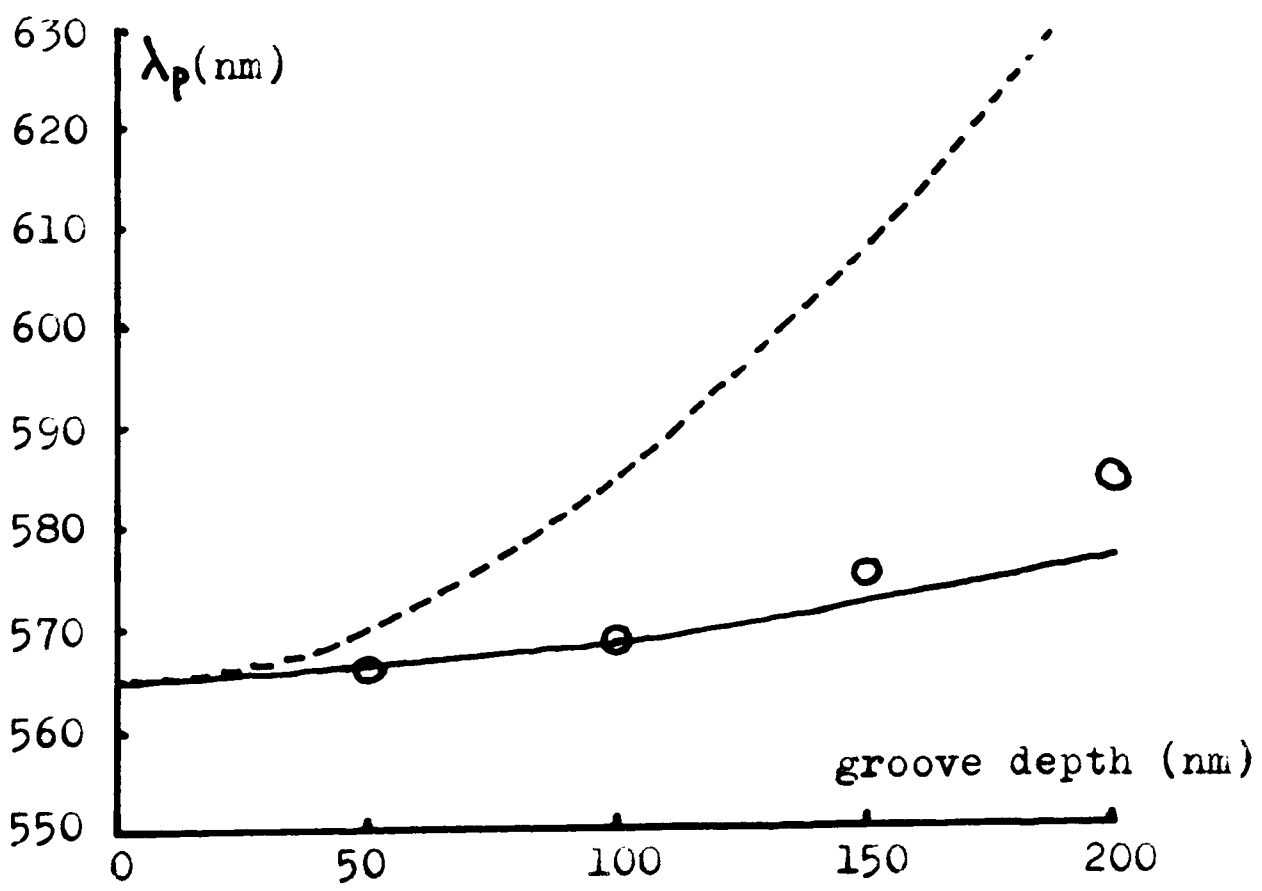


Fig. 6.10

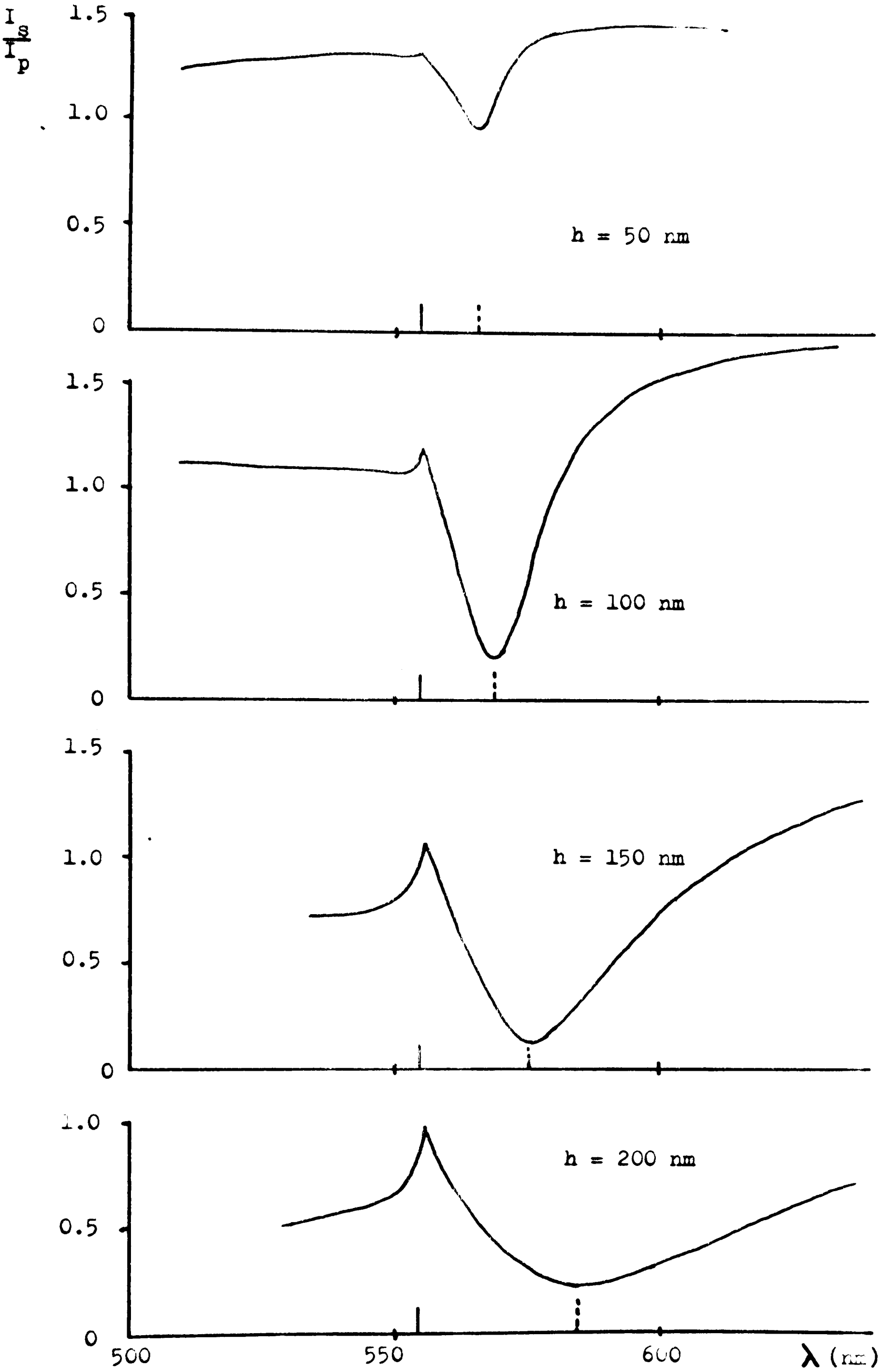


Fig. 6.9 : SP resonances at different groove depths. 1200 l/mm, Al.

This supports the findings of McPhedran (40) who calculates the resonance wavelength using rigorous electromagnetic theory. The solid line in figure 6.12 shows the variation of λ_p taken from figure 5 of (40). It is seen that agreement with the observed points is closer.

McPhedran (41) has suggested that the inaccuracy of the simple theory may be due to the SP not following the surface exactly, but penetrating the tops of the grooves, thereby reducing the effective path length.

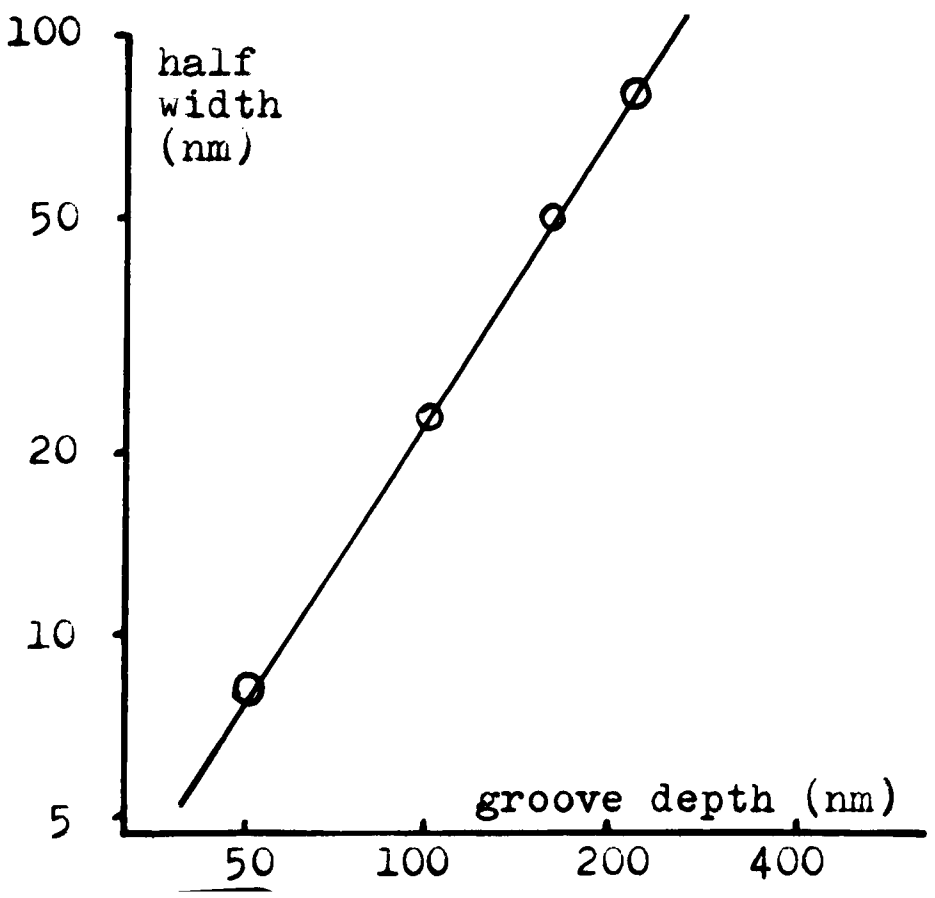
Broadening

Using infinite conductivity theory, McPhedran (42) has found an empirical power law relationship between the half width of a resonance and the groove depth of the grating, viz:

$$HW \propto A^{3.93}$$

where HW is the half width of the resonance and A is the groove depth.

In figure 6.11, the observed halfwidth (measured from the long wavelength half of the resonance curve) is plotted against the groove depth on a logarithmic scale.



An approximate straight line is observed with slope 1.68. This indicates a relationship of the form

$$HW \propto A^{1.68}$$

6.2.2(e) The effect of the metal coating on the performance of diffraction gratings

All of the effects discussed so far have bee n observed on gratings coated with Al. It is obvious, however, that because of the dependence of $\Gamma(\lambda)$ on the properties of the metal, the performance of the grating in the anomalous regions will depend strongly on the metal with which it is coated.

Figure 6.12 shows the polarisation (Is/Ip) in the anomalous region of a 1200 l/mm grating overcoated successively with Al, Au and Ag.

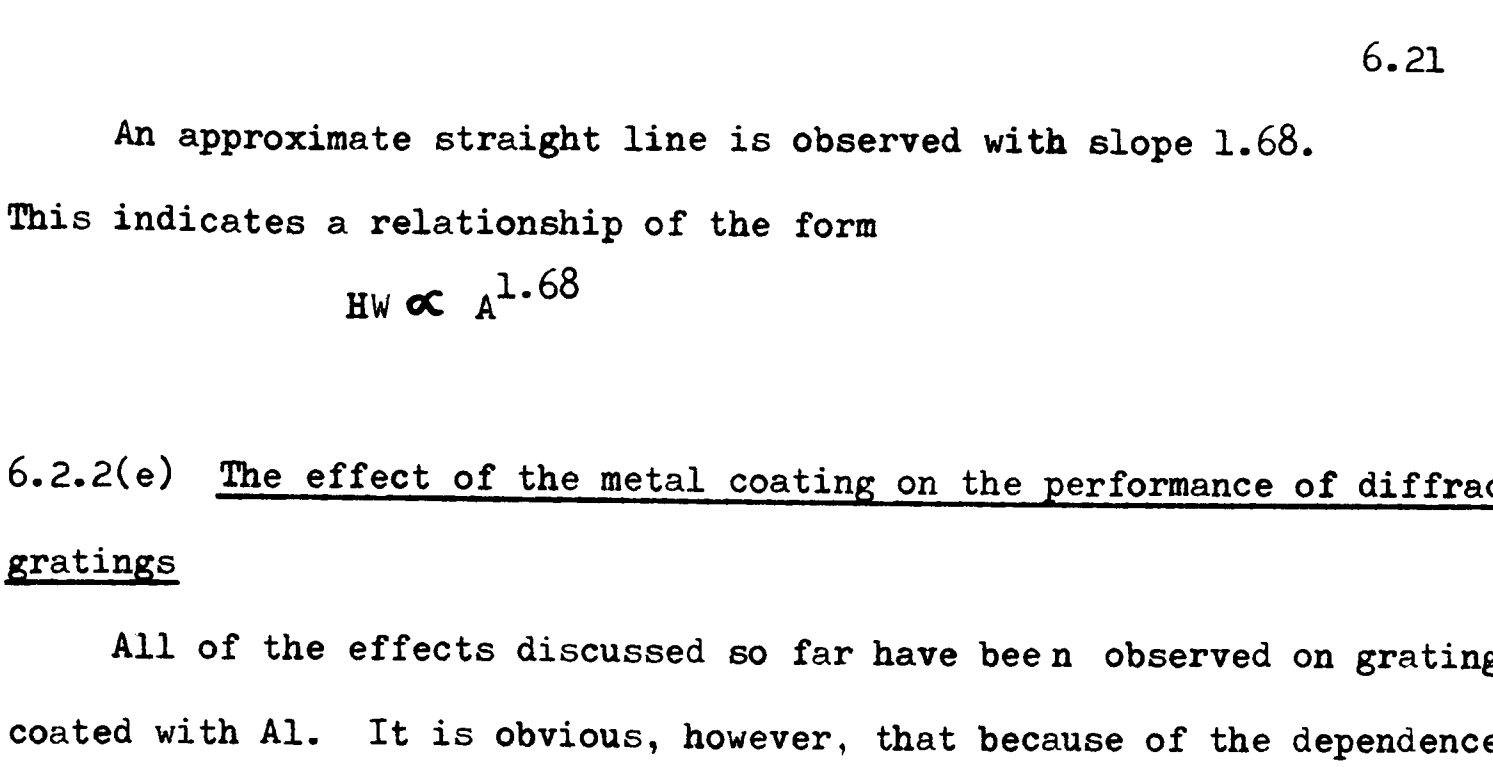


Fig. 6.12: SP resonances. 1200 l/mm, Littrow

It is seen that not only the position but also the width of the resonance depend on the metal.

