**Precision laser beam measurements**

This item was submitted to Loughborough University’s Institutional Repository by the/an author.

**Additional Information:**

- A Doctoral Thesis. Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University.

**Metadata Record:** [https://dspace.lboro.ac.uk/2134/13794](https://dspace.lboro.ac.uk/2134/13794)

**Publisher:** © B.A. Omar

Please cite the published version.
This item was submitted to Loughborough University as a PhD thesis by the author and is made available in the Institutional Repository (https://dspace.lboro.ac.uk/) under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/
Precision laser beam measurements

by

Basil Arthur Omar

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of the Loughborough University of Technology.

May 1990

© by B.A. Omar 1990
Acknowledgements

I would like to thank all members of the Opto-electronics group for the many useful discussions, and in particular my supervisor Dr. David Emmony, for his guidance throughout the project, a special mention also going to Dr. Stuart Clark and Alan Holloway for their advice on a range of topics. I would also like to thank Dr. Noel Kerr who collaborated on the laser induced ripple measurements described in chapter 6. My appreciation goes to the rest of the students and staff of the Physics department most of whom at some point gave useful advice or lent equipment in support of my research.

I would like to thank the Rutherford Appleton laboratory in the guise of Dr. Mick Shaw for the advice and support that I received, together with David Hollis, formerly of the Department of Glasses, Polymers and Ceramics, University of Sheffield, for the supply of the glass samples, towards the development of the laser beam profiling system, described in chapter 1.

My thanks goes to Peter Roderick of 3M (UK) Ltd. for the supply of new and suitably worn video recording-heads for measurement and for the many useful discussions on the results obtained.

The work described in this thesis was supported financially by the Rutherford Appleton Laboratory and 3M (U.K.) Ltd jointly.

Thanks goes to Dr. David Sykes and Dr. Alison Chew of Loughborough Consultants, Surface Analysis Group, for the supply of the SIMS crater samples used in chapter 6.

Finally, I would like to thank all those who helped at the original manuscript stage including David Sheerin and Barry Ward for their help in the preparation and proof reading of the typed texts and to Joanna Edwards for her help with the diagrams.
ABSTRACT

The work described in this thesis is concerned with two main areas of investigation. The first involves the measurement and characterisation of the fluorescence of various doped glasses when excited by a pulsed ultra-violet laser beam, with a view to finding a material which acts as a suitable ultra-violet to visible image converter. A system is described, based on a glass fluorescer, which writes the beam profile of a single-shot KrF laser directly into computer memory and hence permits powerful image processing, and measurements to be made on the laser beam profile. The system was developed primarily for the spatial profiling of 'Sprite', Europe's largest ultra-violet laser, and is currently in routine use at the Rutherford Appleton Laboratory for this purpose.

The second area deals with the development of high resolution optical surface measurement systems. The main purpose of this work is to characterise the surface topology of a magnetic head from a video-recorder, and to measure any subsequent wear arising due to the passage of video-tape across the surface of the head. Two different systems are proposed and built. The first is an optical probe-differential, phase-quadrature, surface-profiling interferometer, in which the phase calculation and sample scanning is performed automatically by a host P.C. type computer. The system attains a height sensitivity of the order of 1nm and lateral resolution of about 10μm with measurements limited to flat surfaces. The second system is a phase-shifting interferometer based on a Twyman-Green optical arrangement, with a sensitivity of better than 10nm, and the capability of profiling highly curved surfaces, such as video-heads. Results taken with this system show it to be a major development in the measurement and understanding of video-head wear, and the system is in current use at the 3M (UK) Ltd. Research Laboratories, Swansea, for the purpose of studying the abrasivity and head-wear characteristics of video-tape. In addition to the measurement of video-head wear, the systems were successfully used for the precision measurement of various other shapes and surface structures including laser-induced ripple structures on the surface of resolidified semi-conductors. These measurements are used to confirm a model proposed to describe the mechanism of the ripple formation.
To my Parents

iii
"When you can measure what you are speaking about and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind."

LORD KELVIN (1824-1907)
PREFACE

This thesis describes work undertaken in the Opto-electronics group at Loughborough University of Technology. The funding for laser measurements, the subject of the thesis; was divided in time between two projects associated with measurements of the laser beam itself and the use of lasers to measure surface profiles. Much of the support electronics and computer work is common to both areas of work.

The first chapter stands as a self contained unit dealing with the measurement of single-shot, high-powered ultra-violet laser-beam spatial profiles. The chapter discusses the importance of these measurements and includes a review of a range of measurement techniques that are employed for this purpose. The fluorescence of specially doped glasses is investigated with a view to their use as uv-to-visible image converters in a video-based profiling system. The hardware and software of a system developed to spatially profile the Rutherford Appleton Laboratory's laser 'Sprite' is discussed, together with results which highlight the capabilities and performance of the system.

Chapter-two forms the start of the second part of this thesis, which deals with the developement and use of precision optical measurement techniques primarily for the measurement of video recorder-head wear patterns. Chapter-two begins with a brief history of the developement of magnetic tape recording from the early audio systems of the late 19th century through to present day domestic video-recorders. The basic principles of video tape-recording are described, together with a description of a typical video-recorder head. Head-wear and the subsequent increase in seperation of head and tape during normal operation with its adverse effect on the recorded signal is discussed. This leads to a review of the basic principles involved in the wearing of two contacting rough surfaces, and the most likely mechanisms involved in the wearing of video-heads. Also given in this chapter is a review of the methods and techniques developed and used by various researchers for the assesement of the life expectancy of magnetic-head materials, and the measurement of the relative abrasivities of magnetic tapes.

In the third chapter a unique optical probe surface profiling
system is proposed for the measurement of the wear on video heads. The chapter begins with a review of traditional surface profiling methods, and proceeds to a comprehensive description of the principles and workings of the proposed phase-quadrature differential-probe interferometer. Results of surface measurements taken with the system are shown; these clearly demonstrate its excellent performance when operated within the constraints imposed by the depth of focus of the imaging optics. Although these conditions are satisfied during the profiling of nominally flat samples, highly curved surfaces, such as video-heads, require sophisticated translation and rotation stages to position the sample with sufficient precision. Despite the fact that these problems are not theoretically insurmountable, the financial limitations of the project required a neater and more cost effective approach.

It was decided that phase-shifting interferometry (PSI) offered the solution to the problem. In chapter four, the principles of PSI are explained. An extensive review is given of the many ways of producing a known optical phase-shift and methods employed to measure the phase of a detected signal. The chapter also reviews the algorithms used by the most common analytical phase detection methods, and compares them. Factors which affect the sensitivity and accuracy of PSI measurements are discussed and methods of reducing their significance and limiting phase errors are given.

Chapter five describes the development of an optical non-contact surface-profiling instrument which employs phase-shifting interferometry as the basis for measurement. The system was designed to overcome the problems of profiling highly curved objects with high sensitivity from which other commercially available PSI systems suffer. The system is used to measure the phase profiles of a video-head before and after the passage of tape across its face. The change in its phase profile is then shown as a wear pattern. Results of the head-wear patterns given by the system are the first known examples of these measurements and highlight the importance of this system in the study of video-head/tape interactions.

Finally, chapter six describes other applications of the phase-shifting interferometer system, mainly in the measurement of the surface topography of resolidified metals and semi-conductors after irradiation by high powered laser pulses. These measurements are used to
verify a model proposed to explain the formation of these surface ripple structures. Another use of the system is the measurement of the depth and shape of craters on the surface of silicon wafers caused by a surface analytical technique termed secondary ion mass spectroscopy. These crater-depth measurements are important in determining the presence of various chemical elements as a function of depth in the sample. The last application of the system is its use for the surface profiling of steel ball-bearings.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>i</td>
</tr>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iii</td>
</tr>
<tr>
<td>Quote</td>
<td>iv</td>
</tr>
<tr>
<td>Preface</td>
<td>v</td>
</tr>
</tbody>
</table>

## CHAPTER ONE

THE USE OF DOPED GLASS FLUORESCERS IN A SYSTEM FOR THE MEASUREMENT OF THE SPATIAL PROFILES OF HIGH POWERED ULTRA-VIOLET LASERS.

1.1 Introduction.  
1.2 Review.  
1.2.1 Photography.  
1.2.2 Beam-scanning.  
1.2.3 Linear and 2-D detector arrays.  
1.3 Ultra-violet and visible converters.  
1.4 Experimental methods.  
1.4.1 Fabrication of glass samples.  
1.4.2 Molar composition of the host glasses and their preparation conditions.  
1.4.3 Summary of glass samples tested.  
1.5 Fluorescence measurements.  
1.5.1 Time response and conversion linearity.  
1.5.2 Fluorescent spectra.  
1.5.3 Results and discussions.  
1.6 Application: Laser beam spatial profiling system.  
1.6.1 Video-format.  
1.6.2 Video camera.  
1.6.3 Video-framestore and interface card.  

1
CHAPTER TWO

VIDEO RECORDING AND VIDEO-HEAD WEAR REVIEW.

2 Introduction 39
2.1 Video tape recording. 39
2.1.1 Principles of magnetic tape recording. 40
2.1.1.1 Magnetic tape recording head. 41
2.1.1.2 Hysteresis loop. 41
2.1.1.3 Transfer characteristics. 45
2.1.1.4 Frequency modulation. 45
2.1.1.5 Head gap and tape speed. 47
2.2 Video recording heads. 49
2.2.1 The Omega wrap. 53
2.3 Video-wear. 56
2.3.1 The causes of wear. 56
2.3.2 Wear of magnetic heads. 60
2.4 Methods of wear measurement. 60
2.4.1 Reduction in gap depth. 61
2.4.2 Weight loss method. 61
2.4.3 Brass shim test. 62
2.4.5 Thin film method. 64
2.4.6 Other methods. 66
2.5 Discussion. 66
References 69

CHAPTER THREE

DIFFERENTIAL PHASE-QUADRATURE SURFACE-PROFILING INTERFEROMETER.
CHAPTER FOUR

REVIEW OF PHASE SHIFTING INTERFEROMETRY.

4.1 Introduction. 111
4.2 Methods of phase shifting or modulation. 112
4.3 Methods of phase detection or measurement. 114
4.4 Phase measurement algorithms. 115
4.4.1 Sampling requirements. 115
4.4.2 General phase measurement techniques. 116
4.4.3 Three-bucket or three-step technique. 118
4.4.4 Four-bucket or four-step techniques. 120
4.4.5 Carre technique. 121
4.4.6 Synchronous detection. 122
4.5 Removal of phase ambiguities. 122
4.5.1 Extended unambiguous range interferometry. 123
4.6 Phase profile to surface contour. 124
4.7 Error analysis. 126
4.7.1 Phase shifter errors. 126
4.7.2 Detector noise and non-linearities. 127
4.8 Conclusion. 130
References. 132

CHAPTER FIVE

PHASE-SHIFTING INTERFEROMETRY FOR THE MEASUREMENT OF VIDEO-HEAD WEAR.

5.1 Introduction 136
5.2 Wyko-Zygo Phase-Shifting Interferometer. 137
5.3 Twyman-Green Interferometer arrangement. 139
5.3.1 Optical arrangement. 142
5.3.2 Matching the curvature of a video-head to a curved reference. 145
5.3.3 Measurement of curvature of video-heads. 149
5.3.4 Imaging of the video-head. 152
5.3.5 Matching the curvature of the video-head to a plane reference. 156
5.4 Experimental arrangement. 158
5.4.1 Phase-shifting method. 158
5.4.2 P.Z.T. Driving electronics. 161
5.4.3 Linearity of the P.Z.T. Power-supply system. 163
5.4.4 Calibration and linearity of P.Z.T. 164
5.5 Data transfer and processing. 172
5.5.1 Phase calculations. 172
5.5.2 Removal of $2\pi$ discontinuities. 173
5.5.4 Replacing video-head. 175
5.5.4.1 Tilt compensation. 175
5.5.4.2 Focus compensation. 176
5.6 Results. 178
5.6.1 Video recording head phase-profile results. 178
5.6.2 Representing results in three-dimensions. 180
5.6.3 $3 \times 3$ filtering. 182
5.6.4 Results of the wear on a video-head. 187
CHAPTER SIX

OTHER APPLICATIONS OF THE PHASE-SHIFTING INTERFEROMETRY SYSTEM.

6.1 Introduction. 204
6.1.1 Laser induced ripple structures. 205
6.1.2 Experimental work. 206
6.1.2a Ripple formation. 206
6.1.2b Ripple topography measurements. 207
6.1.3 Results. 207
6.1.4 Discussion. 213
6.1.5 Conclusions. 219
6.2 Profiling of SIMS sputter craters. 220
6.2.1 Experimental. 222
6.2.2 Results. 224
6.2.3 Conclusion. 232
6.3 Surface profile measurements of steel ball-bearings. 232
6.3.1 Results. 233
6.3.2 Conclusions. 236
References. 240

CHAPTER SEVEN

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK. 244

7.1 Defocus correction system. 245
7.2 Two-wavelength phase-shifting interferometry. 247
References. 249
CHAPTER ONE.

THE USE OF DOPED GLASS FLUORESCERS IN A SYSTEM FOR THE MEASUREMENT OF THE SPATIAL PROFILES OF HIGH POWERED ULTRA-VIOLET LASERS.

1.1 Introduction

Excimer lasers which emit in the ultra-violet region of the electromagnetic spectrum are being used more and more in research, development and manufacture in such varied fields as materials processing, microelectronics and surface treatment (among others).

In most of these applications it is important to know the spatial variation of the laser fluence (J cm\(^{-2}\)) or laser intensity (W cm\(^{-2}\)) at the point of interaction between the laser beam and its target. Furthermore, the fluence or intensity distribution in the near field can give information about gain homogeneity in the laser medium, and the far field (at the focus of a lens) about the divergence of the laser beam.

In this chapter a review of methods of beam profiling of ultra-violet lasers is undertaken. Photographic and beam-scanning methods are discussed together with the use of linear and 2D detector arrays.

Despite their excellent dynamic range, high resolution and real time acquisition, standard commercial imaging devices are not designed to operate with high fluence ultra-violet radiation. This problem may be overcome if the ultra-violet beam is first converted to the visible using a suitable fluorescent screen.

It is well known that fluorescence in glass, when stimulated by U.V. radiation, can arise from the presence of fluorescent ions such as Mn\(^{2+}\) and Cu\(^{2+}\), from rare-earths and actinide series transition metal ions, and from main group heavy metal ions such as tin, thallium and antimony.\(^{1,2}\)

In this chapter the properties of various doped and undoped glasses were investigated with a view to using them as ultra-violet to
visible image converters. The most important property of such converters when used for laser beam profiling is that they produce a linear response over as large a range of laser fluence as possible so that the visible image is a faithful representation of the spatial profile of the incident laser beam.

The result of such linearity measurements are described in this chapter. In addition, the temporal response of the fluorescence to a short laser pulse is determined, so that the application of different doped glass fluorescers to time resolved measurements can be evaluated. The spectral response of the laser induced fluorescence in the glasses are also investigated, since this can have an effect on the choice of imaging lenses and sensors.

A unique method for attenuating an ultra-violet laser beam's fluence, by the use of a passive absorption cell containing low pressures of $\text{NO}_2$ is described. A calibration curve of attenuation (%) of laser beam versus pressure of gas in a 50mm long cell is produced.

Finally, a system is described, based on a glass fluorescer, which writes the beam profile of a single shot KrF laser directly into computer memory and hence permits many powerful image processing routines to be used. It has the capability of producing beam profiles of two different focus positions on a single laser pulse by the use of two cameras and video framestores, controlled by a single computer.

The laser beam profiling system was developed for, and is in normal use on, the Rutherford Appleton Laboratory high power KrF laser, Sprite$^3$. Examples of its beam profile are shown in this chapter.

1.2 Review

In the past, laser beam profiles have been investigated using several techniques.
1.2.1 Photography

Photography has been widely used for laser beam profiling\(^4,5,6\). In this method, a laser pulse is recorded on photographic film having been divided by multiple reflections into a series of images with known intensity ratios. After development of the film, each exposure is scanned by a microdensitometer to yield a 2-D optical density map of the laser pulse. It has been shown\(^7\) that the variation of optical density of photographic film with exposure is highly non linear. Thus a Hurter-Driffield characteristic curve for the film is constructed by plotting relative exposure versus peak optical density for the multiple exposure image. The Hurter-Driffield curve is then employed to convert the optical density map into a map of relative energy per unit area or relative flux of the laser pulse. The final step of analysis relates the energy content of the pulse, measured by a calorimeter, to the spatial integration of the relative flux, and renormalises the relative flux map into absolute units of energy per unit area. The most significant problem associated with this method is that it does not provide real-time results and relies on very reproducible film processing and material calibration.

1.2.2 Beam-scanning

Another technique employs beam-scans\(^8,9,10\). This technique involves scanning a single detector with an aperture, such as a pin-hole, slit or knife-edge, across the beam. The detector output is normalised against the incident beam power and plotted against the horizontal position of the aperture. The use of a pin-hole as the aperture offers greater spatial resolution than that of a slit or knife-edge. However, a single scan with a pin-hole is more likely to miss a rapid change in the laser beam fluence, such as a hot spot. With the use of suitably responsive detectors, this technique can readily lend itself to the spatial profiling of laser beams in the ultra-violet through to the infra-red. Although this technique is useful and easily applied to most cw and stable repetitively pulsed systems, it is limited by the fact that a single-shot pulsed beam profile cannot be captured with a mechanical scanning apparatus.
1.2.3 Linear and 2-D detector arrays

The third technique used employs linear and 2-D imaging detectors such as photo-diode arrays, vidicons, charge injection devices (CIDs), and charge coupled devices (CCDs). Imaging detectors\textsuperscript{11,12} boast excellent dynamic range, high resolution and real time acquisition and may, in their windowless versions, be used to record directly an ultra-violet laser beam. Direct recording of an ultra-violet beam profile onto a CCD device however suffers from two major disadvantages. Firstly, the beam size needs to be matched to that of the image sensor. This can present significant problems where laser beam to be imaged is either very large or very small since the appropriate ultra-violet optics could be very costly. Secondly, silicon based imaging devices are designed to operate at low light levels and are easily damaged by irradiances typical of even heavily attenuated, pulsed ultra-violet lasers.