Ag, which is essentially a perfect metal in terms of free-electron theory, shows a sharp resonance. Au absorbs strongly at the wavelength of the resonance and this shows a broad, relatively weak resonance. Al is an intermediate case, and its properties may also be affected by the thin layer of oxide which forms on the surface on exposure to air.

6.3 Surface plasmons and scattered light

In section 6.1.2 it was predicted that the SP resonance would affect the level of scattered light produced by a grating. We now examine in detail the properties of this scattered light (here referred to as surface plasmon scattered light or SPSL).

First we consider the variation of the total level of SPSL as a function of angle of incidence, at constant wavelength. We then examine the spectral distribution of SPSL in white light and finally consider the angular distribution of SPSL.

6.3.1 The total intensity of scattered light

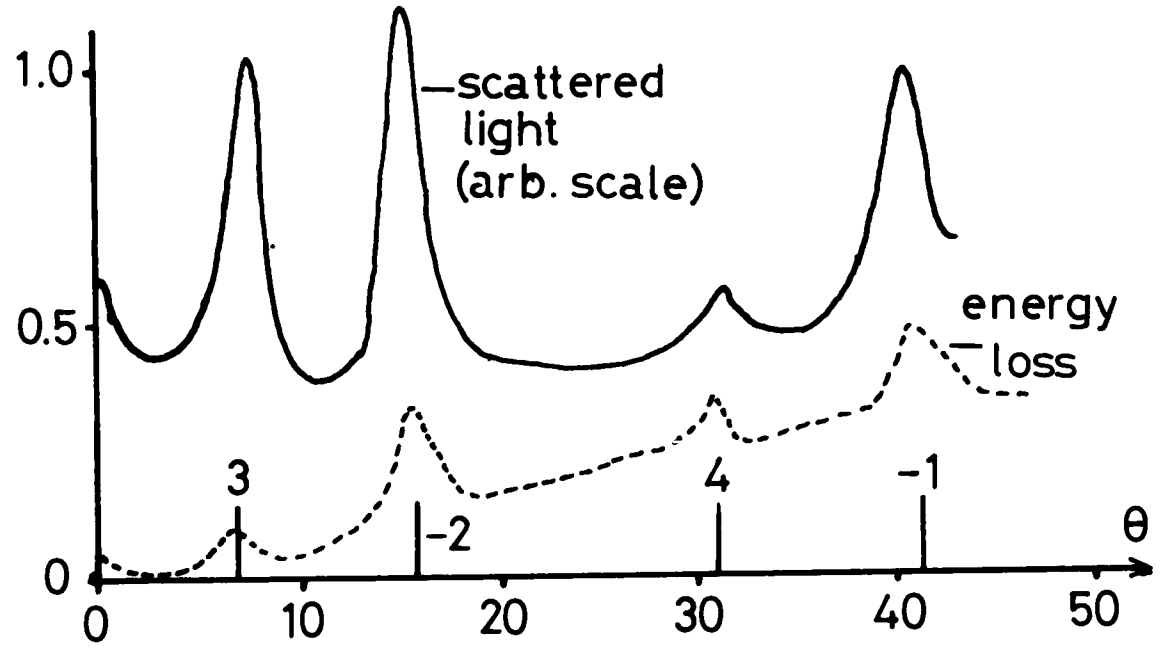
6.3.1(a) Variation with angle of incidence at constant wavelengths

If a SP is scattered before it decays, the resulting photon will travel in a direction other than the diffracted orders and will contribute to the level of scattered light produced by the grating. The intensity of this scattered light must be proportional to the density of SP's on the grating, thus at resonance, when large numbers of SP's are produced, it will increase.

As with all SP effects, this will only occur if the incident light is S polarised.

The total intensity of SPSL is difficult to measure absolutely, but a relative figure may be obtained by measuring the intensity in a sufficiently large solid angle at a fixed direction to the grating normal. The way in which this is done is described in 4.2.1.

Figure 6.13 shows the variation of intensity of scattered light with angle of incidence (solid line) for a 800 line/mm grating at 488 nm. The vertical lines indicate the calculated resonance angles.



It is seen that at resonance the level of scattered light increases dramatically, reflecting the rise in the SP flux.

Also shown in the diagram is the measured energy loss of the grating (i.e. the total loss of efficiency from the diffracted orders). The similarity between the two curves indicates that both effects are due to the decay of SP's.

It should be emphasised that the small fraction of SP's which give SPSL is not sufficient to account for the large energy loss at resonance; this must be explained by decay within the metal leading to Joule heating.

Figure 6.14 shows how the intensity of SPSL changes with angle of incidence for a number of different wavelengths. The measurements were made on a 1200 l/mm grating coated with Ag. The resonance angles are indicated.

As a final observatinn on this section it has been noted that if the surface of a grating deteriorates, either through aging or contamination the intensity of SPSL increases. This is to be expected since the poor surface increases the scattering of SP's.

6.3.1 (b) Spectral distribution of SPSL in white light

When a resonance occurs with white light incident on the grating, the SPSL should contain only those wavelengths which fall under the resonance curve. Thus the spectral distribution of SPSL should be related to the efficiency of the grating as a function of wavelength.

It is difficult to measure SPSL in white light, since the intensity per unit bandwidth of the source is now much less than that of the laser used in the previous section. Also the light must be analysed by a monochromator, which introduces further losses.

To overcome these problems, the following technique was used. The grating used had a frequency of 2800 line/mm and was held fixed relative to the incident beam. The angle of incidence, 34^{o}, was chosen so that the first order resonance occurred at 570 nm, the peak sensitivity of the detection system.

The zero order efficiency was measured as a function of wavelength as described in 4.1.3, and then using the collector described in 4.2.1, the spectrum of SPSL was analysed using a motor driven monochromator. Because the signal was very weak, it was averaged in a MCS and the scan made over a long period.

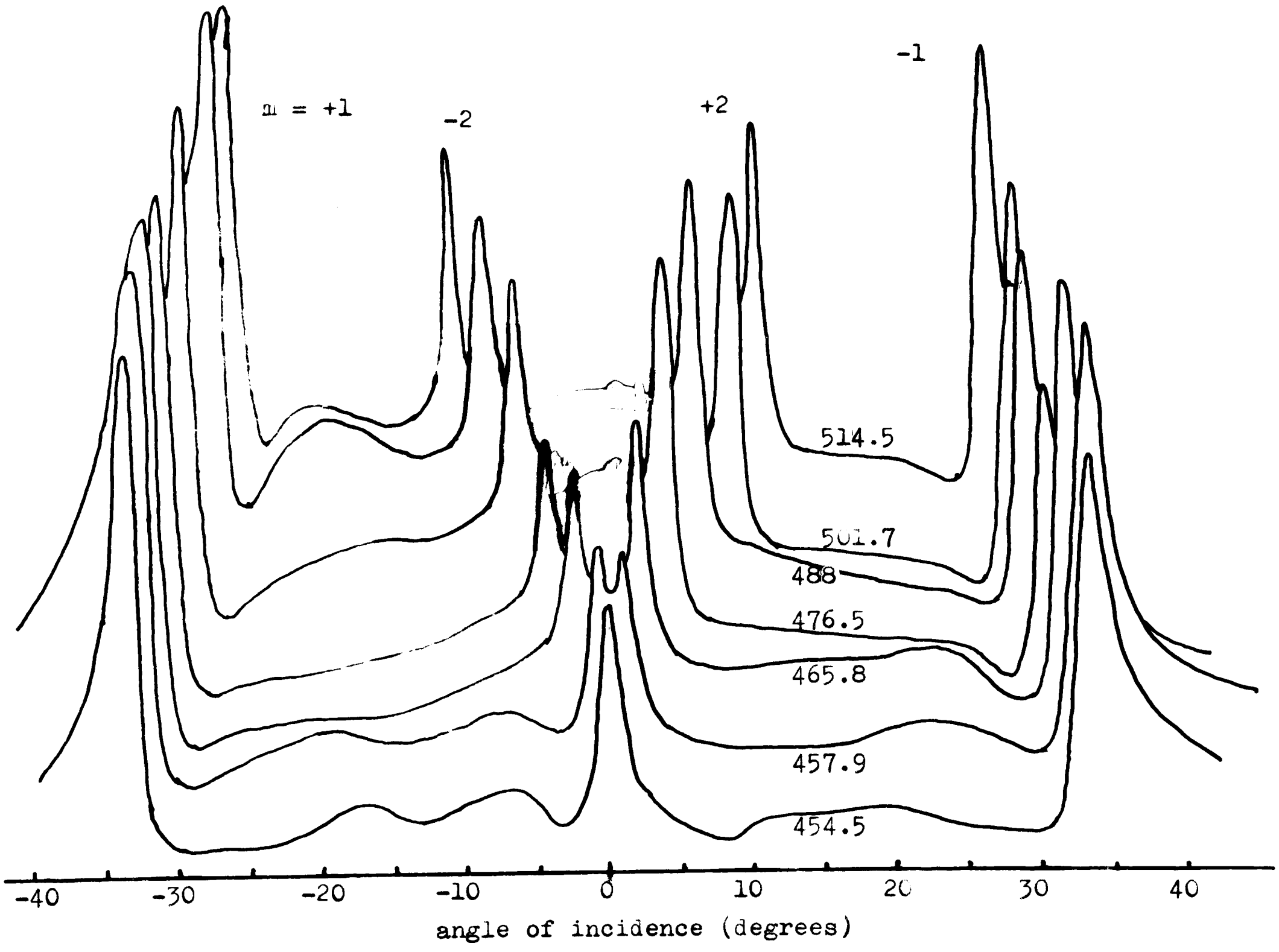
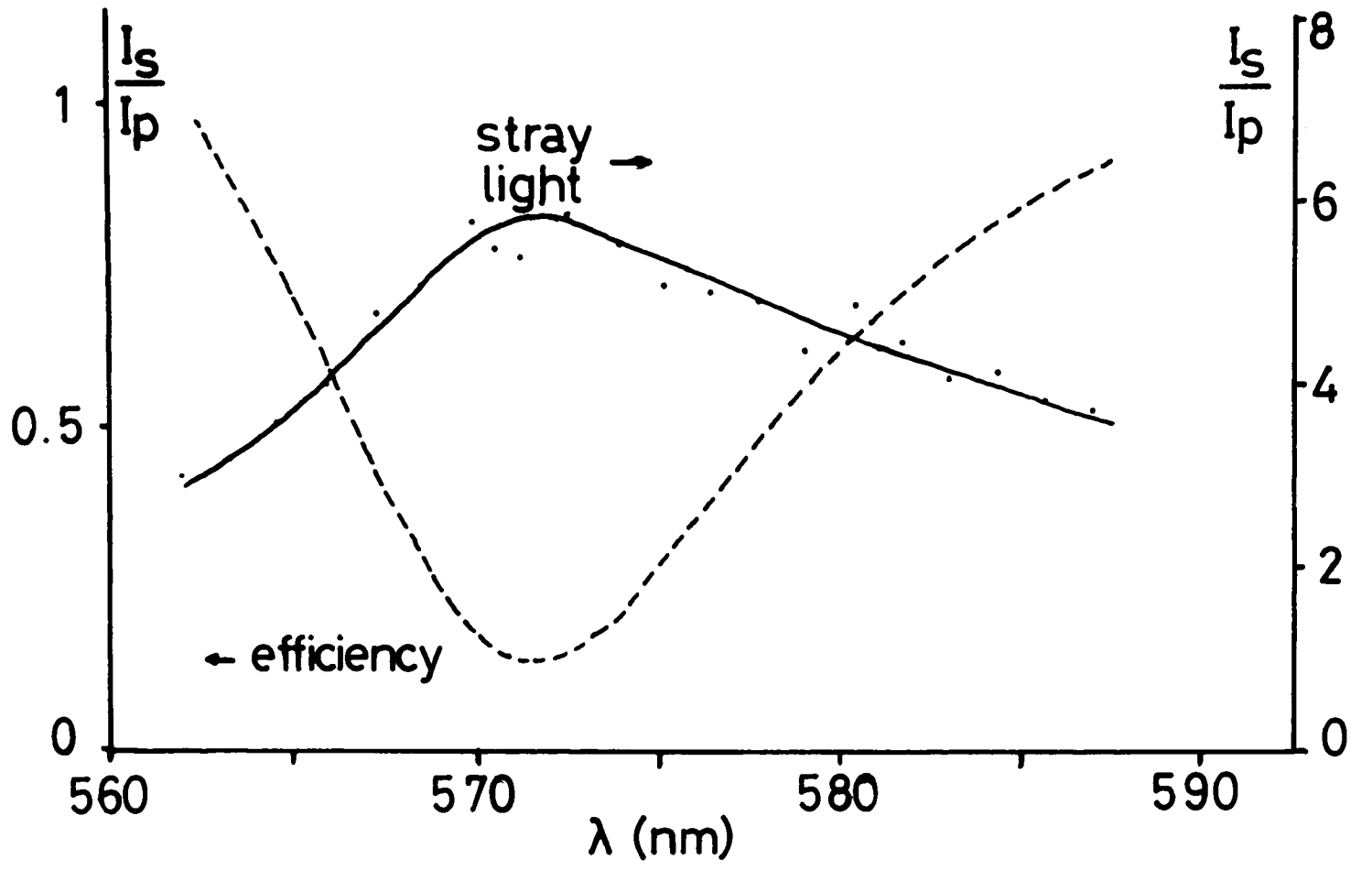


Fig. 6.14 : Intensity of scattered light vs. wavelength and θ_i

1200 l/mm grating, Ag coated

The results are plotted in figure 6.15.