1.3 Ultra-violet to visible converters

The above mentioned problems are overcome if the ultra-violet beam is first converted to the visible using a suitable fluorescent screen. The image on the screen can then be recorded by a standard video camera after suitable magnification or demagnification using standard photographic lenses. The desirable properties of the fluorescent screen are that it converts the ultra-violet laser radiation into visible light with a good degree of linearity, has a suitable spectral and temporal response, and is highly damage resistant, but easily replaced if damaged. Traditional fluorescent organic materials such as sodium salicylate and rhodamine 6G deposited on a fused silica substrate were tried, but were found to be damage prone and easily bleached and, in fact, the high quantum yield of such fluorescers is not required in this application. In looking for alternatives it was noticed that different types of glass produced (to the eye) markedly different fluorescent yields but in general appeared to suffer relatively little damage or bleaching when irradiated by quite high fluences from a pulsed KrF laser. It was therefore decided to investigate the properties of various doped and undoped glasses with a view to use them as ultra-violet to visible image converters. The most important property of such converters when used for laser beam profiling is that they produce a
linear response over as large a range of laser fluence as possible so that the visible image is a faithful representation of the incident laser beam profile.

1.4 Experimental Methods

1.4.1 Fabrication of glass samples

A number of glass samples was produced and kindly supplied by Hollis\textsuperscript{13}. These were produced using standard small batch techniques. Since batch materials and preparation conditions strongly influence the final characteristics of the glass, a short description of the process is given here.

The host glasses are made from the following materials:

Local aline sand provided the silica, with an iron content of about 0.04\% wt Fe\textsubscript{2}O\textsubscript{3}/wt SiO\textsubscript{2}. ICI soda ash or potash provided the oxides of sodium or potassium. ICI minifil limestone gave the calcium oxide, with an iron oxide content also of about 0.04\% wt Fe\textsubscript{2}/wt CaCO\textsubscript{3}. Boric acid was introduced as 'dehybor' Na\textsubscript{2}B\textsubscript{4}O\textsubscript{7}. This was supplied from the '20 Mule Team' Borax Manufacturing Company, USA. Cookson Industries red lead gave the lead oxide. Dopants were added as the lowest valence oxides of the metals concerned, low-doped glasses having a 0.5\% molar dopant metal oxide to total moles of host glass oxides, high doped glasses having 5.0\% molar dopant metal oxide to total host glass oxides.

About 300 grammes of batch was filled into a mullite crucible, and heated for 7 hours in a gas-fired furnace to the correct temperature to cause the batch components to break down into the molton oxides and react. Since the tops of the crucibles were open to the gas-air combustion atmosphere of the furnace, considerable control of the redox state of the dopant in the glass was possible, thereby ensuring that the correct valence state of the dopant for efficient fluorescence (lowest valence state for all dopant ions chosen here) was possible. Then the glass was poured into disc moulds, cooled, annealed, and slowly cooled. The discs were 4 centimetres in diameter, and 1 centimetre thick. A summary of the samples tested appears in table (1.1).
1.4.2 Molar compositions of the host glasses, and their preparation conditions

The 70% molar SiO₂, 20% molar Na₂O and 10% molar CaO glass, i.e. 'crown glass', was melted at 1450°C in a reducing atmosphere for 7 hours, and annealed at 550°C for 1 hour before slow cooling overnight. All the samples with the lettering SLS (Soda Lime Silica) as part of their identification number are typical of this type.

The lead crystal glass (Pb 3/7) contained 12.6% molar K₂O, 12.6% molar PbO and 74.8% molar SiO₂. It was melted at 1500°C for 7 hours in a reducing atmosphere, annealed at 600°C for 1 hour and cooked slowly overnight.

1.4.3 Summary of glass samples tested

Table 1.1:

<table>
<thead>
<tr>
<th>Identification number</th>
<th>Dopant and concentration</th>
<th>Visual coloration¹⁴</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLS 7/6</td>
<td>Pb (low)</td>
<td>Green disc (Fe²⁺ and Fe³⁺ contamination)</td>
</tr>
<tr>
<td>SLS 7/7</td>
<td>Pb (high)</td>
<td>Blue disc</td>
</tr>
<tr>
<td>SLS 7/8</td>
<td>Sn (low)</td>
<td>Blue disc</td>
</tr>
<tr>
<td>SLS 7a/7</td>
<td>Sn (high)</td>
<td>Blue disc (Fe²⁺ contamination)</td>
</tr>
<tr>
<td>SLS 7a/8</td>
<td>Cu (high)</td>
<td>Strongly blue disc  (some Cu²⁺ present)</td>
</tr>
<tr>
<td>SLS 1c/1</td>
<td>No doping</td>
<td>Blue(Fe²⁺ contamination)</td>
</tr>
<tr>
<td>Pb 3/7</td>
<td>Pb (crystal)</td>
<td>No coloration</td>
</tr>
</tbody>
</table>
1.5 Fluorescence measurements

All measurements reported here were done using a 'Lambda Physik' EMG 103, KrF laser, emitting in the ultra-violet at 249 nm.

Results are expected to be comparable for the XeCl wavelength of 308 nm and the ArF wavelength of 193 nm since most glasses absorb strongly below 350 nm. It is expected that the laser fluence at which the glass damages will vary with absorption depth and so the maximum useable fluence at 193 nm would, in most cases, be less than that at 308 nm.

1.5.1 Time response and conversion linearity

The apparatus used to study the laser induced fluorescence is shown in figure (1.2). A fast photodiode D₂ (Type ITL TF1850 UV) placed behind the glass sample recorded the fluorescence directly. Part of the laser beam was split off to provide a reference signal on another photodiode D₁. The laser beam energy was varied using a variable pressure cell filled with NO₂ gas, see section (1.6.4).

To obtain absolute measurements of the laser fluence, the reference diode, D₁, was calibrated against a pyroelectric joulemeter. A neutral density filter, ND, was used to approximately equalise the signals from the two photodiodes. The NO₂ cell attenuated the 249 nm laser pulse fluence over four decades, from 10μJ cm⁻² to 100mJ cm⁻². The linearity measurements were done with the photo-diodes operating in the integrating mode so that the CRO signal on both diodes was proportional to the total optical energy detected. Thus a measure of the linearity of detected fluorescence in terms of laser fluence on the sample was obtained. Since the response of a CCD device to a short optical pulse depends on detected total fluence, the fluence linearity is the important check rather than the linearity in terms of instantaneous intensity which could well be very different. To measure the time response of the fluorescence, the photo-diode D₂ was terminated in 50Ω at the CRO. This gave a sub-nanosecond response time and the variation of fluorescence power with time could be measured accurately.
Figure 1.2 The apparatus used to measure the linearity and time response of the glass fluorescer.
1.5.2 Fluorescence spectra

A series of measurements using a continuous ultra-violet source with a wavelength close to that of the excimer laser was carried out to determine the fluorescence spectra of the glass samples. The source of exiting radiation was the 253.7 nm line from a 'Pen-ray' Hg discharge lamp operating with a d.c. supply.

A computer controlled spectrometer, the Rofin Spectralyser 6000, was used to measure the broad band spectra over a range 300-600 nm. The spectral sensitivity of the spectrometer system was calibrated using a standard tungsten lamp operated at a known black-body temperature (2200K). At this temperature, using values for the emissivity of tungsten as a function of wavelength\(^1\) multiplied by values of black body spectral intensities for 2200K\(^2\), it was possible to plot a theoretical curve of the black body radiation emitted from the tungsten at 2200K. This was compared to the curve obtained using the spectrometer and the correction factors were evaluated, figure (1.3). Thus the plotted spectra were corrected to give a true relative spectrum for the different samples.

The undoped SLS glass produced a signal which was too weak to be recorded accurately, as compared to the doped glasses which produced much stronger signals, results of which appear in Figure (1.4).

1.5.3 Results and discussions.

Dopants of lead, tin and copper were tried and figure (1.4) shows that a significant increase in the fluorescence signal can be achieved over undoped glass. The time response of the fluorescence is, in general quite complex with no single decay time dominating. It is clear that the glasses are unsuitable as fast fluorescers which will follow the time response of a short laser pulse. However, in imaging, this is not important since it is the integrated fluorescence which is recorded (that is, the area under the curves in Figure 1.4) and in this respect lead dopants clearly give the highest fluorescence yields. This result is confirmed by the fluorescence spectra shown in Figure (1.5). The advantage of lead dopants is not so strong if one considers the spectral sensitivity of the CCD image sensor, which in the non-intensified versions is much more sensitive in the red
Figure 1.3 Comparison of the a) theoretical curve of black-body radiation from a tungsten filament at 2200K, to b) curve measured using the Rofin spectrometer.
Figure 1.4 Plot showing the relative fluorescence power as a function of time for various doped glasses irradiated with a KrF laser pulse at about 1mJ cm⁻².
Figure 1.5 Plot showing the relative fluorescent spectra of glasses doped with various materials when excited at 254 nm by a Hg lamp. Curves have been corrected for the spectral response of the spectrometer and photomultiplier.
than in the blue\textsuperscript{17}. In this case tin dopants look more promising. For maximum sensitivity, the combination of a lead-glass fluoresecer and an image-intensified CCD device would be best.

The linearity of the various doped glasses is shown in Figures 1.6 and 1.7. The relative fluorescence yields from the different glasses are not to scale in these figures. The influence of doping concentration on linearity is apparent. Curiously, the concentration affects the linearity differently with different dopants. For example, high doped tin and low doped lead both show significant non-linearity. However, it should be noted that between 100 \( \mu \text{J} \) and 10 mJ cm\(^{-2} \), all of the glasses are adequately linear to be used with a 6-bit resolution frame store. For ultra-linearity over a very wide range, soda-lime silica glass highly doped with lead appears to be a very good image converter.

1.6 Application: Laser beam spatial profiling system

Although the undoped SLS glass had a significantly lower fluorescence efficiency than the doped glasses, it proved to be sufficiently high for use as a laser beam profile image converter since, in this application, efficiency is not of prime concern. Ultra-violet radiation below 300nm is strongly absorbed in this glass and consequently the fluorescence is generated within a few \( \mu \text{m} \) of the entrance face, by the \( \text{Fe}^{3+} \) impurities\textsuperscript{2}. Since it is readily available as very thin plates (microscope cover slips) it is possible to record the profiles of extremely small laser spots (\( \approx 10\mu\text{m} \)) by viewing the fluorescence transmitted through the plate with a microscope objective fitted to the CCD TV camera and focused on the front face of the cover slip.

1.6.1 Video format

A TV picture is built up by an electron beam scanning the face of a cathode ray tube (CRT) in a pattern of horizontal lines. The beam makes one left to right trace to produce each line and its passage takes just 52\( \mu \text{s} \). After completing one line the beam is cut off (line blanking) and flies back from right to left, taking about 12\( \mu \text{s} \) to do so, ready to begin the next line. While this horizontal scanning process is taking place, the
Figure 1.6 Linearity of fluorescence energy with incident laser fluence for copper and tin doped glasses.
Figure 1.7 Linearity of fluorescence energy with incident laser fluence for undoped and lead doped glasses.
beam is simultaneously being slowly drawn downwards (by the action of the field deflection coil) until it reaches the bottom of the screen after 20ms. The beam then flies back to the top left corner, to begin another complete scanning pattern. This second fly-back period is termed field or frame blanking, and is 800μs long.

In Europe, the TV screen, and hence the picture, consists of 625 lines of information. To achieve this number of lines in 20ms (50 Hz), without needing a large band-width, a system known as interlaced scanning is used. With a picture frequency of only 25 Hz, alternate lines are scanned in each vertical sweep, the remainder being interlaced between them in the following vertical sweep. The picture is now made up of two interlaced vertical scans - called fields - each covering half the total number of lines. In the 625 line picture each field will be of 312½ lines.

1.6.2 Video camera

The video camera used to image the fluorescence was an EEV P4310 with a charge-coupled device (CCD) sensor, with a dynamic range of greater than 200. This small solid-state sensor 'chip' contains a matrix of many thousands of tiny silicon photodiodes or 'pixels', each of which effectively forms a capacitor whose charge is proportional to the brightness of light falling on it. The charge on each pixel of the array is read sequentially and forms the video signal by the addition of timing or 'sync' pulses to synchronise the TV receiver.

The EEV P4310 camera is a frame-transfer type device (FTD). This implies that the chip is active for recording for 18 ms, and is inactivated by the application of bias pulses to the pixels. During the last 2 ms of a field the charge on each pixel is transferred on a point-by-point basis to an adjacent part of the chip, known as the storage area which is identical to the imaging part except that it is shielded from light. In a FTD, the read out of the pixel charges and the addition of sync pulses is performed from the storage area and takes place during the frame after the imaging. Hence the video frame coming from an FTD is one field or 20ms behind the actual event.

A feature of most standard commercially available video cameras
is the presence of automatic gain control (ACC) i.e. built in electronic circuit to change the amplification of the output signal depending upon the amount of incident light. This feature would obviously invalidate any absolute measurements made. Fortunately the EEV P4310 has no ACC. Another potential camera problem is blooming. Blooming occurs when any pixel whose charge exceeds its charge-holding capacity, i.e. when it becomes saturated due to too much incident light, spills over and corrupts the information in the surrounding pixels, and distorts the image. However, if the chip or any part of it is not saturated, then blooming will not occur. In some more up-to-date video cameras, anti-blooming circuitry is built into the chips. Hence, when a pixel becomes saturated, the excess charge is 'drawn out' and subsequently does not affect the adjacent pixels.

As explained previously, a TV frame is made up of two interlaced fields each 20ms long. To effect this interlacing, each pixel on the CCD is actually comprised of two horizontally split halves, so that the data for successive fields comes from different halves of the pixel and is thus spatially displaced by 10 μm. This results in what is termed 'half pixel shift', which occurs when the data for successive fields at the same point in the video image has come from a different point in space and is therefore slightly different. This is not a problem if the images are to be used for simple viewing or beam profiling. If however, two images need to be subtracted, such as in the case for background subtraction (section 1.7.3), then it is important that this shift be eliminated by ensuring that the related images are always captured with time delays of multiples of 40 ms, which ensures the same field is captured each time.

1.6.3 Video framestore and interface card

The beam profiling system is based on an Image III video framestore produced by Eltime Ltd, England. The framestore is based upon the VISOO single board field store developed by British Telecom Research Division and is a computer-controlled digital picture store with variable picture resolution and 6 bit A/D and D/A converters giving 64 intensity levels or 'grey levels'.

The framestore is connected via a ribbon cable to an interface card, which is inserted into the standard I/O expansion slots at the back
of an IBM PC computer. It performs the functions of re-organising pin allocations, buffering selected signals and generating the frame store control signals. A DIL switch on the interface card enables the frame store to occupy any block of sixteen address locations in the I/O range HEX 000 to HEX 300, by setting the desired base address. The boards are supplied with the address HEX 300 preselected. For the two-beam profiling system, the second framestore address is needed to be different, so HEX 200 was chosen.

For beam profiling purposes the framestore is operated in the two-store mode, where two frames of 256 (vertical) by 512 (horizontal) pixels are capable of being stored simultaneously in memory. In the work described in chapter 5, the framestore was set to operate in the four-store mode, with four frames of 256 by 256 pixels being stored. The framestore required some hardware modifications so that both the controlling computer, as well as an external trigger signal from the laser, could initiate a frame capture or 'snatch'. The snatch is synchronised to and starts on the first field sync pulse after the request, and the entire incoming video signal for that field is captured and digitised. The digitisation is performed so that level 0 corresponds to 0.3 volts on the video signal and level 63 to 0.7 volts.

The incoming data is stored in the framestore memory in a grid format, with the top-left-hand corner of the memory having a cartesian type coordinate of (0,0), and the bottom-right-hand side corner with a coordinate of (512,256). By the use of suitably written computer programmes, it is possible to write to and read from any location in the framestore memory, thus permitting many powerful image processing routines to be used, in addition to performing such operations as reading and writing images to and from computer disc storage.

1.6.4 Laser beam attenuation

In measuring the conversion linearity of the glass fluorescence, it was necessary to measure the fluorescence signal at several laser fluence (J cm⁻²) levels. It was also important to have a means of adjusting the fluence of the 'SPRITE' laser when profiling, to obtain the correct light level exposure by the video camera, and to prevent saturation of the
CCD detector elements. Various methods have been tried by several researchers for attenuating the beam of KrF lasers. A system proposed by Bennett and Byer\textsuperscript{18} uses two pairs of counter-rotating fused silica wedges. This makes use of the principle of variation in transmittance of a fused silica beamsplitter with angle of incidence. This system has a dynamic range of 40dB and a high laser-induced damage threshold. It does, however, require large incident angles (>60') to achieve significant attenuations. This limits its use with large output aperture lasers, such as SPRITE.

In a review by Wiseall\textsuperscript{19}, various other methods of attenuation are discussed. Among these is the use of thin plastic films, commercially available as 'cling-film'. The attenuation can easily be varied by increasing the number of layers. The attenuation increases as \( \exp(N\alpha) \) where \( N \) is the layer number and \( \alpha \) is the absorption coefficient for a single layer of film. However, it is shown that the laser damage threshold of the film is 0.5 Jcm\(^{-2}\). This level would be easily exceeded in profiling the laser SPRITE. An alternative to thin film absorbers are volumetrically absorbing solutions, which were considered because of their improved resistance to laser-induced damage at 248nm. Wiseall proposed a mixture of hydrated nickel sulphate, (NiSO\(_4\).6H\(_2\)O) and cobalt sulphate (CoSO\(_4\).7H\(_2\)O) placed in a perspex reservoir attached to an aluminium cylinder via a hollow brass collar. The aqueous solution was simply drawn into or expelled out of the cell region by a hand-driven piston. The fused silica windows on the front of the piston and aluminium housing had a clear aperture of 44mm. The liquid cell length could be varied in the range 2.5 to 60mm.

The KrF laser beam attenuator in use at present at Loughborough University is based on the above system, but uses an organic dye solution (Rhodamine 6G) as the absorber.

An alternative to a liquid volume attenuator is a gas absorber. This method permits the use of a fixed cell length with the attenuation increased by raising the pressure of the gas inside the cell, figure (1.8). The gas used was nitrogen dioxide (NO\(_2\)), supplied by BDH at 99.5% purity. This method was based on a system used by Armandillo etal (1982)\textsuperscript{20} for measuring the gain of the discharge of rare gas halide lasers. This is found by including in the laser cavity a cell containing a passive absorber gas, and increasing the pressure until the laser action is extinguished,
Figure 1.8  NO$_2$ gas filled KrF laser beam attenuator.
thus relating the absorption of the gas to the gain of the laser.