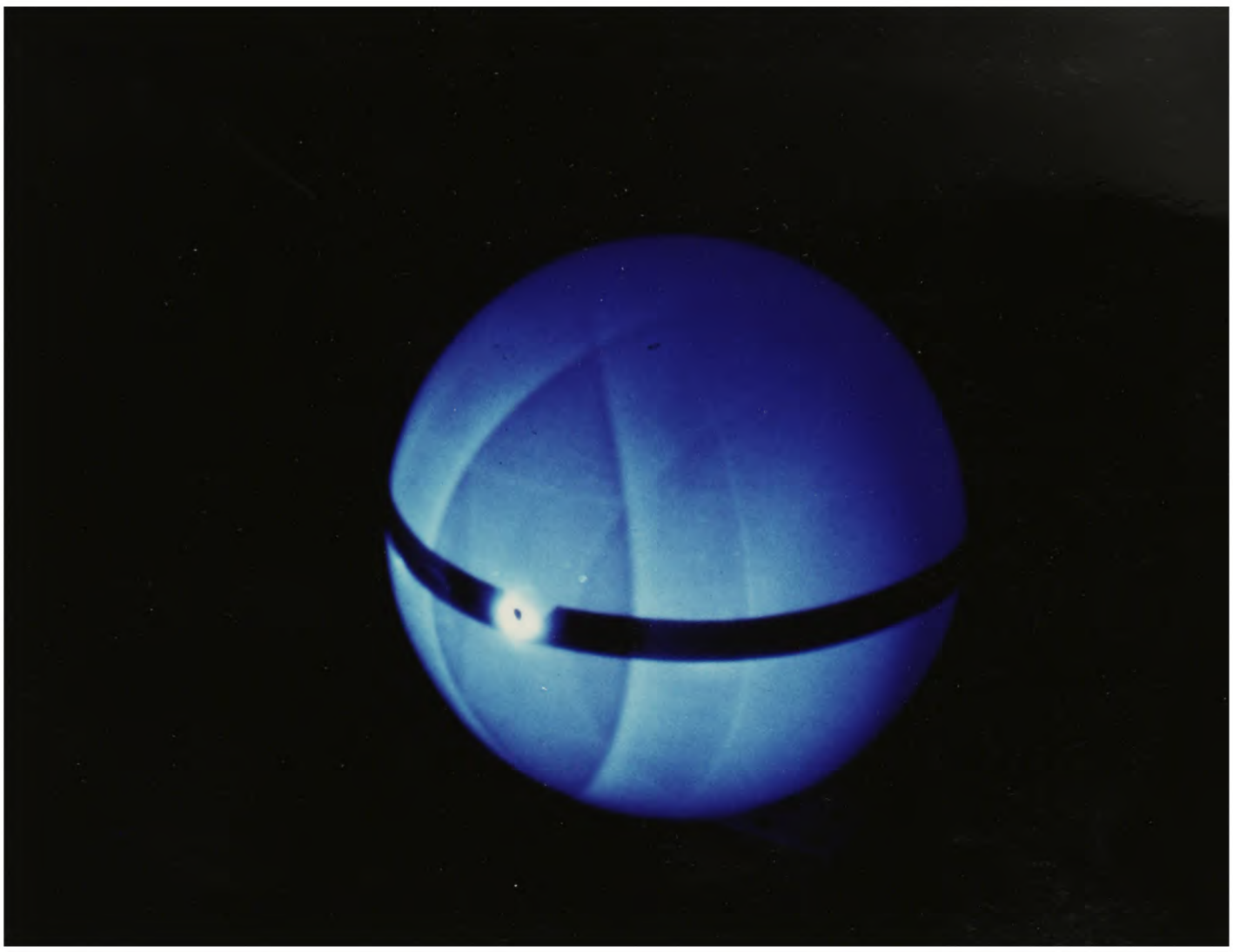


The broken line shows the ratio of the S and P efficiencies and the solid line shows the mean ratio of the S and P SPSL. It can be seen that the S polarised SPSL goes through a maximum corresponding to the minimum in the efficiency curve. This is further confirmation of the connection between scattered light and SP's.

6.3.2 The angular distribution of scattered light

6.3.2(a) Qualitative description

When a SP decays to give a photon, the angle at which the photon is emitted must be such that momentum is conserved in the emission process. Hence the SPSL will show a well-defined angular structure. If a laser is used as a light source, the SPSL is sufficiently intense to be observed visually and a convenient way of doing this is shown in figure 6.16.



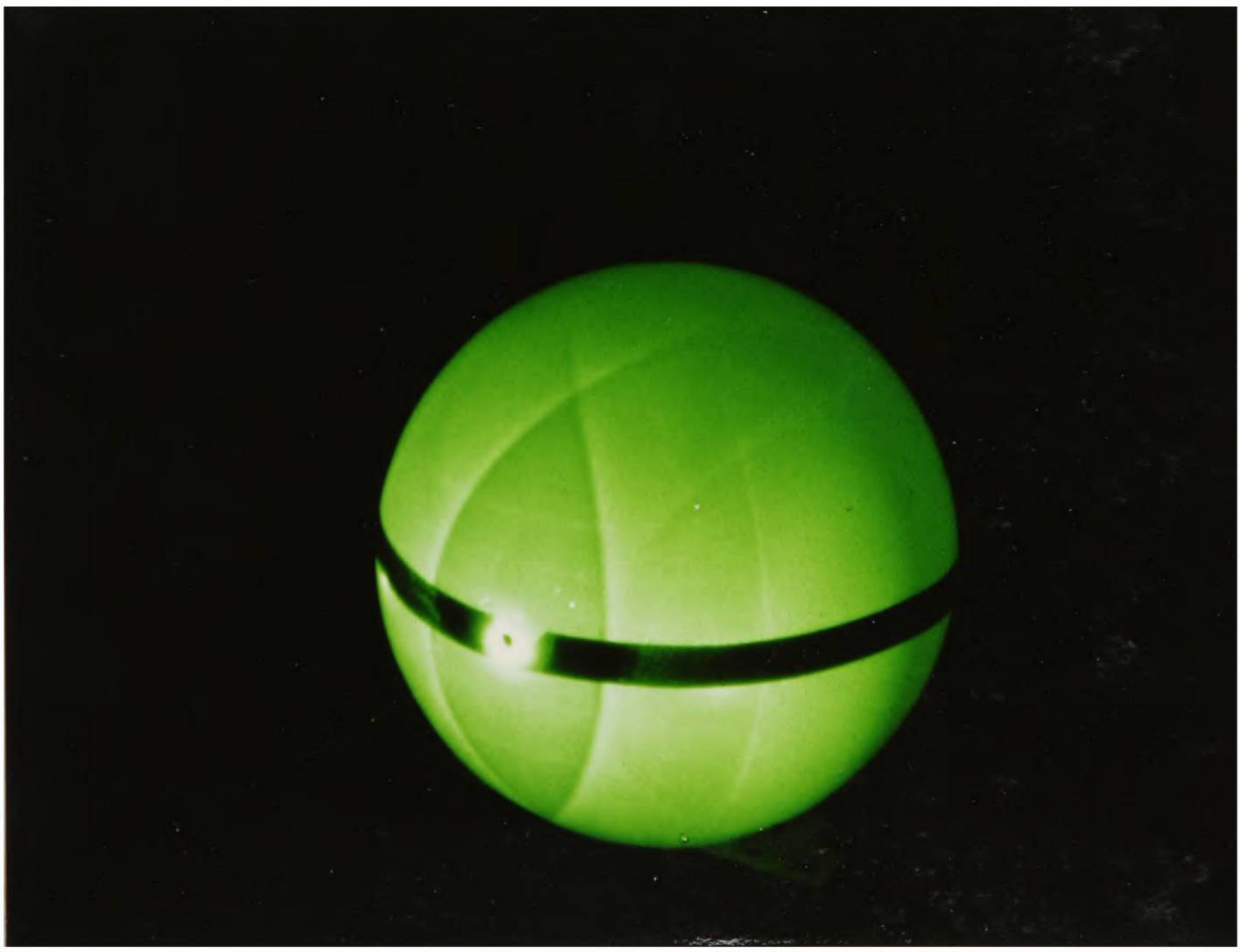


Fig. 6.16

The grating is mounted at the centre of a translucent glass globe which has a small hole drilled in it to admit the laser beam. The plane of incidence is masked with black tape so that scattering from the intense orders does not mask the pattern. As shown in the illustration, a striking pattern of bright and dark bands is formed on the globe.

In general, patterns are seen to consist of two distinct kinds of feature:

i) "Sharp" arcs, consisting of a single bright band of small angular width;

ii) "Diffuse" arcs consisting of a bright band adjacent to a dark band, usually separating areas of high and low intensity.

The intensity in the two types of feature is indicated schematically in figure 6.17.

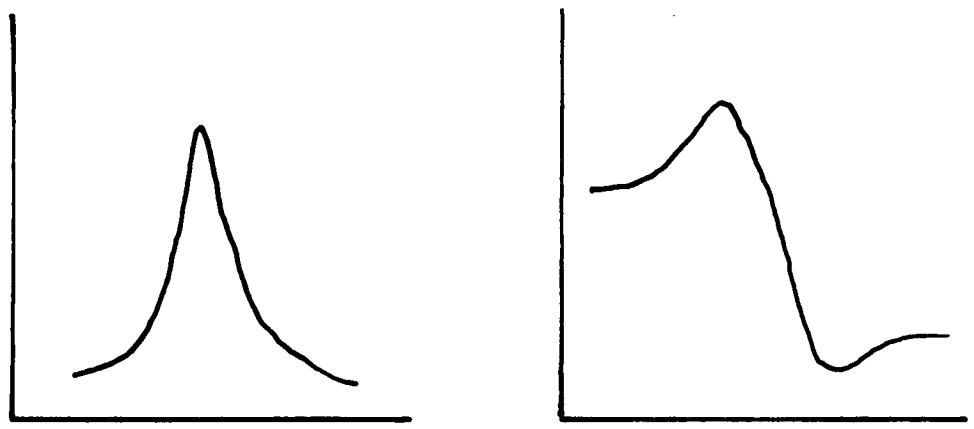


Fig. 6.17

The dependance of the two types of feature on the rotation of the grating and the polarisation of the incident beam is summarised in table 2.

The fact that the diffuse lines are associated with the SP resonance and S polarisation identifies them as SPSL.

Table 2

Dependance on	Sharp	Diffuse
Rotation of grating	Position: occur in pairs about each order and move at same speed as orders. Intensity: Constant	Position: Symmetric about grating normal. Rotate at same speed as grating. Intensity: Low except at SP resonance angles when large increase is seen.
Polarisation of incident beam	Intensity: Constant Polarisation: Mainly that of the incident beam	Intensity: Low in P High is S Polarisation: Mainly S

The sharp lines were eventually identified as the holographic reconstruction of light scattered in the resist layer when the grating was formed. As explained in Appendix 4, this is reconstructed as pairs of arcs centred on the diffracted orders.

6.3.2(b) Calculation of the position of the SPSL lines

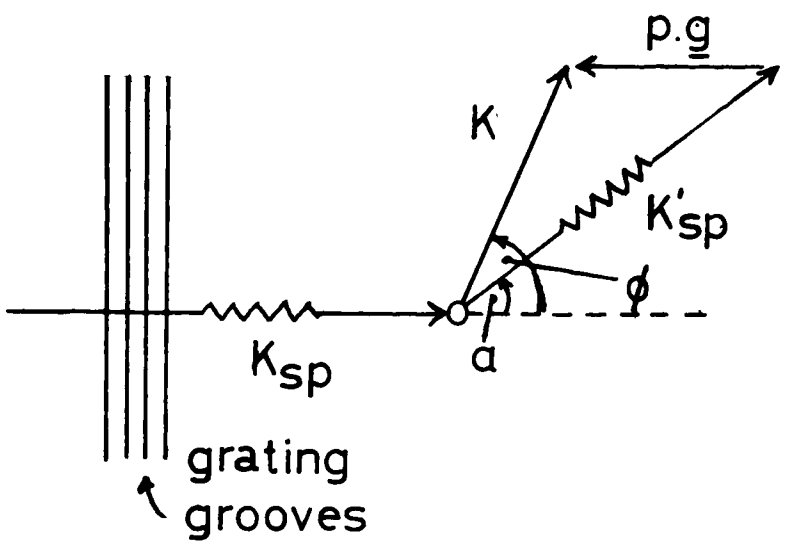


Fig. 6.18

In figure 6.18, a SP with wavevector k_{sp} normal to the grooves is scattered through an angle α. It then decays in conjunction with a vector $p.\overline{g}$ ($g = 2\pi/d$) to give a photon with a parallel component of wavevector K at an angle ϕ to the original SP.

The value of K is given by the vector equation:

$$\overline{K} = \overline{K'}_{sp} + p.\overline{g}$$

If the SP is scattered elastically, $k'_{sp} = k_{sp}$ and also the wavelength of the decay photon is equal to the incident wavelength.

Thus if θ is the angle of the emitted photon to the grating normal, we may write

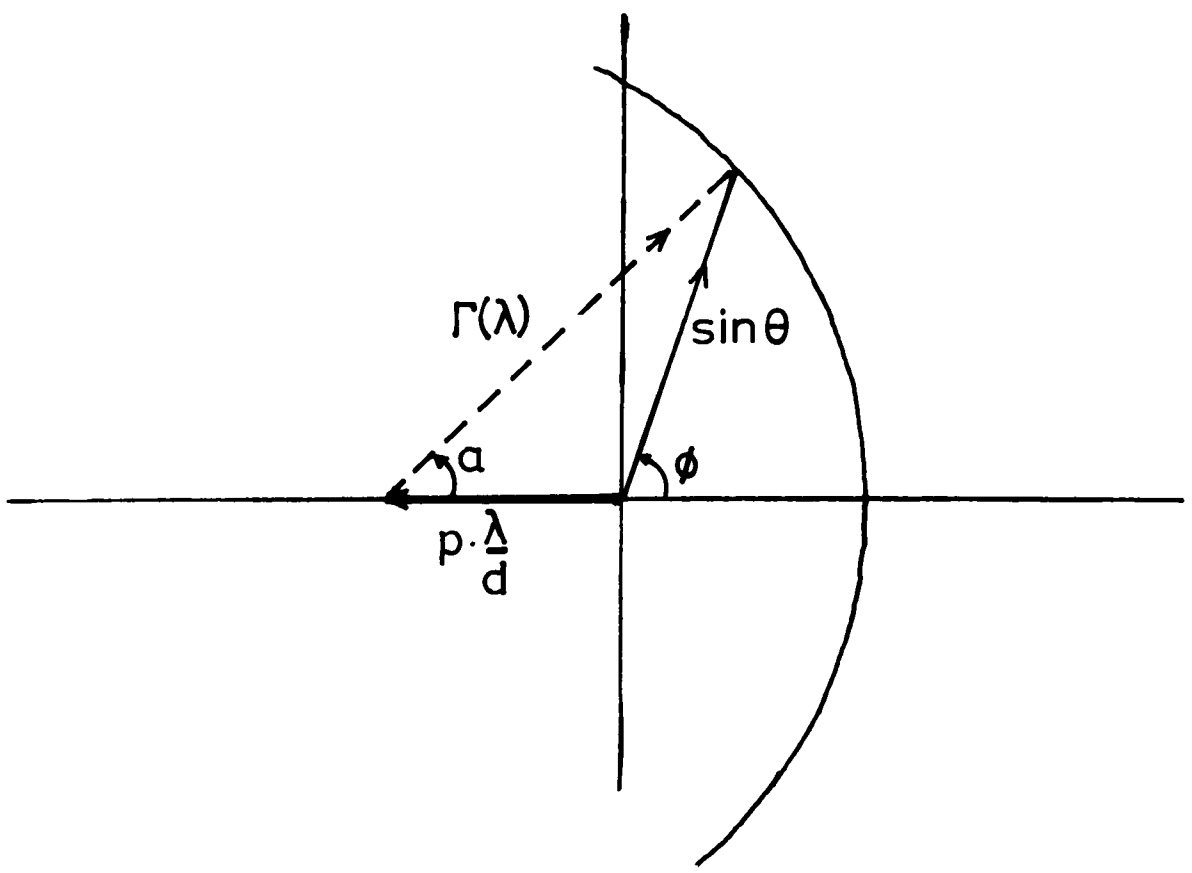
$$\overline{\sin\theta} = \overline{\Gamma(\lambda)} + p.\overline{\frac{\lambda}{d}}$$

where the vectors have the same directions as in the previous equation. Thus we may show that

$$\sin\theta = \left[\Gamma^2(\lambda) + 2p\Gamma(\lambda)\cos\alpha.\frac{\lambda}{d} + p^2\frac{\lambda^2}{d^2}\right]^{\frac{1}{2}}$$

This determines the positions of the SPSL lines. It is more convenient, however to find $\sin\theta$ by the following graphical method.