Figure (1.9) shows the calibration curve of percentage attenuation of the KrF laser beam versus the pressure in torr of the absorbing gas. This curve was produced using the arrangement similar to that explained in section 1.5.1. It must be noted that the numbers shown on the curve are only correct for the specific cell length used, i.e., 50 mm. It is seen from figure (1.9) that at about the 50-torr mark, there appears to be a nonlinear pressure dependence of the absorption. This arises because at very low pressures the gas exists in the form NO₂, but at higher pressures the dominant molecular species is N₂O₄. At 1 bar and 27 °C, about 80% of the gas is in the form of N₂O₄ (Plekhotkin, 1970)²¹. The species NO₂ and N₂O₄ have absorption coefficients which differ considerably in magnitude and spectral dependence over the near ultra-violet region (Nakayama et al 1959)²².

1.6.5 Timing requirements

It was shown in section 1.5.3 that the decay time of the KrF laser-induced fluorescence of the glass is of the order of 1 µs; this is obviously much shorter than the 20 ms TV frame period. It is therefore necessary for the video framestore to be synchronously triggered with the laser pulse in order to capture the TV frame which contains all the information. It is also important for this event to occur in the middle of a TV frame to avoid the laser being fired during a video fly-back period, which would result in the loss of the information.

To control the timing of these events an electronic synchronisation box was built. Figure (1.10) shows the basic circuit diagram of this box. The first part of the circuit, adapted from previous work²³, comprises three BC108 transistors and a 355 operational amplifier which amplify the input signal from the video camera. This signal is then input into the 311 voltage comparator which separates the data and line sync pulses from the field sync pulses. Thus the output of the 311 chip is one pulse every 20 ms corresponding in time to the presence of a field sync pulse. The 7408 AND gate simply serves to shape the pulse. On the arrival of the leading edge of an external pulse, which indicates the wish to fire the laser, the 74121 monostable multivibrator outputs a 20 ms-long positive
Figure 1.9 Calibration curve showing percentage attenuation of the KrF laser beam versus pressure of absorbing gas (NO₂) in torr for a cell 50 mm in length, at 25°C.
Figure 1.10 Circuit diagram of electronic event synchronisation/trigger box.
going pulse, the length of which is determined by the values of the resistor and capacitor external to the chip by the equation \( t = 0.6RC \). The arrival of the leading edge of the 20ms pulse activates the 7474 dual D type flip-flop, so that on the arrival of the next, and only the next, field sync pulse, does it give an output. This pulse is then passed to two 74121 monostable multivibrators. The first of these is triggered by the leading edge of the pulse, and outputs a negative pulse 6ms long. The positively going trailing edge of this pulse triggers the second 74121 which outputs a pulse delayed by about 6ms to the field sync pulse, (i.e. roughly a third of the way into the TV field). A dual input monostable multivibrator outputs two pulses of differing lengths, to suit the framestore and laser triggering requirements. The signal to the laser is amplified by the 531 op-amp up to 20V before it is used, together with the framestore signal, to drive the two opto-isolators. These diodes are used to isolate the circuit from any reflected signals on the outputs, which had been found to cause the framestore to trigger a second time and over-write the original data.

The pulse to the framestore is 150\( \mu \)s long and negatively going, as recommended by 'Eltime', the framestore manufacturer, for correct triggering\(^{24}\). The pulse needed to trigger the firing circuit of the laser 'Sprite', was a positive going 20V pulse with a duration of 10\( \mu \)s\(^{25}\). These pulses are both provided by the outputs of the timing circuit in the correct manner, with the framestore pulse in synchronisation with the start of a new video field, in order to capture the correct frame, and the laser firing at a point a third of the way into the TV field, to ensure consistent recording of the image. Figure (1.11) shows the pulse timing sequences.

1.7 Control and data processing

1.7.1 Computer control

The computer used to control the system and collect the data was an Amstrad PC 1520. This is an IBM PC compatible, with 520k byte of RAM and a built in 20M byte Winchester hard-disc plus a single floppy disc drive. At the back of the computer are three Input/Output ports, one of which was
Figure 1.11 Pulse timing sequence. The frame-store is triggered on the falling edge of a pulse synchronised to the start of a field sync pulse, while the laser fires on the rising edge of a 6ms delayed pulse. This prevents the laser being fired during a fly-back period.
taken up by the hard-disc controller, the other two of which were occupied by the framestore interface cards. The hard-disc is able to hold up to 150 TV images of 512 by 256 resolution. The floppy disc drive was used to make back-up copies of the more interesting images.

All the computer control and data processing routines were written in 'Turbo Pascal'. This is a fast version of the standard high-level computer language Pascal, and was developed and marketed by Borland International Inc, California.

Initially, 80286 assembly language was to be used in the programming, for its speed. However, when the programs written in Turbo Pascal are compiled and executed, it is found that they perform their tasks in times not much slower than routines written in assembly language, with the added advantage of the structure and simplicity of use of a high-level programming language.

For convenience to the operator, the computer used was dedicated to the beam profiling system. On switching on, the screen displayed the amount of free memory available on the hard disc. This is to prompt the clearing of memory, if full, in anticipation of the arrival of new data. On pressing any key, the system software is automatically loaded from the hard disc and run. This then loads the program MAIN.COM, which displays an option to select framestores (1) or (2). On selection of either these two options, the particular framestore is initialised, and the program BTP300.CHN or BTP200.CHN is loaded and run. On exiting from either of these two programs, the option program MAIN.COM is loaded and run.

The programs were all written to be menu driven, and user friendly.

1.7.3 Data handling and processing routines

A suite of programs was written to perform the following tasks:
(a) Acquisition, storage and display of data
(b) Image processing
(c) Intensity profiling in x and y directions
(d) Three-dimensional contour map
(e) Numerical integration
(f) Statistical analysis
(g) Calibrate dimensions

By using (g) it is possible to calibrate an image in absolute units. This is done by replacing the glass slide with a grid of known dimensions, such as a piece of graph paper or microscope objective graticule, so as to be in the same plane as the front surface of the fluorescer. On capturing this image on the framestore, cursors are moved to suitable positions on the grid. By entering into the computer the distance between the cursors in the x and y directions, the computer counts the number of pixels, and so calculates the number of pixels per unit length for that particular magnification in the vertical and horizontal directions. These are found to be different, due to the way in which the TV and framestore combine to present an image, with an aspect ratio of 2 to 3. These values are then automatically stored on the hard disc in one of two files, depending on which framestore is activated at the time. These will overwrite any other values that might have been stored previously, and are used in calculating the area of the beam and laser fluences. It is necessary to recalibrate the absolute screen dimensions when changing the magnification of the imaging system. Figure (1.12) shows the image of a grid with cursors together with the calculated screen dimensions.

Option (e) provides a histogram of laser fluence (grey-level) which shows the radiation distribution and hence the most likely fluence. The histogram also provides a good indication of the background/image cut-off point. As the computer integrates over the whole screen to calculate fluences, it is necessary to set a background level below which the computer discriminates against, and thus integrates over the area occupied by the beam only and not the whole screen. On selection of the appropriate cut-off point, the computer replots the histogram with the background numbers removed, normalised to fill the screen. This is shown in figure (1.13).
distance in X direction is 111 pixels ....

.... which gives 93 pixels/cm.

distance in Y direction is 162 pixels ....

.... which gives 135 pixels/cm.

Figure 1.12 Image of grid of known dimensions shown with cursors positioned at suitable points on the grid. Also shown is an example of a computer screen print-out giving the absolute dimensions.
Figure 1.13 Histogram of laser fluence (grey-level)

a) with background,
b) with background removed.
Another source of possible error is due to the camera 'dark level' noise. This may be eliminated by using option (b) which includes a routine to subtract the contents of two frames, one of which is taken momentarily before the laser pulse, and consequently only contains dark level noise. The other frame contains the image plus dark level noise. When the first frame is subtracted from the second, the frame which contains the image becomes free of dark level noise. The time interval between the two frames was set to be 40ms, to avoid the 'half pixel shift' problem, see section 1.6.2.

It had been noted that the increase in temperature of the camera allowed the 'dark level signal' to rise appreciably, hence the camera was turned off between laser firings, and a fan was used to direct cool air at the camera when in use.

Option (b) also contained routines for enhancement of low contrast images, by 'stretching' the range of intensity levels to fill the 64 levels available, and another routine to display the difference (in intensity levels) between two frames.

On entering the values of the estimated 'cut-off' background level and the energy of the laser pulse $T_{tot}$ (monitored by a joulemeter) into the computer together with the number of pixels per centimetre in the horizontal ($N_x$) and vertical ($N_y$) directions, option (f) enables the computer to give such parameters as average fluence and peak fluence ($J \text{ cm}^{-2}$) for that image.

In order to convert grey levels (0-63) into fluence ($J \text{ cm}^{-2}$), it is necessary to determine what fluence a certain grey level corresponds to. For a grey level $g_i$ there is a fluence ($J \text{ cm}^{-2}$) $F_i$ such that:

$$F_i = f \cdot g_i$$

Let the total digitised object be divided up into pixels so that there are $N_x$ pixels/cm in the $X$ direction and $N_y$ pixels/cm in the $Y$ direction. [n.b. these are dimensions on the object: the magnification of the lens and the size of image sensor are not needed]. Thus the area (all referred to object) of a pixel is
and thus the energy per pixel is

\[
\frac{1}{N_xN_y}
\]

Now let there be \( n_1 \) pixels having the same energy \( \frac{F_1}{N_xN_y} \), the total energy at fluence \( F_1 \) is thus

\[
\frac{n_1 F_1}{N_xN_y}
\]

and summing over all the other fluence levels we get the total beam energy

\[
E_{\text{tot}} = \frac{i \sum n_i F_i}{N_xN_y}
\]

Hence

\[
E_{\text{tot}} = f \frac{\sum n_i g_i}{N_xN_y}
\]

or the fluence/greylevel \((f)\) is:

\[
f = \frac{E_{\text{tot}} N_xN_y}{\sum n_i g_i}
\]

Now the average fluence \( F_{\text{av}} \) is simply

\[
F_{\text{av}} = f \frac{\sum n_i g_i}{\sum n_i}
\]

Thus

\[
F_{\text{av}} = \frac{E_{\text{tot}} N_xN_y}{\sum n_i}
\]

The ratio of peak to average fluence is a simple measure of overall uniformity in the beam.

The profiling option allows the user to select the location of a line profile by moving a cursor across the screen in the \( x \) and \( y \) directions and produces graphs of recorded intensity versus position which are then plotted on the computer screen ready for printing.
Option (d) gives a three dimensional contour map of the image, by displacing the position of the pixel in the y direction by an amount proportional to its intensity level thus simulating a z-direction.

Finally, option (a) contains routines for writing TV frames to the floppy and hard discs for future referencing and processing.

Figure (1.14) shows the output from a routine which assembles all the most frequently required data concerning a laser shot on a single A4 sheet. A thermal copy of the actual TV image is included on the shot data sheet. The two white lines crossing the image represent the position of the plotted intensity profiles in the horizontal and vertical directions shown. Also included are a 3D profile of the image and a histogram of laser fluence (grey level).

In this particular example the laser was the Rutherford Appleton Laboratory high power KrF laser, SPRITE and the interaction experiment was conducted with the beam focused to approximately 8 mm in diameter. The experimental lay-out of the system is shown in figure (1.14).

1.7.4 Printer and video copier

A dot matrix EPSON FX 100+ line printer is connected to the centronics parallel port at the back of the computer. This gives the ability to print the contents of the computer screen which contains a record of a laser shot as shown in figure (1.15).

In order to obtain a hardcopy of the image captured on the frame-store, a Mitsubishi P60B video thermal copier is used. On the press of a button, the image in the frame-store is produced on paper with an intensity resolution of 16 levels. This produces an unavoidable loss of detail, when compared to the original 64 intensity level image in the frame-store, but was found to be adequate for simple viewing purposes.

1.8 Conclusions

The effect of the type and concentration of various dopents, in
Figure 1.14 Optical set-up to measure a pulsed ultra-violet laser beam profile.
Laser Shot Number...
...5127a
Date...27/11/87
Energy of Laser Shot...
16 JOULES
Peak Fluence is
122 J/cm2
Average Fluence is
38 J/cm2
Area of beam is
46 mm2
FULL SCREEN DIMENSIONS
H - 20 mm  V - 13 mm

**Figure 1.15** Record of laser shot showing horizontal and vertical profiles, 3-D profile, fluence histogram together with derived values for the laser spot size, average and peak fluence.
the host glasses, on ultra-violet radiation induced fluorescence has been shown. This has demonstrated the viability of using these glasses as effective ultra-violet to visible image converters for the purpose of pulsed ultra-violet laser beam profiling. Also described in this chapter is the hardware of a system developed specifically for the beam profiling of the Rutherford Appleton Laboratory high power KrF laser, Sprite, and the computer routines used to process and display these profiles.

Finally, a unique method for attenuating the fluence of an ultra-violet laser, by the use of a passive absorption cell containing low pressures of NO₂ gas, is described. It has been shown to attenuate the fluence of a KrF laser over four decades, from 10μJ cm⁻² to 100mJ cm⁻², and is again in everyday use at the laser division of the Rutherford Appleton Laboratory.
References


36


CHAPTER TWO.

VIDEO RECORDING AND VIDEO-HEAD WEAR REVIEW

2 Introduction.

In this chapter the history of video tape recording is traced from its origins by discussing the early development of audio tape recording in the late 19th century through to the use of such developments as rotary head technology and frequency modulation as used in present day video recorders. The principles of magnetic tape recording are discussed and a description is given of a typical recording head.

The effect of head wear on the signal to noise ratio of a typical recording system is highlighted. This leads to a comprehensive review of the major fundamental cause of wear, due to the rubbing of two rough surfaces, and the discussion of the most likely causes involved in magnetic recording head wear.

Finally in this chapter, a review of the methods and techniques that have been developed over the years to establish relative abrasivity of various tapes and the life expectancy of various head materials is given.

2.1 Video tape recording.

The emergence of video tape recording is a direct consequence of the earlier development of audio tape recording. The basic idea of recording a signal in terms of an alternating magnetic field along some magnetic material was first put forward in 1899 by Valdemar Poulsen of Copenhagen. By this time the relationship between electricity and magnetism was well understood, and in 1900 a US patent on Poulsen's apparatus, the Telegraphone, was awarded. Initially the medium was magnetic wire; this gave way in the 1930s to 6mm-wide steel tape travelling at 1.5 metres per second. In 1932 the BBC broadcasted a programme of the Economic Conference in Ottawa, for which seven miles of steel tape was used, edited by means of a hacksaw and soldering iron!
A great step forward in magnetic recording was achieved when it became possible to coat a thin and flexible insulating base with a fine layer of magnetic particles, as shown by Pfleumer of Germany in 1928. The first commercial sound tape recorder, using a cellulose acetate tape coated with carbonyl iron powder, developed by Wilhelm Caus, was demonstrated at the Berlin Radio Exhibition in 1935. This and other developments caused the performance of these sound recording machines to steadily improve during the '30s and '40s to the point where much radio broadcast material was from pre-recorded tape, and indistinguishable from live programmes. These professional machines, which were a key component in the commercial development of hi-fi, used open reels of tape and were somewhat awkward to load. Magnetic tape recording for the home truly came into being with the invention, by the Dutch electronics group Phillips, of the early audio cassette which eliminated the loading problems encountered with open-reel tape.

Once audio magnetic recording had become established in the radio industry, attention was turned towards the possibility of recording television images on tape. The problems were formidable, mainly because of the relatively large bandwidth of a television signal. In order to explain the above statement it is necessary to describe the principles of magnetic tape recording and the main component involved, the recording head.

2.1.1 Principles of magnetic tape recording.

All magnetic materials, contain magnetic dipoles. In the natural state these dipoles are randomly aligned within the material so that their fields cancel one another out, and no external magnetic force is present. To magnetise the material, an applied external magnetic force is required to align the dipoles, which then point in a specific direction. When the external field is removed, most of the magnetic dipoles in the material remain in alignment and the material now exhibits magnetic properties of its own.

This ability to magnetise or 'mark' certain materials, namely magnetic tape, in a controlled fashion, by applying an external field, forms the basis of magnetic recording.
2.1.1.1 Magnetic tape recording head.

The changing magnetic patterns corresponding to the input signal are recorded on the tape by the recording head. The recording head consists of a near circular ring of a ferrous magnetic material with a very narrow gap which interrupts the otherwise complete ring, as shown in figure (2.1). The gap is filled with a nonmagnetic material, such as glass. A coil of fine wire is wrapped around the ring. An alternating current is applied to the coil which induces a changing magnetic field in the metal ring. A changing magnetic field causes bridges to form around the interruption of the magnetic circuit which occurs at the gap. The magnetic tape passes by the gap, and the magnetic energy bridges this gap by travelling through the magnetic coating on the tape, figure (2.2). A residual, or remanent, magnetism remains which is the recorded signal. The magnetisation induced onto the tape is proportional to the electrical current in the coil of the recording head.

On replay, the magnetic field recorded on the tape passes by the gap. This creates a changing magnetic field which induces a voltage at the terminals of the coil. The same head can thus be used for both writing and reading the recorded information. It is necessary for the head material to be magnetically soft to avoid magnetising the tape after signal currents have ceased.

2.1.1.2. Hysteresis loop.

When the magnetising force applied to a magnetic material is increased, it causes an increase in magnetism until the saturation point is reached. If the magnetising force is reduced the material retains a certain level of magnetism, depending on the remanence of the material. When the force is reduced to zero the remanent magnetism represents the peak magnetism that the material can retain. This is known as the retentivity of the material. If the force is reversed in a negative direction another saturation point is reached. If the force is again reduced to zero, the material retains magnetism of reversed polarity. If the cycle of magnetising forces continues, the magnetism will follow the curves shown in figure (2.3). This is known as a hysteresis loop. The area of this loop is an important characteristic of magnetic materials. The coercive force is
Figure 2.1 Side view of a magnetic head used for tape recording and playback. A magnetic circuit is created in the near circular metal assembly by the signal at the coil of wire. This magnetic circuit is broken by a thin gap of non-magnetic material. The magnetic energy bridges this gap by travelling through the magnetic coating on the tape. A residual, or remanent, magnetism remains which is the recorded signal.
Figure 2.2 The flux distribution across the pole tips of a conventional head.
Figure 2.3 Hysteresis loop.