If a set of polar co-ordinates is chosen where angle represents the angle from the groove normal in the plane of the grating, ϕ, and radial distance represents $\sin\theta$, it is possible to redraw figure 6.18 so that $p.\lambda/d$ is represented by a distance from the origin along the $\phi = 0$ axis. The SP vector may now be represented by a circle of radius $\Gamma(\lambda)$ centred on the end of $p.\lambda/d$ (figure 6.19).



The value of $\sin\theta$ for any angle ϕ may now be found by the intersection of the circle with the ϕ co-ordinate.

A full graphical construction for a typical case is shown in figure 6.20. It should be noted that the position of the pattern is independent of angle of incidence and so it remains fixed relative to the grating.

6.3.2(c) Comparison between the predicted and observed positions of the lines.

Using the techniques described in 4.2.2, it is possible to measure the intensity of scattered light as a function of ϕ at constant It is possible to build up a full intensity distribution by taking different values of θ .

This is done in figure 6.21. The verticial lines marked on the curves are the predicted positions of the lines taken from figure 6.20.

It is seen that the calculated angular position corresponds to the position of maximum slope of the line - i.e. the boundary between the bright anddark bands.

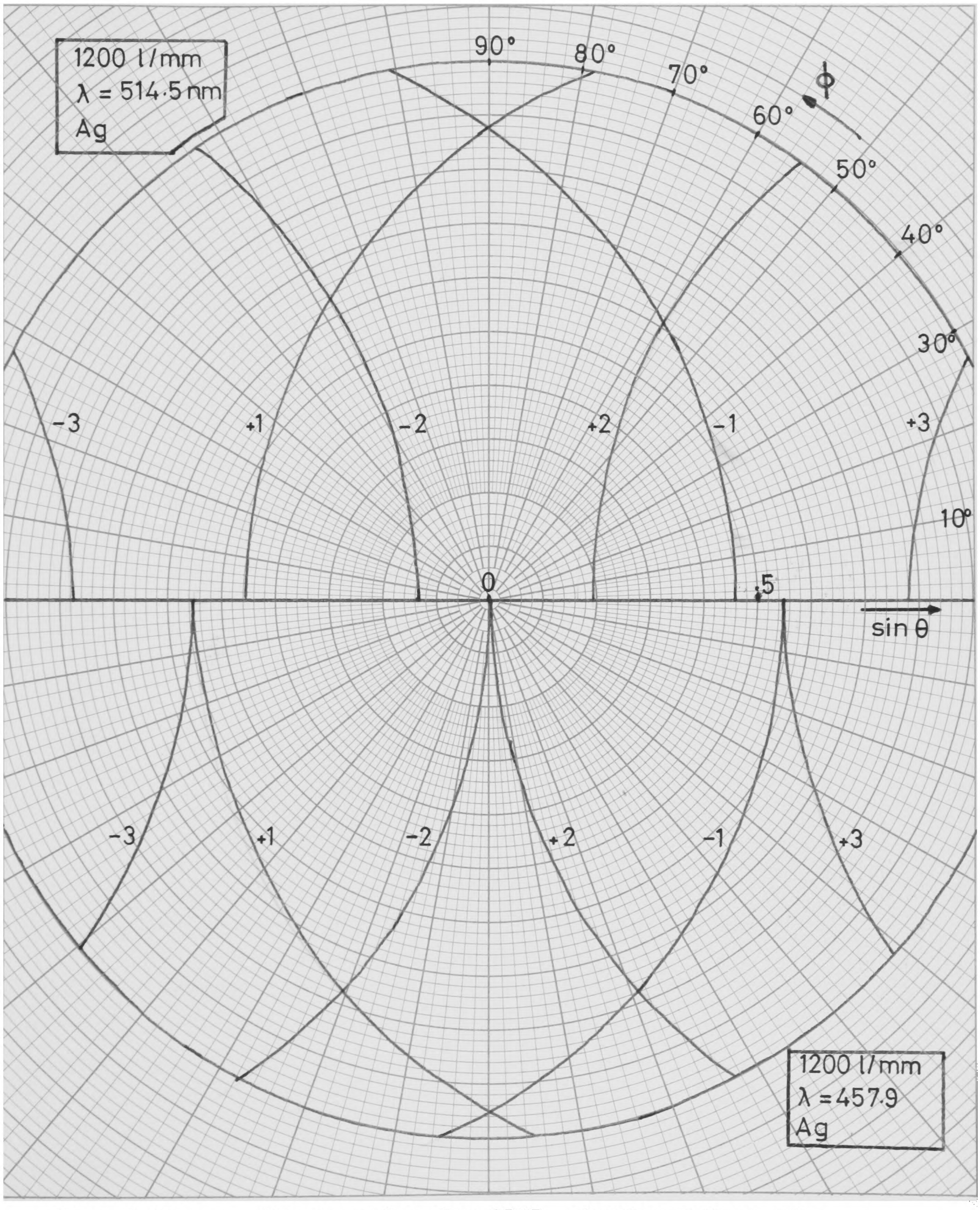


Fig. 6.20: Construction for SPSL, showing values of p. 1200 l/mm; silver; λ = 514.5 nm and 457.9 nm

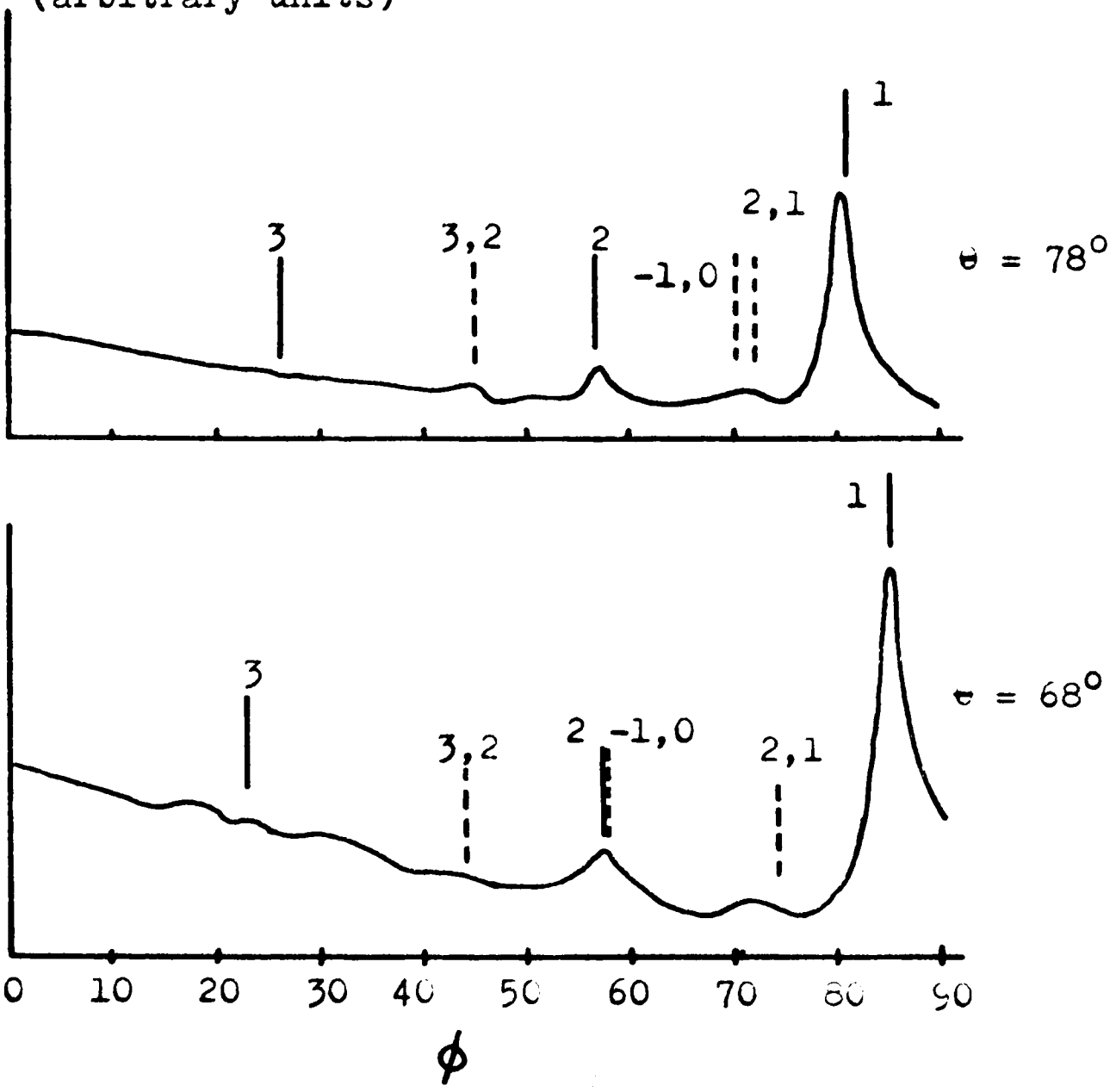


Fig. 6.21: Angular distribution of scattered light.
1200 l/mm; silver; normal incidence
λ = 514.5 nm.

Solid lines indicate predicted angles of SPSL emission (Fig. 6.20)
Broken lines indicate predicted angles of holographically reconstructed scattered light (Fig. A4.1)

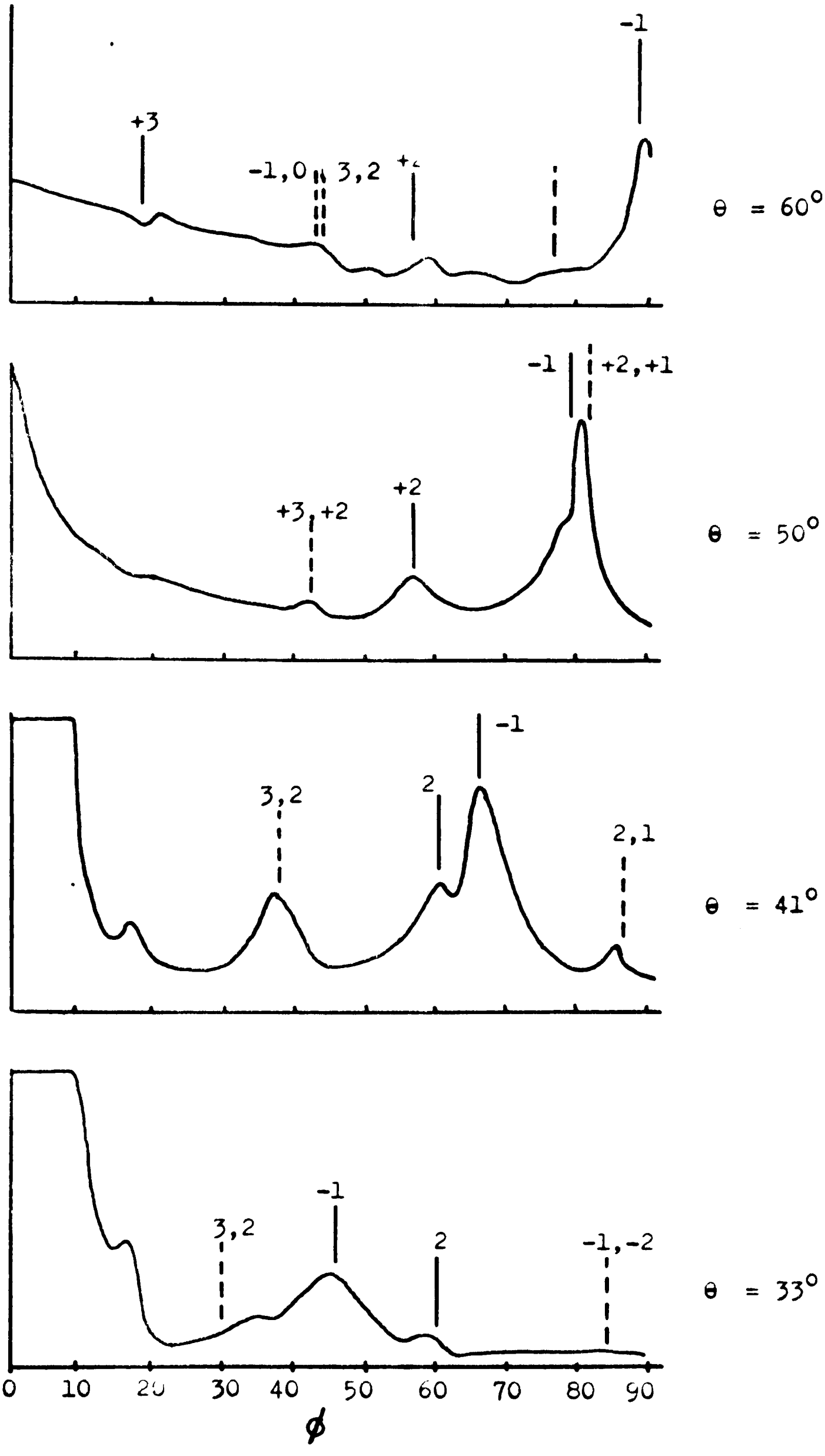


Fig. 6.21 (cont'd)

6.4 The determination of SP properties from measurements on diffraction gratings

Any of the phenomena referred to in this chapter may be used to study the properties of SP's on periodic surfaces. The information obtainable is in the form of a dispersion curve ($\Gamma(\lambda)$ as a function of λ), which can then be used to deduce the effective optical constants of the metal. The method would be particularly convenient for studying effects at a metal-dielectric boundary (by overcoating the grating with a thin layer of dielectric) or for studying effects in semiconductors (by coating the grating with a semiconductor). Other information concerning the damping of electron waves in the metal may be obtained from the width of the resonance.

The most convenient effects to study from this point of view are

i) The total intensity of SPSL as a function of angle at constant wavelength.

ii) The zero order efficiency as a function of wavelength at constant angle of incidence.

The former is convenient to observe - it is only necessary to rotate the grating until the SPSL peaks and read off the angle, but one is restricted to laser wavelengths.

In the latter, the angle of incidence is set up, and the wavelength is scanned until the resonance is found, when a drop in response will be observed. The resonance wavelength is read off. This method allows measurements to be made over the whole visible region, but the groove spacings and angles of incidence must be carefully chosen.

Once an angle of incidence and a wavelength have been found for a resonance, the order may be deduced and a value of $\Gamma(\lambda)$ calculated.

Figure 6.22 shows dispersion curves for two Al gratings with different groove depths measured by technique (ii). The dotted line is the free-electron theory prediction. Figure 6.23 shows the dispersion curve for an Ag grating measured by a variety of methods. The dots represent measures by method (i) at the laser wavelengths (these were taken from figure 6.14) and the open circles represent measurements on other properties .