Magnetization.

$H$ →

Magnetic field strength.

$I$
the force required to reduce the magnetism of the saturated condition of the material to zero. The higher this force, the higher the value of signal that can be recorded.

In addition to being the means for recording information, hysteresis is also used to erase the recorded signal completely. This is done by placing the magnetic material in a gradually diminishing a.c. field, which makes successively smaller hysteresis loops until finally no magnetisation remains.

2.1.1.3 Transfer characteristics.

The relationship between the magnetizing field in the gap $H$ and the residual, or remanent, magnetic induction $B_r$ in the tape is highly nonlinear, as shown in figure (2.4). This means that the magnetism in the tape would be a highly distorted replica of the signal to be recorded. There is, however, a portion of the curve which is very nearly linear. The solution thus is to operate in this linear region by adding a high-frequency tone (sine wave) to the input signal. This high-frequency wave is called the bias current.

2.1.1.4 Frequency modulation.

A problem encountered with recording video signals is the large range of frequencies encompassed by the video signal, from 30 Hz to about 5 MHz, a range of nearly 18 doublings in frequency, or octaves. This is very much greater than audio tape recorders, which use a range of only 10 octaves. Since each octave results in a doubling of the amplitude of the voltage of the playback signal, the dynamic range of the largest output voltage to the smallest output voltage would be 262,144 to 1, or 108 dB. Magnetic tape can be satisfactorily used over 10 octaves, which corresponds to about 60 dB, above which it is difficult to separate the low frequencies from the noise of the system.

The solution is to modulate the video signal onto a carrier frequency. If a 6 MHz carrier is modulated with 5 MHz, the upper and low sidebands will range from 1 MHz to 11 MHz. The octave range is now reduced to just over five and easily handled.
Figure 2.4 The transfer characteristics between the magnetising force $H$ applied at the gap and the remanent magnetization $B_r$ in the tape is highly nonlinear. The solution is to add a high-frequency bias to the input signal to cause operation in the linear range of the characteristic.
Although amplitude modulation, AM, could have been used, frequency modulation FM, was adopted because it confers other advantages, particularly in the area of noise reduction. An FM signal can be recorded at constant level regardless of the modulating signal amplitude, so that head losses and the effect of imperfect head-to-tape contact cause fewer problems.5

2.1.1.5 Head gap and tape speed.

The linear relationship between head field strength and stored flux in the tape, as described in section 2.1.1.3, holds true provided the wavelength of the signal to be recorded is long compared to the width of the gap. However, when the wavelength of the signal on the tape becomes comparable with the head gap width, the flux imparted to the tape diminishes, reaching zero when the signal wavelength is equal to the width of the head gap6. This is shown in figure (2.5), where it can be seen that during the passage of a single point on the tape across the gap, the applied flux has passed through one complete cycle, resulting in a cancellation of the stored flux in the tape. This point is known as the extinction frequency, and sets an upper limit to the useable frequency spectrum. As a practical matter, the recorded wavelength should be no shorter than half the width of the gap.

Video signals contain very high frequencies, which are impossible to record at the very slow speeds at which tape passes by the record head as used with audio recording.

The speed of the head with respect to the tape is called the writing speed4. The frequency \( f \) of a recorded waveform is related to the writing speed \( v \) and the wavelength \( \lambda \) by the relationship:

\[
f = \frac{v}{\lambda}.
\]

Assume that a maximum frequency of 4 MHz is to be recorded. A practical gap width is 0.6 \( \mu \text{m} \). The longest wavelength that could be recorded is thus 0.3 \( \mu \text{m} \). The required speed of the tape passing the head can then be calculated from the preceding equation as the product of the frequency and the wavelength, or 4 MHz \( \times \) 0.3 \( \mu \text{m} = 1.2 \text{ metres per second.} \)
Figure 2.5 Storing flux on the tape. At a) the flux appearing across the head gap is penetrating the oxide surface to leave magnetic patterns stored on the tape. b) shows the effect when one complete cycle of the recorded waveform occupies the head gap - no signal transfer will take place.
would be impossible for a small cassette, or even a practical size of open-reel tape, to contain enough tape to record a reasonable amount of time at such a fast tape speed. How this is achieved will be explained in due course. Other HF losses also occur during recording. The head is by definition inductive, so losses will increase with frequency. Eddy currents in the head will add to these losses, as will any shortcomings in tape-to-head contact. High frequency signals give rise to the generation of short 'magnets' in the tape itself, and it is the nature of these to tend to demagnetise themselves. For all these reasons, the flux imparted to the tape tends to fall off at higher frequencies, as in the solid line of the graph shown in figure (2.6). To counteract this, recording equalisation is applied by boosting the HF part of the signal spectrum in the recording amplifier, as per the dotted line. This is termed recording equalisation, and its aim is to make the amplitude/frequency characteristics of the signal stored on the tape as flat as possible.

2.2 Video recording heads.

In the early days of video development attempts were made to produce longitudinal recording machines, such as the BBC's VERA (Video Electronic Recording Apparatus) of 1956. Such machines were wasteful of tape, and frightening to anyone who happened to be in the room in which they were operating.

It soon became clear that another approach was needed. The answer was Rotary-Head Technology.

In a rotary-head machine there are two simultaneous motions: the tape is transported in the normal manner and the heads, mounted on a drum, rotate at high speed. Figures (2.7a&b) show the two main types of rotary head arrangements. In the original Ampex "quadruplex" recorders, the axis of rotation of the drum was parallel to the direction of the tape motion, and the heads therefore, described tracks almost transversely across the tape. In later professional and all consumer machines, the drum axis is inclined so that the heads describe tracks that are more nearly parallel to the tape motion; this type of scan is called helical. In the transverse quadruplex recorders each track was, of necessity, short and only a portion
Figure 2.6 Losses in the recording process. The dotted line shows a compensating 'recording equalisation' curve.
Figure 2.7a  Transverse scan. A) shows a quadruplex head drum. B-a) vacuum unit, B-b) tape guide. C) Video track pattern. If the tape were not moving the recorded tracks would be perpendicular to the edge of the tape. The tape movement gives the tracks a slight slant.

51
Figure 2.7b Helical scan: two headed wrap.
A) rotating drum. B) tape wrap greater than 180°. C) tape climb less than tape width. D) first head tracks indicated by continuous lines, and second head tracks indicated by dotted lines.
A video field could be recorded. Slight differences from one tape scan to the next caused visible defects in the picture. The later helical recorders overcame this problem because longer tracks could be scanned and one complete television field could be scanned on each pass.

In helical-scan or slant-track recording, as it is sometimes known, the drum is rotated at 25 revolutions per second. Two heads are normally mounted opposite to each other on the drum. The heads protrude slightly from the drum to make contact with the surface of the tape. In the VHS format, the drum has a diameter of about 62mm, and the length of each pass along the skewed tape is about 121mm. Since the tape has a width of 12.65mm, the slant along the tape is thus about 5°.

Sound is recorded conventionally along one or two tracks at the top of the tape for mono and stereo recordings respectively. A control track consisting of pulses at the frame rate is recorded conventionally along a track at the bottom of the tape. The tape moves at about 23.4 mm/s for normal playing times.

The video heads are slightly tilted as they record and play back the signal. This tilt is called azimuth. The heads are tilted in opposite directions to each other, as shown in figure (2.8). One head is +6° and the other -6° from the perpendicular. The result is to minimise interference from adjacent tracks so that the tracks can be recorded immediately next to each other with no physical guard bands. The technique is called a zero band-guard system and greatly increases the length of time that can be recorded on a fixed amount of tape in cassette.

2.2.1 The Omega wrap

As stated earlier, two video heads are normally used in the drum of a helical VTR and they can be seen in figure (2.9). The two head system means that the tape needs only to be wrapped around half the video drum perimeter, with one head joining the tape and beginning its scan as the other leaves the tape after completing its task. To give a degree of overlap between the duty-cycles of the two heads, the tape wrap is in fact about 186°, slightly more than half a turn. This is known as an omega (Ω) wrap. Details of the tape threading, and path through the various rollers and guiders of a commercial VHS video recorder, are shown in figure (2.10).
Azimuth of the heads adjusted to -6° from perpendicular.

Figure 2.8 VHS format azimuth recording system.

Figure 2.9 Omega wrap for a two-head drum. The tape occupies rather more than a half turn of the drum.
Figure 2.10 Tape path and 'linear' sequence of deck components.

55
2.3 **Video-head wear.**

Signal-to-noise ratio (and, therefore, dynamic range) in the tape recording process is a function of head-to-tape separation as defined in the equation

\[ \text{loss} = (54.6 \frac{d}{\lambda}) \text{ dB} \]

where

- \( d \) = length of separation.
- \( \lambda \) = wavelength of signal on tape.

The above equation shows that on the reproduction of low frequency signals in particular, the tape to head distance needs to be kept as small as possible, to maximise the signal to noise ratio. Consequently, this involves contact, friction and erosion of both the tape and the head. The situation is compounded by the need to incorporate very shallow gap depths in order to maximise flux concentration in the gap. It is a fact that heads on wide bandwidth systems, such as video-recorder, wear out after a few thousand hours of use. The replacement of these heads is a costly operation. It follows, also, that abrasion and wear of the coatings on the magnetic tape will also take place. However, an application in which the same piece of tape will be used repeatedly to the point where wearing of the tape becomes noticeable is rare. Consequently, the emphasis is on head wear.