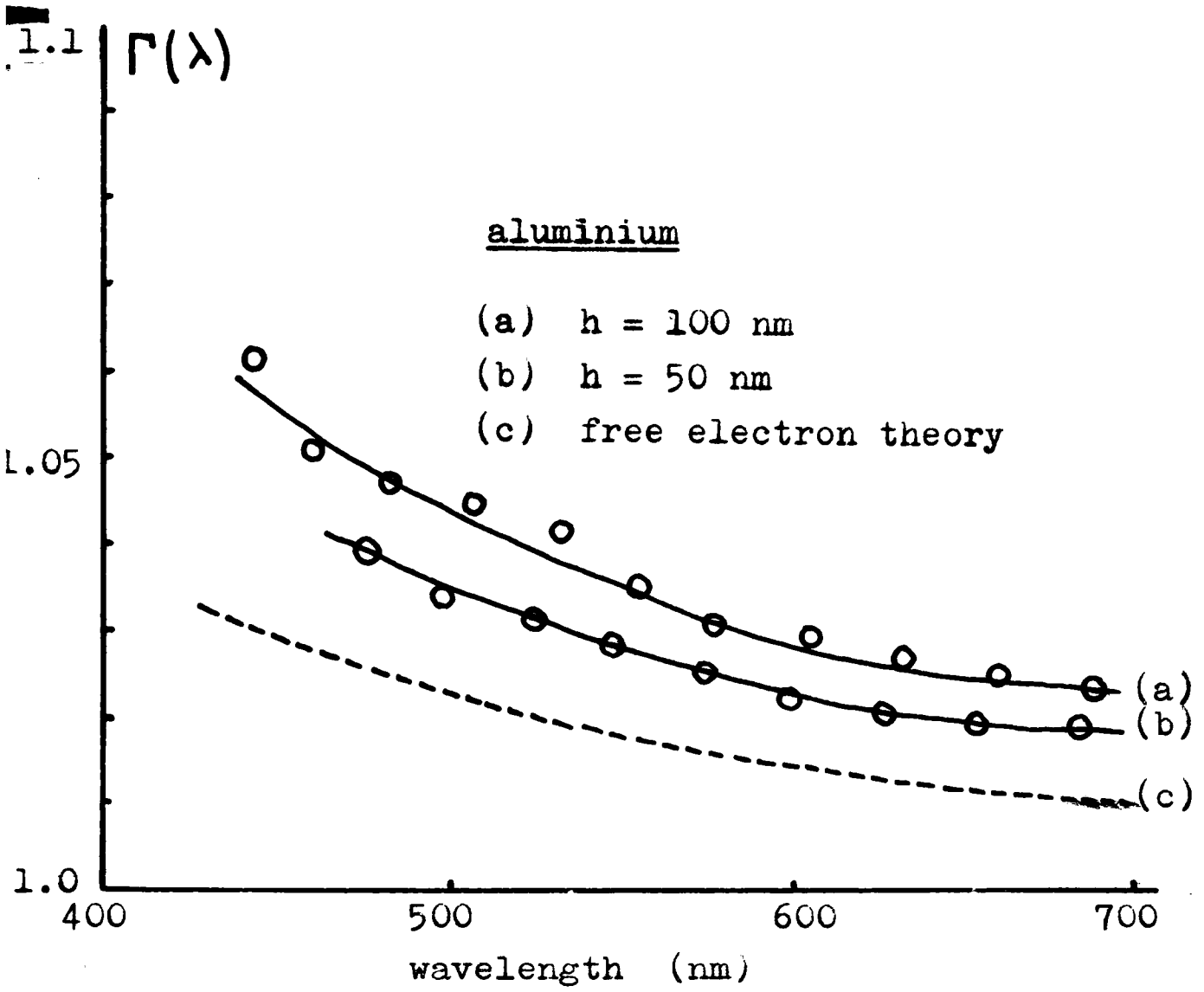


Fig. 6.22

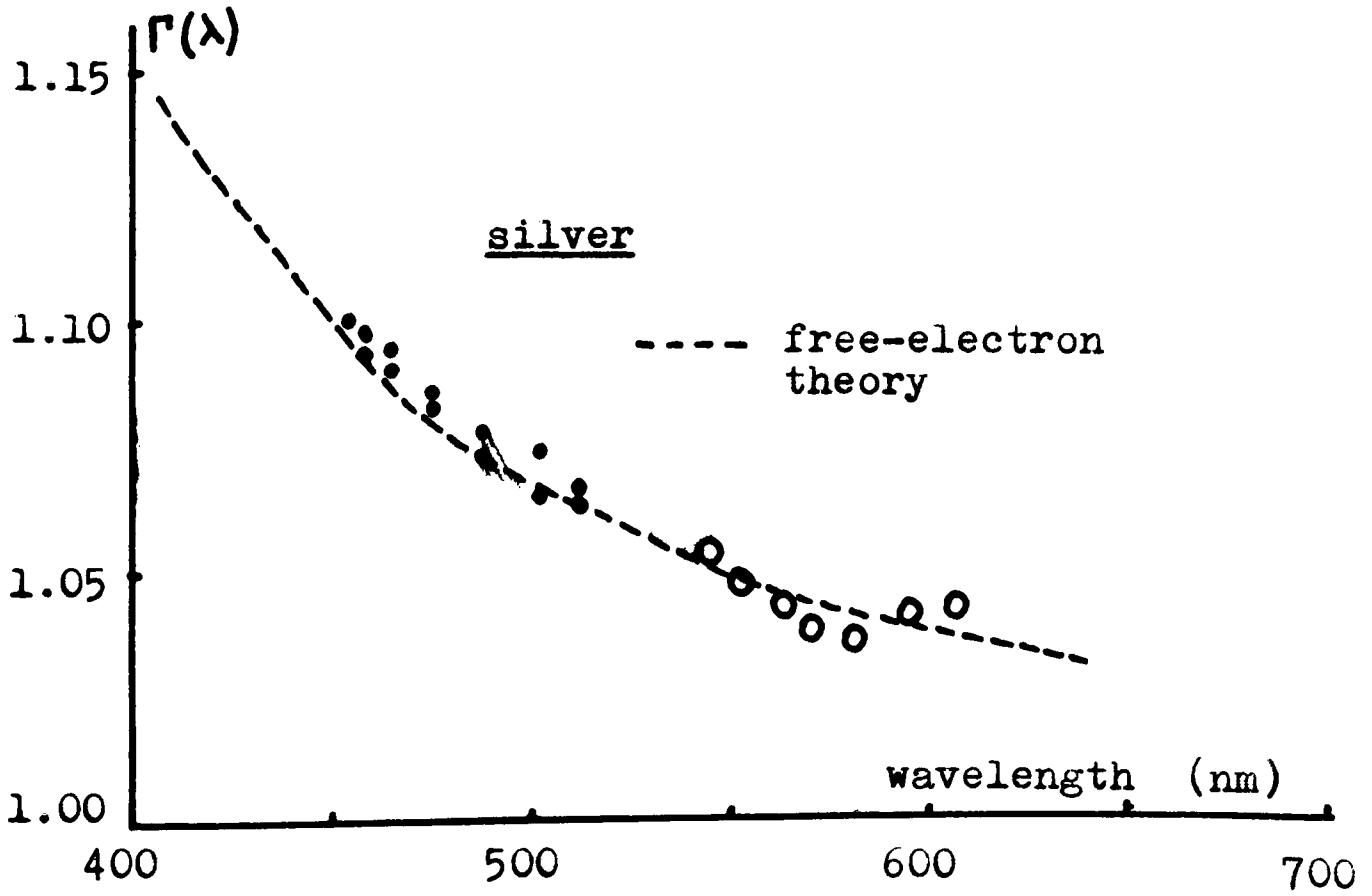


Fig. 6.23

Chapter 7

APPLICATIONS OF HOLOGRAPHIC DIFFRACTION GRATINGS

7.1 Radiation resistant gratings for use with synchrotron radiation

7.2 Gratings for tunable lasers

- 7.2.1 Dye lasers
- 7.2.2 CO_2 lasers

7.3 A wide range multichannel spectrometer using a composite HDG

7.4 Holographic zone plates

In the preceding chapters, HDG's have been treated mainly from a theoretical standpoint. It is apparent however, that they may be employed in spectroscopic instruments and may advantageously replace ruled gratings in some applications.

There are also applications of HDG techniques in which ruled gratings cannot be used.

In this chapter we discuss two applications of HDG's in situations where ruled gratings could be used (radiation resistant gratings for synchrotron radiation and gratings for tunable lasers) and two unconventional applications of HDG techniques (a composite grating for a multichannel spectrometer and holographic zone plates).

7.1 Radiation resistant gratings for use in synchrotron radiation

In response to a request from the synchrotron radiation users at the SRC electron synchrotron "NINA" at Daresbury, some experiments have been carried out on the performance of HDG's in the v.u.v. especially at high electron energies. The reason for this work was the fact that the replica ruled gratings used in the monochromators were suffering extreme damage in the form of blistering of the gold film after relatively short exposures to the beam.

This damage is assumed to be due to the breakdown of the replicating resin by the hard x-ray flux releasing gas which then expands and distorts the gold layer.

Preliminary experiments with photoresist had indicated that the resist might be less prone to this kind of damage. In addition the thickness of the organic layer in HDG's is less than the thickness of the epoxy layer in replica gratings, so it appeared that the use of HDG's might alleviate the problem.

To determine whether this was the case, a HDG was made as a direct replacement for the replica grating in the horizontally dispersing Wadsworth monochromator (46).

The specification of the grating was as follows:

radius of curvature:	2998.3 mm
ruling frequency:	1200 lines/mm
blaze wavelength:	75nm
blank diameter:	63.5 mm
ruled area:	32 x 52 mm^2
metal coating:	gold

In order to optimise the efficiency for 750Å, it is assumed that scalar theory is valid and so the efficiency is given by

$$\eta = RJ_1^2(x)$$

$$\text{with } x = \frac{2\pi h}{\lambda}$$

where R = reflection coefficient of metal

h = peak-peak groove depth

(see appendix 1)

This function has a maximum when x = 1.84, and so the optimum groove depth is given by

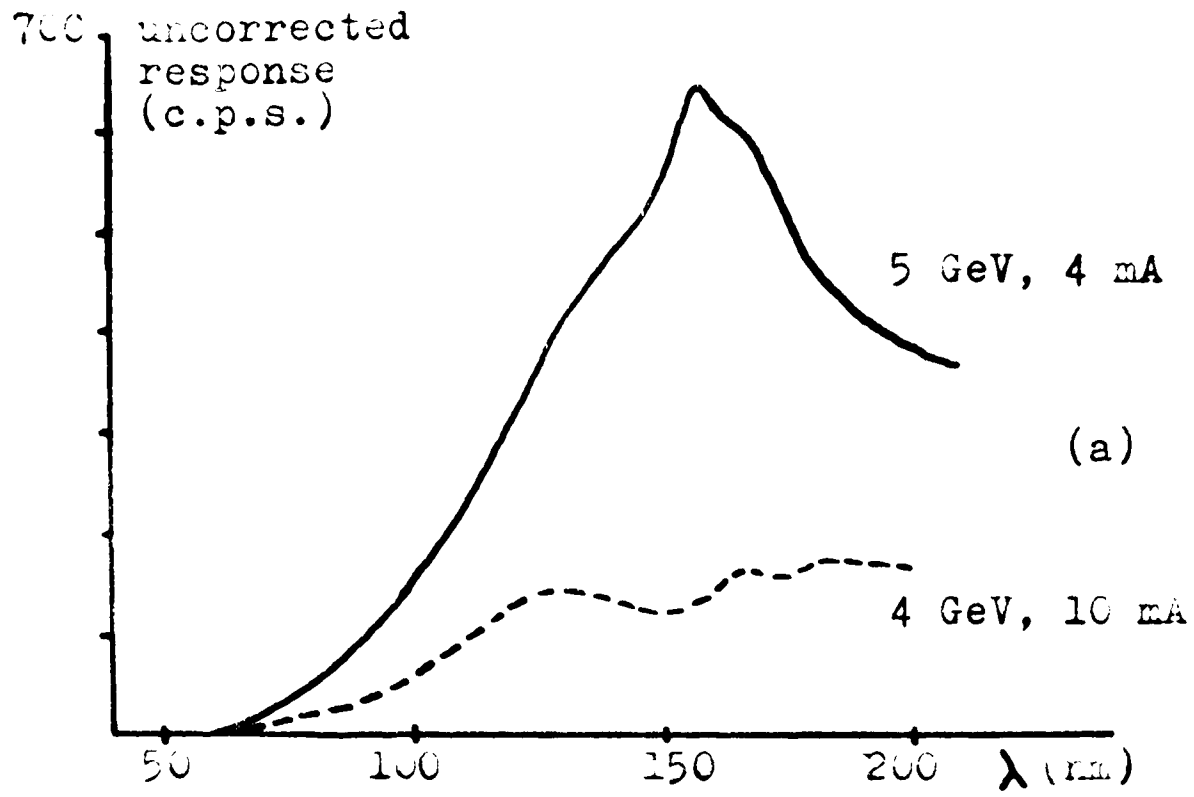
$$h = \frac{1.84}{2\pi}\lambda_p = 0.293\lambda_p$$

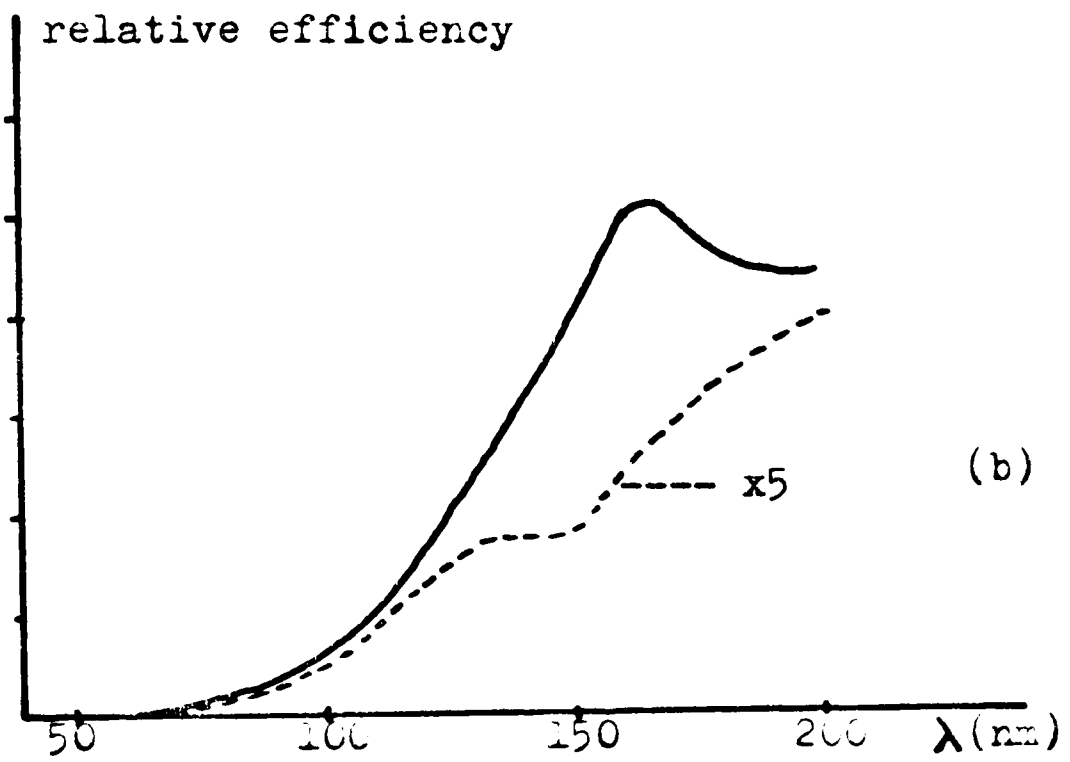
where λ_p is the wavelength of peak efficiency, Two gratings were made with $\lambda_p \simeq$ 150nm and $\lambda_p \simeq$75nm.

The results of the trial have been encouraging. The 150 nm grating has operated in the machine over a cycle of 5 weeks and has shown none of the blistering effects seen in replica gratings. During part of this time the grating has received radiation from 5GeV electrons, which would have damaged a replica in a short time.

The efficiency of the grating has been satisfactory, although a shift of λ_p to longer wavelengths has been observed. Figure 7.1 shows (a) the response of the grating to synchrotron radiation at different times during its exposure and (b) the responses corrected for the spectrum of the radiation, giving the relative efficiency of the grating.

Following the success of this experiment, a series of gratings has been made with peak wavelengths between 50 and 150 nm.





——— approx. 30 hrs exposure

------ approx. 100hrs exposure

Fig. 7.1 : Performance of a HDG in synchrotron radiation.
(Data kindly supplied by Dr. J.B.West (Daresbury Laboratory))

7.2 Gratings for tunable lasers

7.2.1 Gratings for pulsed dye lasers

In a nitrogen pumped dye laser (47), the very short excitation time (typically 10ns) means that only a few light passes can be made during the pulse.

This means that cavity modes cannot build up and so it is possible to tune the output by a filter of very narrow bandwidth. The losses of such a system are overcome by the high gain of the dye (≃30dB/mm).

In the dye lasers built in this laboratory (48), the tuning elements consit of a high dispersion HDG as a cavity end reflector and a single intra-cavity etalon.

The groove spacings of the gratings have been chosen to give the maximum possible dispersion while maintaining a reasonable angle of incidence. For wavelengths up to 500 nm a grating of 3600 lines/mm is used and for wavelengths up to 650 nm the spatial frequency is 2400 1/mm.

In order to take advantage of the full resolving power of the grating, the beam is expanded to a diameter of 50 mm. The gratings are 50 mm in diameter.

The bandwidth due to the grating alone is given by (47)

$$\Delta\lambda_G = \frac{\lambda^2}{2\pi w \tan \phi}$$

where w is the width of the expanded beam and ϕ is the Littrow angle of the grating.

Thus for a 3600l/mm grating at 500 nm, the bandwidth is ~0.04Å. This is further reduced by the use of an intra-cavity etalon.

HDG's are advantageous in this application for two reasons:

i) The high spatial frequencies can be obtained with no extra difficulty;

ii) The freedom of the gratings from groove errors means that a high resolving power is obtained.

The lack of blaze of HDG's is not a serious disadvantage because the gratings are working in the region where only the zero order and the order of observation are propagating. Under these conditions, efficiencies are high.

Mean powers of several mW have been obtained from the dye lasers.

7.2.2 Gratings for CO_2 lasers

In infra-red molecular lasers, the use of a grating is mandatory because of the difficulty of obtaining prisms with low loss and sufficiently high resolution to separate the rotational transitions (49). The gratings used are usually of the order of 80 lines/mm, blazed for the required wavelength.

Experiments have recently been undertaken to determine whether HDG's could be made with characteristics suitable for use in CO_2 lasers. If this proves possible, it would not only remove a major item in the cost of a CO_2 laser, but would permit the production of gratings with non-standard groove spacings on curved substrates designed to suit particular cavity configurations. The HDG process as described in chapters 2 and 3 is unsuited to making coarse gratings because:

i) it is difficult to obtain fringes with spatial frequencies less than 150 lines/mm.

ii) the deep grooves required are difficult to record.

The problem of forming the fringes was overcome by setting up the mirrors and the beam splitter as a Twyman-Green interferometer. Then

by suitable tilts of the mirros, spatial frequencies from 0 to around 200 lines/mm can be obtained over an aperture of several centimetres.

In order to record the deep grooves, a thick film of photoresist was built up by successive applications. Long development times (~10 mins) were used to obtain the required groove depths.

After development the gratings were gold coated.

As a preliminary experiment, a grating was made for use on the 10.6μm transitions of a CO_2 laser. The specifications were:

Substrate: 25 mm diam. plane

Ruling frequency: 100 lines/mm

Estimated groove depth: 1.5μm

Detailed information of the performance of the gratings is not yet available, but initial results have proved encouraging.

7.3 A wide range multichannel spectrometer using a composite HDG

A multichannel spectrometer combines the precision and speed of photoelectric detection with the advantages of a spectrograph. This is done by replacing the photographic plate of a spectrograph with an array of photodetectors which can be scanned electronically. In one commercial instrument (50) this consists of a silicon vidicon, in which case the detectors are scanned by an electron beam.

The advantages of this technique are:

i) The spectrum can be acquired very rapidly - the detectors can be scanned in less than 100 ms.

ii) Time dependant effects overthe whole spectrum may be studied, or alternatively, for very waak sources, the signal can be stored and integrated over long periods.

iii) The data is immediately compatible with computer processing.

The disadvantage of the instrument is that for a given detector configuration, the wavelength range and the resolution are inversely related (e.g. for the SSR instrument (50), the resolution is 1.5Å/channel over a range of 400Å and 6Å/channel over 1600Å). In order to increase the resolution while maintaining the range of coverage, the spectrum must be imaged in a format which is more compatible with the detector. This is done using the grating shown in figure 7.2



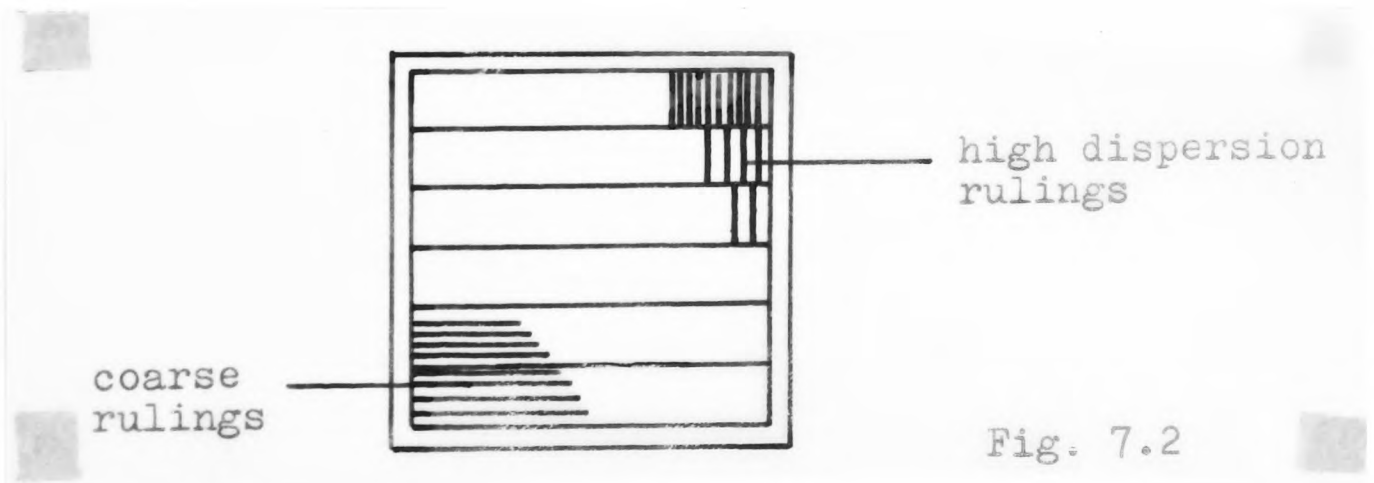


Fig. 7.2

The grating consists of a number of horizontal strips of grating with the rulings running vertically. The spatial frequency of each strip is chosen so that the spectra imaged onto the detector by each successive strip give continuous coverage of the spectrum

$$\text{i.e. } (\lambda_{max})_N = (\lambda_{min})_{N+1}$$

or if the ends of the detector subtend angles θ_1 and θ_2 at the grating and the angle of incidence is θ_i.

$$d_N (\sin \theta_i + \sin \theta_2) = d_{N+1} (\sin \theta_i + \sin \theta_1)$$

Vertical separation of the spectra formed by the strips is provided by an overall coarse ruling with the grooves horizontal.

A prototype grating was made by successive exposures on a single blank. The parameters were:

Number of strips: 6

Spatial frequency: d_1 = 2120 line/mm

d_6 = 1190 line/mm

Coarse grating frequency: 300 l/mm

When used in a Littrow configuration with a lens of 40 cm focal length, the spectral region 400-800 nm was imaged as six segments stacked one above the other in an area 40 x 40 mm^2, as shown in figure 7.3.

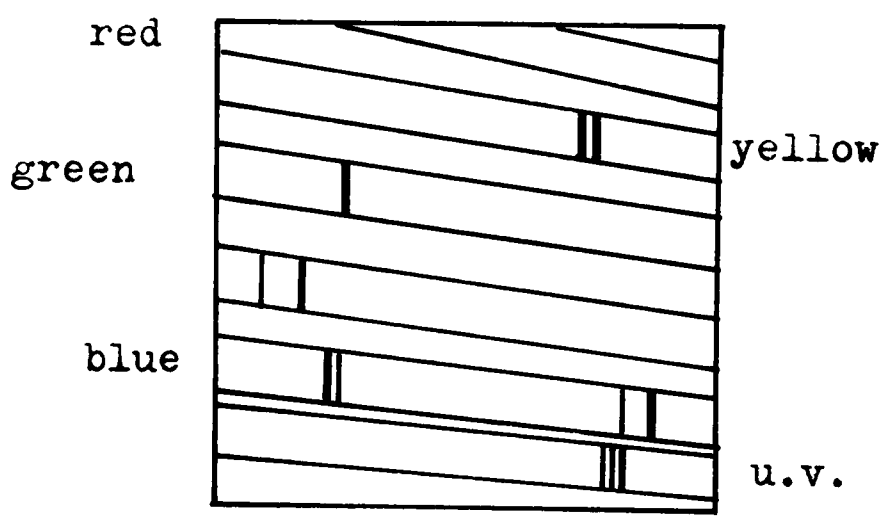


Fig. 7.3: Spectrum of a mercury lamp

This format is particularly convenient for use with a vidicon as the full face area can now be used. An instrument using this technique could be made very compact since no motion of the grating is required.

7.4 Holographic zone plates

In the spectral region below 180 nm, the only imaging device available is the Fresnel zone plate (51). Until recently, these devices have been made by conventional photofabrication techniques, but by using holographic techniques, it is possible to form zone plates with a significantly improved performance (52). As a preliminary to experiments with X-ray zone plates, several plates were made for use in the visible.

The principle of formation is the interference of two point sources located on the normal of the blank (see figure 2.1). If the distances of the sources from the blank are a and b respectively, then it can be shown that the focal length of the zone plate is given by:

$$f = \frac{\lambda_o}{\lambda}\left(\frac{1}{a} + \frac{1}{b}\right)^{-1}$$

where λ_o is the recording wavelength

and λ is the wavelength of use.

In order to get sufficiently short focal lengths in the v.u.v, a and b must be very short, so the optical system shown in figure 7.4 was used.

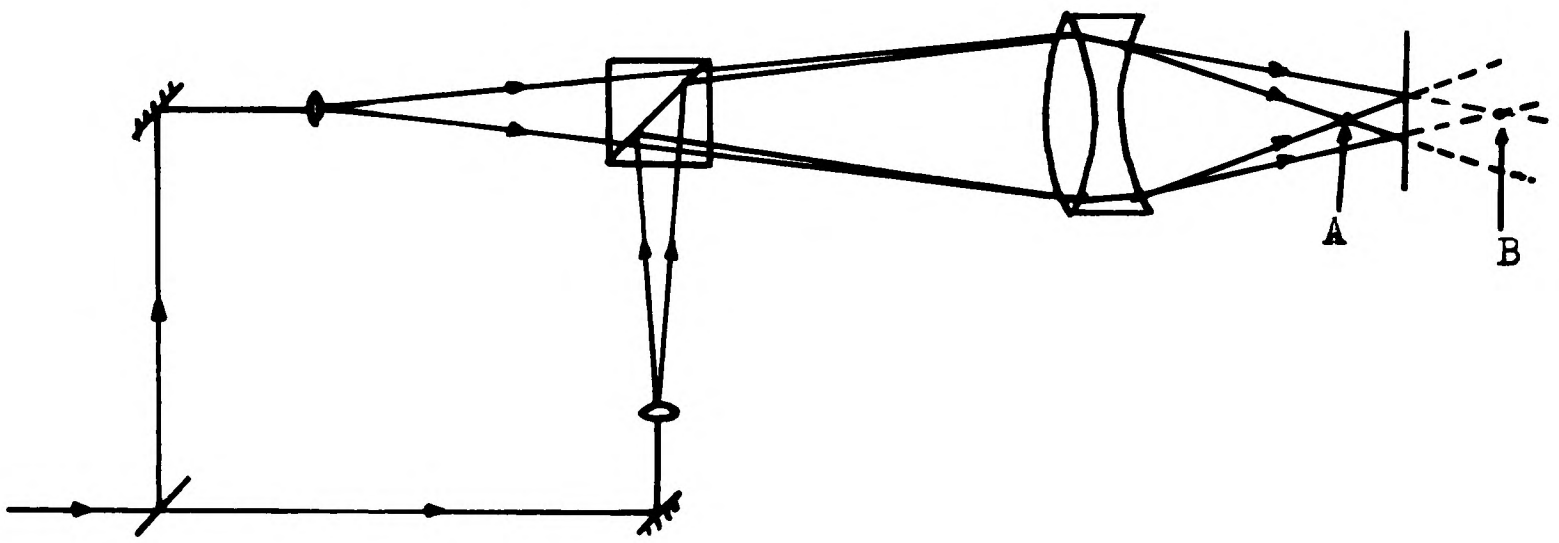
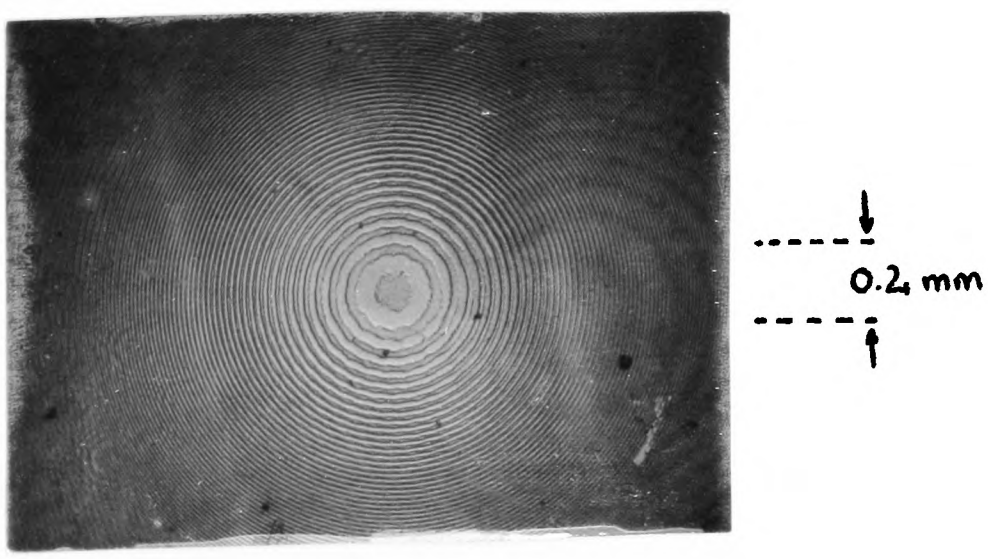


Fig. 7.4

The point sources are formed by two microscope objectives which expand the beams from the laser. The spherical waves are recombined by a semi-reflecting cube and are brought to a focus by a wide angle lens (f = 25 mm; f/0.9). The separation of the foci depends on the positions of the beam expanders.

The blank is mounted on a precision translation stage so that it can be set accurately between the sources which may be less than 1 mm apart.

Figure 7.5 shows a zone plate made with this system.

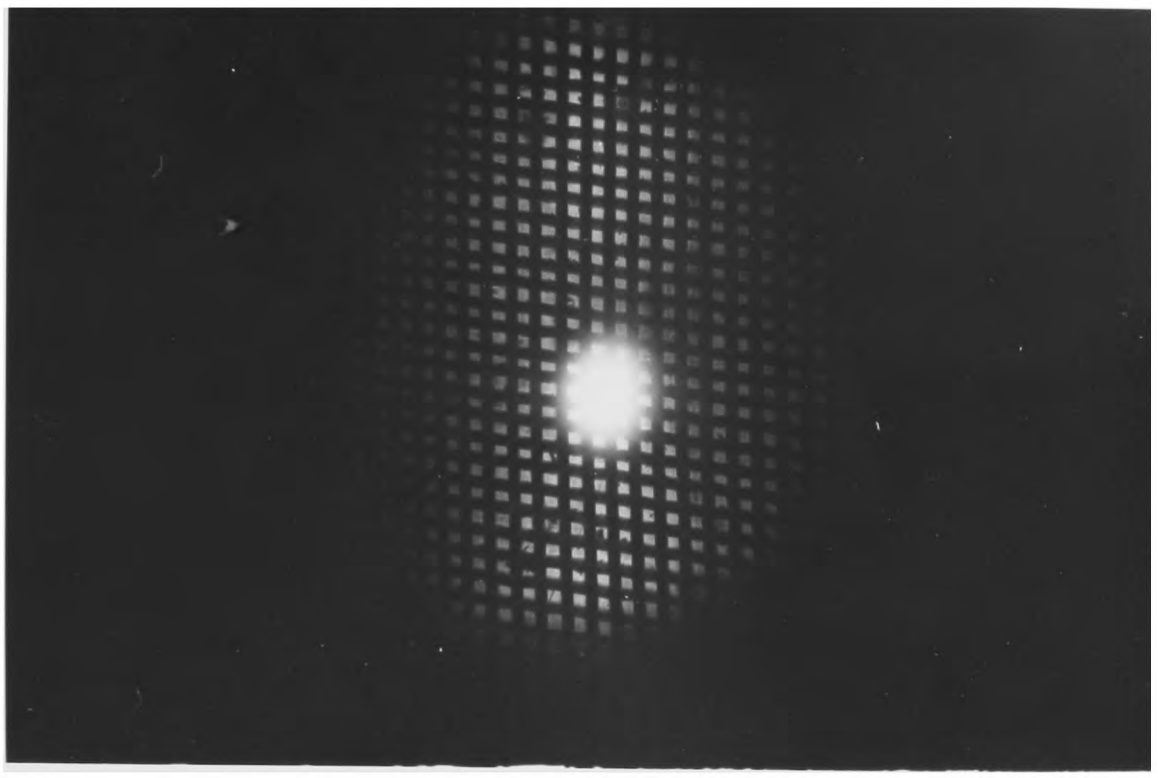


The rings are formed from the resist remaining after the exposed parts of the film have been etched away, and so in the visible, the device acts as a phase reversal zone plate.

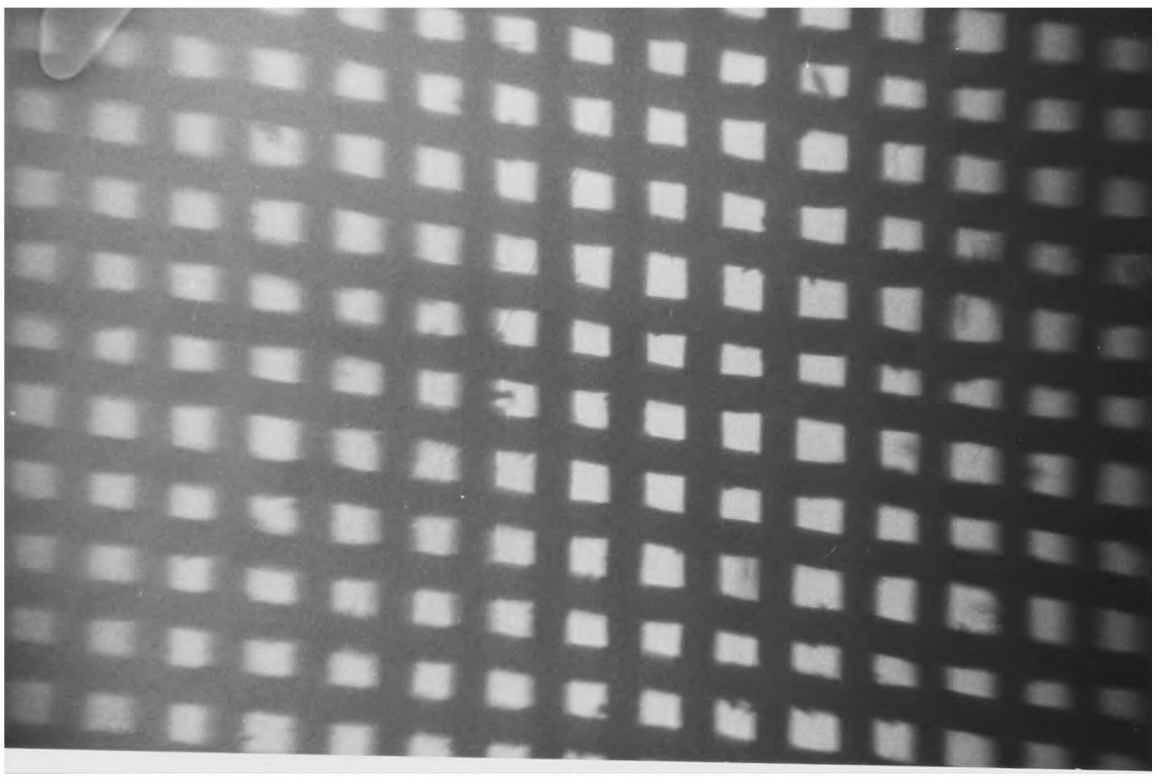
To test the performance of the zone plate, it was set up as a microscope to project an image of a 60μm wire gauze directly onto a film. Figure 7.6 shows prints of the negatives obtained using a wavelength of 546 nm.

The low magnification image (Fig.7.6(a)) shows the bright zero order, which is a disadvantage of all zone plate systems.

In figure 7.6(b), the zone plate has been moved closer to the gauze to increase the magnification. The off-axis aberrations of the system now become apparent.



(a)



(b)

Fig. 7.6

APPENDICES

1. Scalar Theory of Grating Efficiency

2. The grating equation and anomalies in constant deviation mountings

3. Measurement of spatial frequency

4. The holographic reconstruction of scattered light

Appendix 1

Scalar theory of grating efficiency

When the wavelength of the radiation is much smaller than the groove spacing of the grating, the diffraction process may be treated by a scalar theory in which the amplitude of the light at any point is represented by a single scalar function.

Consider the arrangment of figure A1.1.

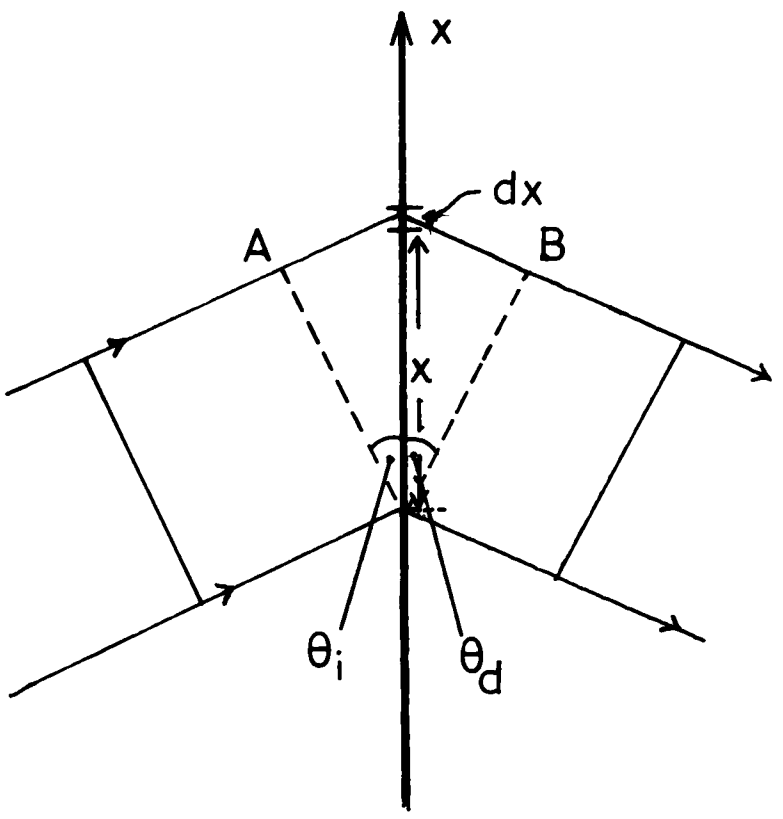


Fig. A1.1

This represents the diffraction of a plane monochromatic wave incident at an angle θ_i by one-dimensional screen whose complex amplitude transmission function is T(x).

Each small element of the screen contributes an amount $d\psi$ to the total amplitude diffracted at an angle θ_d, where

$$d\psi = T(x)\ \exp - ikx\ (\sin\theta_i + \sin\theta_d)$$

Here the exponential term represents the phase introduced by the extra path length $[AB]$, and $k = 2\pi/\lambda$.

Thus the total amplitude of the diffracted wave is given by

$$\psi(u) = \int_{-\infty}^{\infty} T(x)\ \exp(-iux)dx$$

where

$$u = K(\sin\theta_i + \sin\theta_d)$$

This expression is the Fourier transform of T(x).

If T(x) is periodic, the calculation may be simplified by expressing it as the convolution of the function taken over one period with an array of δ-functions with a spacing equal to the groove period. Then by use of the convolution theorem (43), the transform of the convolution may be written as the product of the transforms of the two functions. This gives the following expression:

$$\psi(u) = F\left[T(x)\right]_o^d \times F\left[\sum_0^N \delta\ (x-nd)\right]$$

where $F[\ ]$ denotes the Fourier transform and N is the total number of grooves.

The second term in the expression is evaluated in many text books (43) and represents the effects of interference. This term gives a series of constructive interferences or orders when

$$u = \frac{2n\pi}{d}$$

where n is an integer

i.e. when

$$\sin\theta_i + \sin\theta_d = \frac{n\lambda}{d}$$

This is the well known grating equation. The first term of the

expression gives the amplitude of the order and so may be used to calculate the efficiency of the grating. viz:

$$\eta_n(u) = |\psi_n(u)|^2 = \left| \int_o^d T(x)\ \exp(-iux)dx \right|^2$$

when $u = \frac{2n\pi}{d}$

Assuming that HDG's are sinusoidal phase gratings T(x) may be written:

$$T(x) = \exp -i(\phi_o + \phi_m \sin \frac{2\pi x}{d})$$

The Fourier transform of this is a standard form, giving

$$\eta_n(\lambda) = J_n(\phi_m)^2$$

where $J_n(x)$ is the nth order Bessel function of argument x.

Here ϕ_m is the amplitude of the phase variation. For a reflection grating

$$\phi_m = \frac{2\pi h}{\lambda}$$

where h is the peak-to-peak groove depth and a transmission grating

$$\phi_m = \pi(n-1)\frac{h}{\lambda}$$

where n is the refractive index of the grating material.

Appendix 2

The grating equation and the occurrence of Wood's anomalies in constant deviation mountings

In the constant deviation mounts, the directions of the incident and diffracted beams are fixed and the grating is rotated to change the wavelength passing through the exit slit. When the angle between the two beams is zero, the arrangement is known as a Littrow mount. Otherwise it is an Ebert or a Czerny-Turner mounting. The quantities used to describe a Czerny-Turner are defined in figure A2.1.

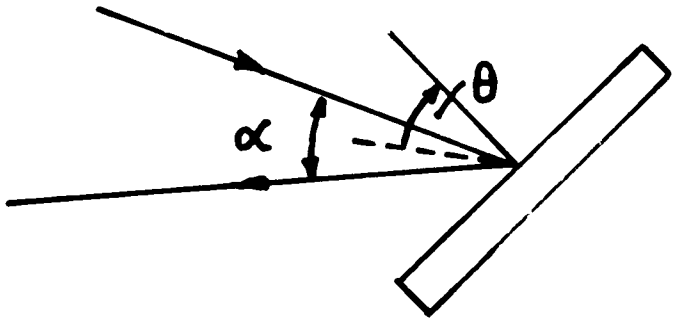


Fig. A2.1

It is assumed that observation is to be made in nth order at wavelength λ with a grating of period d. θ is the angle of rotation of the grating normal from the zero order position (when it bisects the angle between the beams) and α is the angle between the beams.

Thus the angle of incidence is

$$i = -\theta - \alpha/2$$

and the angle of diffraction

$$\delta = -\theta + \alpha/2$$

Putting these quantities into the grating equation gives:

$$\sin(-\theta-\frac{\alpha}{2}) + \sin(-\theta+\frac{\alpha}{2}) = \frac{n\lambda}{d}$$

This reduces to the required equation

$$\frac{\lambda}{d} = -\frac{2}{n} \sin\theta \cos\frac{\alpha}{2}$$

In the particular case of the Littrow mounting, when $\alpha = 0$, this becomes

$$\frac{\lambda}{d} = -\frac{2}{n} \sin\theta$$

Wood's Anomalies

These rapid variations of efficiency occur when an order other than the order of observation is diffracted at $\pm 90^{o}$.

If the passing off order is m, then the grating equation is

$$\sin i \pm 1 = \frac{m\lambda}{d}$$

But $i = -(\theta+\alpha/2)$

Thus $- \sin(\theta+\alpha/2) \pm 1 = \frac{m\lambda}{d}$

Also, if the order of observation is n, we have

$$-\sin\theta \cos\frac{\alpha}{2} = \frac{n\lambda}{2d}$$

Elimination of θ between these equations yields a quadratic equation in λ/d. The roots of this are

$$\frac{\lambda}{d} = \left| \frac{(m-\frac{n}{2}) \pm \sin\frac{\alpha}{2}\left[(m-\frac{n}{2})^2 - n^2\right]^{\frac{1}{2}}}{(m-\frac{n}{2})^2 + n^2 \tan^2\frac{\alpha}{2}} \right|$$

This expression gives the wavelength at which the mth order anomaly is observed in nth order.

In the Littrow mount, when $\alpha = 0$, this becomes

$$\frac{\lambda}{d} = \left| \frac{1}{m - \frac{n}{2}} \right|$$

It should be noted that for a given value of n, this will give the same value of λ/d for two values of m (e.g., n = 1, m = -1 or +2). Thus in the Littrow mounting, two anomalies occur simultaneously.

As α increases from zero, the Rayleigh wavelengths for the two values of m move apart.

Appendix 3

Measurement of groove spacing

It was necessary to measure the spatial frequency of the gratings made in this work, sometimes to a high accuracy ($\ll$ 1 part in 10^3).

This was done by measuring the autocollimation angle of the grating for a spectral line of known wavelength. The instrument used was a Zeiss autocollimating spectrometer, which has a turntable with a scale which may be read to 10 seconds of arc.

The procedure was carried out as follows:

i) The instrument was aligned with a low pressure Hg discharge. The grating was mounted and adjusted until its grooves were parallel to the axis of rotation. The lens of the telescope was adjusted until the spectral lines were in focus in the plane of the cross-wires.

ii) The angles of autocollimation in positive and negative nth orders of a wavelength λ were measured, θ_+ and θ_-.

iii) The autocollimation angle in nth order is then given by

$$\theta_n = \frac{\theta_+ + \theta_-}{2}$$

and the groove spacing is given by

$$d = \frac{n\lambda}{2 \sin\theta_n}$$

The error in this measurement is given by

$$\frac{\Delta d}{d} = - \frac{\Delta\theta_n}{\tan\theta_n}$$

Thus for $\theta_n = 45^o$ and $\Delta\theta = 10''$ or 4×10^{-5} radians, the fractional error in d is of the order of 1 part in 10^4.

Appendix 4

The holographic reconstruction of scattered light

While investigating the angular distribution of scattered light from HDG's, a set of arcs was discovered whose occurrence could not be explained either as Rowland ghosts or as surface plasmon effects.

They were eventually found to be similar to the two ring diffraction phenomenon observed by Moran and Kaminow (44) and later explained by Forshaw (45). These rings were shown to be due to volume effects in thick media caused by the interference of light scattered within the medium with the incident recording beam. It was shown in (45) that the wave vector of the scattered light is given by

$$\bar{S}_i = \bar{S}_m - \bar{S}_R + \bar{S}_c \qquad \text{(primary image)}$$

$$\bar{S}_i = -\bar{S}_m + \bar{S}_R + \bar{S}_c \qquad \text{(conjugate image)}$$

where

$\bar{S}_i$ = wave vector of scattered light

$\bar{S}_R$ = " " of recording beam

$\bar{S}_m$ = " " of beam caused by scattering of S_R within the medium.

$\bar{S}_c$ = wavevector of light illuminating the developed hologram.

$$|S_R| = |S_m| = 2\pi/\lambda_o$$

$$|S_i| = |S_c| = 2\pi/\lambda$$

where λ_o is the recording wavelength

and λ is the reconstructing wavelength

In the case of HDG's, it is assumed that the scattering is confined to the plane of the resist. We may then write:

$$\bar{S}_R = \pm \frac{2\pi}{\lambda_o} \sin\theta_R \qquad \text{normal to the grooves}$$

$$\bar{S}_m = \frac{2\pi}{\lambda_o} \qquad \text{in any direction}$$

$$\bar{S}_c = \frac{2\pi}{\lambda} \sin\alpha \qquad \text{normal to the grooves}$$

and $$\bar{S}_i = \frac{2\pi}{\lambda} \sin\theta$$

where θ_R is the recording angle,

α is the angle of incidence of the reconstructing beam

θ is the angle between the scattered light and the grating normal

We must also include an extra term of $2\pi n/d$ to allow for the effect of the diffraction grating.

Under these conditions, the equations reduce to

$$\frac{2\pi}{\lambda_o} \sin\theta = \frac{2\pi}{\lambda_o} \pm \frac{2\pi}{\lambda} \sin\theta_R + \frac{2\pi}{\lambda} \sin\alpha + \frac{2\pi n}{d}$$

We recall that for a HDG, the groove spacing d is given by

$$d = \frac{\lambda_o}{2\sin\theta_R}$$

and so

$$\sin\theta = \frac{\lambda}{\lambda_o} + \frac{\lambda}{d}\left(n \pm \frac{1}{2}\right) + \sin\alpha .$$

To find sin θ, a graphical construction similar to that used in 6.32(b) for the SPSL lines may be used. Using the same co-ordinate system, a circle of radius λ/λ_o is drawn with centre on the $\phi = 0$ axis at a distance $\frac{\lambda}{d}(n \pm \frac{1}{2}) + \sin\alpha$ from the origin.

The value of sinΘ for any value of Ø is now found by the intersection of the Ø radius with the circle. Figure A4.1 shows the construction for a typical case.

This theory has been found to account well for the positions of the sharp arcs inthe scattered light distribution.

The cause of the scattering in the film is likely to be air bubbles or dust in the photoresist or scratches on the glass surface. Effects are also seen due to scattering from the edges of the blank.

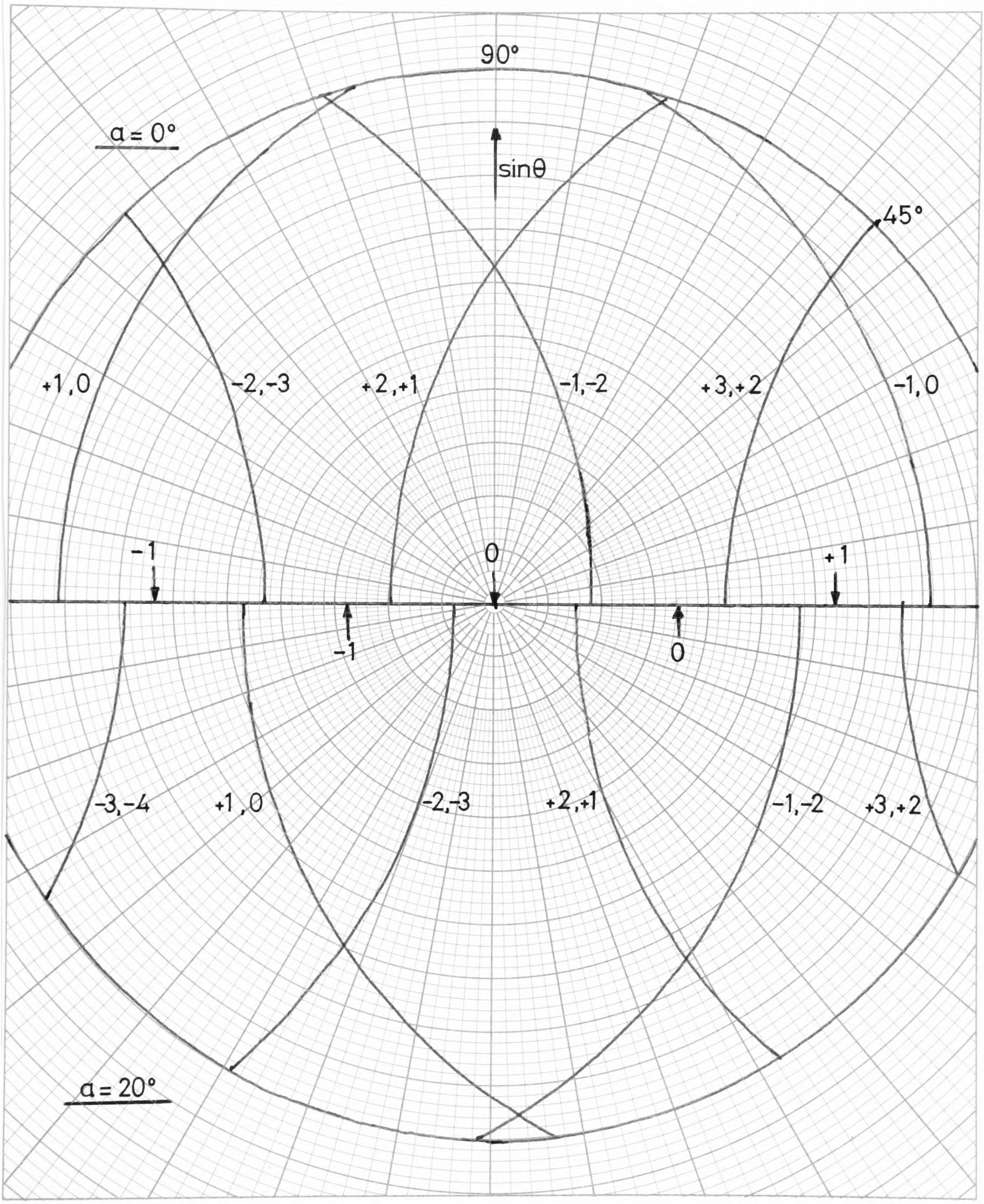


Fig. A4.1 : Construction for holographic scattered light, showing values of n and positions of orders.
1200 l/mm; λ = 514.5 nm; α = 0° and 20°

REFERENCES

1. J.M.Burch. "Photographic production of scales for Moire fringe applicatinns", Brussels colloquium on Optics in Metrology, May 1958. (London: Pergamon Press).
2. J.M.Burch. Research, Lond., 13, 3 (1960).
3. J.M.Burch and E.W.Palmer. Optica Acta, 8, 73 (1961).
4. N.George and J.W.Matthews. Appl. Phys. Letts. 9, 212 (1966).
5. T.A.Shankoff and R.K.Curran. Appl. Phys. Letts, 13, 239 (1968).
6. D.Rudolph, and G.Schmahl. Umschau in Wissenschaft und Technik, 7, 225, (1967).
7. A.Labeyrie and J.Flamand. Opt.Commun. 1, 5, (1969).
8. M.Pouey. "Imaging Properties of Ruled and Holographic Gratings". Invited Paper at ICO conference on Spectroscopic Instruments. Tokyo. Aug. 1975.
9. H.Noda, T.Namioka, and M.Seya. J.Opt.Soc.Am. 64, 1031, 1037, and 1043 (1974).
10. M.C.Hutley. Opt.Act. 20, 607, 1973.
11. M.C.Hutley and V.M.Bird. Opt.Act. 20, 771 (1973).
12. M.C.Hutley. Opt. Act. 22, 1, (1975).
13. R.J.Collier, C.B.Burckhardt and L.H.Lin. "Optical Holography" Academic Press, New Y rk, 1971). p.156.
14. Ibid., p.169.
15. Ibid., p.164.

16. G.W.C. Kaye and T.H.Laby. "Tables of Physical and Chemical Constants" 13th ed. (Longmans, London 1968), p.80.

17. G.Schmahl, and D.Rudolph. Optik, 30, 606 (1970) (German).

18. D.Rudolph and G.Schmahl. Contribution to the one-day meeting on diffraction gratings. Imperial College. September 1974 (unpublished).

19. J.F.Verrill. J.Phys.E. 6, 1199 (1973).

20. R.S.Longhurst, "Geometrical and Physical Optics" (Longmans, London) 2nd ed. 1967, p.155.

21. Shankoff, T.A. App.Opt. 7, 2101 (1968).

22. J.P.Kirk and G.L.Fillmore. App.Opt. 11. 2347 (1972).

23. M.T.Gale and D.L.Greenaway. RCA internal report number ZRRL-71-TR-022 (1971).

24. M.J.Beesley and J.G.Castledine. Appl. Opt. 9, 2720 (1970).

25. R.W.Wood. Phil. Mag. 4, 396,(1902).

26. C.V.Shank and R.V.Schmidt. Appl. Phys.Letts. 23, 154, 1973.

27. G.F.Damon. Collected papers from Kodak seminars on micro-miniaturization 1965/1966. p.34. (Eastman Kodak Company, 1969).

28. Lord Rayleigh, Phil. Mag. 14, 60, 1907.

29. M.Born and E.Wolf. "Principles of Optics" (Pergamon Press, London, 1959), Chap. VIII.

30. A.Hessell and A.A.Oliner. Appl. Opt. 4 1275 (1965).

31. D.Maystre. Opt.Commun 6, 50,(1972).
Opt. Commun 8, 216 (1973).

32. R.C.McPhedran, I.J.Wilson and M.D.Waterworth. Optics Commun. 7, 331, (1974).

33. R.C.McPhedran and D.Maystre. J.Spect.Soc.Japan 21, Supplement 1, p.47 (1975).

34. C.Kittel. "Introduction to solid state physics", 3rd ed. (Wiley, N.Y. 1968). Chapter 8.

35. W.Steinmann. Phys.Stat.Sol. 28, 437 (1968).

36. R.C.McPhedran and D.Maystre. "Etude theorique des anomalies de diffraction des reseaux holographiques". To appear in Nouvelle Revue d'Optique Appl.

37. R.H.Ritchie, E.T.Arakawa et al. Phys.Rev.Letts. 21, 1530, (1968).

38. E.A.Stern. Phys. Rev.Letts. 19, 1321, (1967).

39. G.W.Stroke. "Diffraction gratings" in Handbuch der Physik (ed. S.Flugge) (Springer Verlag 1967) vol. 29, p.588.

40. R.C.McPhedran and D.Maystre. Opt. Act. 21, 413, 1974.

41. R.C.McPhedran. Private communication (1973).

42. R.C.McPhedran and M.D.Waterworth. Opt.Act. 19, 877 (1972).

43. S.G.Lipson and H.Lipson. "Optical Physics" (C.U.P. 1969), Ch.3.

44. J.M.Moran and I.P.Kaminow. Appl. Optc. 12, 1964 (1973).

45. M.R.B.Forshaw. App.Opt. 13, 2, 1974.

46. I.H.Munro. International Symposium for Synchrotron Radiation Users, Daresbury, Jan. 1973. p.22.

47. T.W.Hänsch. Appl. Opt. 11. 895 (1972).

48. P.E.G.Baird; K.Gardner. Clarendon Laboratory, Oxford.

49. R.McClellan and F.Denton. Optical Spectra, July/August (1968) p.49.

50. Princeton Applied Research Corp., Princeton N.J. Optical Multichannel Analyser (OMA) system.

51. A.V.Baez. J.Opt.Soc.Am. 51, 405 (1961).

52. B.Niemann, D.Rudolph and G.Schmahl. Opt.Commun. 12, 160 (1974).