**2.3.1 The causes of wear.**

A committee of the Institution of Mechanical Engineers decided on the following definition of wear: 'the progressive loss of substance from the surface of a body brought about by mechanical action'.

Several mechanisms may be involved in wear, either separately or in unison, and this is particularly true of metals. However, despite the fact that the individual mechanisms are well understood, wear as a general phenomenon remains somewhat unpredictable in quantitative terms.
The generally accepted modes of wear are given below\textsuperscript{11}:

1) Adhesive Wear.
2) Abrasive Wear.
3) Corrosive Wear.
4) Contact Stress Fatigue.

1) Adhesive wear involves the joining, or welding of small interfacing projections in mating surfaces, where material is then torn away from the weaker material by the continuing motion.

The contacting surfaces touch each other at the asperities, and under normal loads, these asperities deform plastically, until the real contact area has increased sufficiently to support the load. Cold welding at these points takes place. On continuation of the tangential motion, these junctions are sheared, and new junctions are formed. Adhesion will clearly occur in cases where head and tape have been in prolonged stationary contact.

A simplified law of adhesive wear was proposed by Archard\textsuperscript{12}, where the wear volume, per unit distance of sliding is given by:

\[
Q = \frac{kW}{3p_0}
\]

where

- \( k \) = probability of an asperity contact producing a wear particle.
- \( W \) = total load.
- \( p_0 \) = yield pressure.

This equation leads to three laws of wear.

1. The volume of wear material is proportional to the distance of travel.
2. The volume of wear material is proportional to the applied load.
3. The volume of wear material is inversely proportional to the yield stress, or the hardness, of the softer material.
The second law is found to be true only at low rates of load. If the load is increased, plots of the form in figure (2.11) are obtained\textsuperscript{13}. These show $k/H$, the adhesive wear coefficient plotted against the average pressure (that is, load over apparent area of contact) for steels of different hardnnesses. It can be seen that $k$ remains constant up to a pressure of about $H/3$ where $H$ is the hardness of the steel, and above this pressure $k$, and hence the wear rate, increase rapidly. It is found that at these higher loadings, large scale welding and seizure occur. These high load effects are not relevant in the case of video head/tape contact.

ii) Abrasive Wear is a term which covers three types of situation. In all cases wear is caused by the ploughing-out of softer material by a harder surface. The first case is when a hard rough surface rubs against a softer surface and is termed gouging abrasion, and the other case is when loose hard particles are caused to slide between rubbing surfaces, this is termed grinding abrasion. The third case is termed erosion abrasion and is due to the passage of small sharp hard particles parallel to a surface. These abrasive particles may be suspended and carried in a fluid such as air or water.

iii) Corrosive Wear involves both corrosion and rubbing. The rate of growth of, say, an oxide film on a steel will decrease exponentially with time, and therefore unless the oxide film is removed by rubbing, the metal-to-oxide reaction will rapidly become negligibly small. The detailed operation of corrosive wear is extremely complex. The reaction products depend on the exact composition of the environment. Small quantities of water vapour in air cause the reaction product to be hydroxide rather than oxide, with the result of an increase in wear.

Corrosion is not always a deleterious phenomenon. Oxide films and other corrosion products prevent adhesion of metal asperities, and can act as a lubricant.

iv) Contact Stress Fatigue can be divided into two main effects\textsuperscript{14}, scuffing and pitting. The cause of scuffing is basically the high temperatures produced at points at intimate contact as a result of high load and sliding speed. The heat generated by the friction leads to localised welding of the opposing surfaces and their subsequent rupture.
Figure 2.11 The variation of wear coefficient with apparent pressure for steel.

a) Brindell hardness 223; b) Brindell hardness 430 [13].
leading to the removal of relatively large fragments from the surfaces. Pitting is a common cause of failure of rolling elements, and is characterised by the formation of small pits on the surface. This is believed to be due to fatigue of the material in the region of maximum shear stress which normally occurs at some depth below the surface.

2.3.2 Wear of magnetic heads.

The wide range of possible causes for wear makes the analysis of head wear, not surprisingly, difficult. Many factors are involved, and need to be carefully measured for a proper analysis, such as the initial shape and surface roughness of the tape and head. Chemical interactions may also be involved. The environmental conditions such as temperature and humidity. The abrasivity of the tape and the speed and length of its passage across the head, together with the applied load all play a part.

Jorgensen reports that abrasive wear and some amount of corrosive wear are largely responsible for head wear. The adhesive wear is observed in the stick-slip motion of a tape at slow speeds, and may in part be the cause of the increase in frictional forces, or drag, with the number of passes, whereby head and tape develop wear patterns which are replicas of one another with resulting larger contact areas and reduced wear.

Abrasive wear is also observed as severe scratches on head and coating surfaces, caused by debris formation or foreign particles. This mode can lead to a destruction of the otherwise perfectly straight gap, and occasionally drags material across the gap (gap-smear), and causes a magnetic short of the head. This method of failure can only be remedied by re-surfacing the head by passing a highly abrasive tape across it.

2.4 Methods of wear measurement.

The problem of magnetic-head wear is as old as magnetic recording. This problem has become more demanding in video however, as the relative head-to-tape speed has increased. The development of more abrasive longer lasting tapes, and market demands for a clearer reproduced image with less noise have also contributed.

60
In order to develop a wear resistant video recording head, it is required to be able to measure the wear on existing heads to evaluate and measure the life expectancy or resistance to wear of various head materials and shapes.

Several techniques have been developed over the years to make these measurements, and some of the most widely used will be reviewed in this section.

2.4.1 Reduction in gap depth.

One of the earliest attempts at measuring the wear of magnetic recording heads was performed by Retinger in 1955. The wear was found by measuring the reduction in the front gap depth. He concluded that the wear rate was proportional to the pressure of the tape on the head.

2.4.2 Weight loss method.

In 1966 Carroll and Gotham published a paper dealing with an abrasion test utilising a 1/2 inch metal rod made from 80-20 Ni-Fe alloys (MuMetal). This rod was cleaned, polished and carefully weighed. After the passage of several thousand feet of tape across it, the rod was removed and weighed again. The weight loss represented the material removed, and is therefore a measure of the wear. The results were published in milligrams per 1000 feet.

The graph shown in figure (2.12) is taken from their paper. It shows that the rate of wear of a head is greatest when the head is new, but is reduced by the passage of an increasing amount of tape. This may be explained by the fact that high points on the head and tape are worn off, which reduces the abrasivity, and the two surfaces begin to conform in shape to each other, so the pressure per unit area is reduced with a resultant reduction in wear.

The authors also investigated the effect of running the tapes in a humid environment. They found that wear of 80-20 Ni-Fe alloys increased rapidly with increasing humidity. Other metals were tried, such as aluminium and copper, and were found not to respond in the same way. Brass
Wear Rate

Tape A, new

Tape A, worn

Tape B, new

Tape B, worn

Figure 2.12 Graph of the rate of wear of a half inch diameter metal rod as a function of tape length passed over the surface. Results taken from [17].
showed approximately half the response of MuMetal. This would seem to suggest that the presence of water causes corrosive wear in the head material.

Although this method clearly gives some very interesting results, there is no evidence to suggest that the wear of a 1/2 inch diameter bar bears any relation to the wear of a video recording head, which has a totally different geometry.

2.4.3 Brass shim test.

The brass shim test first proposed by Daniel and Ragle, was adopted by the U.S. General Services Administration for the measurement of magnetic recording head wear, and is described in detail in Federal Specification No. WT-001553. A summary of this test procedure is described below.

A sample of 0.6 mil thick brass shim with a specific hardness is placed over the reproduced head. A 50000-ft long tape, prerecorded with a 50 kHz signal at 60 inch./sec. is played back in the normal play mode with a tape tension of 8±0.5 oz/inch. The output of the signal is measured initially with the shim in place, and measured again after 50000 feet of tape has passed over the brass shim. The increase in output signal is related to the wear of the shim, which can be calculated using the spacing formula as follows:

Shim wear (mil) = 1.2 × increased output level(dB)/54.6

since the recorded wavelength is 1.2 mils. The wear of the shim due to the abrasivity of the tape should not cause an increase of the output level greater than 2.5 dB for standard and high resolution tape. This corresponds to approximately 55µinch. or 140µm.

2.4.4 The Radicon test.

A more elaborate method called the Radicon test, was proposed by Buchanan and Tuttle in 1968. Radicon derives its name from the Radicon Co., consultants on the design of the tape transport and radioactive heads.
This technique involves fabricating a dummy recording head from Kovar. This is a Carborundum Co. alloy that contains a nominal amount of cobalt (17%), which has been irradiated for a short period, yields abundant gamma radiation, and has a half-life of 5.62 years.

The principle behind this system is that, as the tape is passed over the Kovar head, it wears the surface of the head, and the worn particles become imbedded in the tape. After the passage of about 200 feet of tape, the tape is removed and placed in a special chamber and its radioactivity measured by use of a scintillation counting system. After accounting for background radiation levels, the count is then proportional to the amount of material removed, indicating the abrasivity of the tape used. This method is used in current U.S. National Security Agency specifications, for the measurement of the abrasivity of magnetic tapes, and appears in NSA Specification No. L14-3-75.

Figure (2.13) shows a plot of the variation of abrasivity as a function of wear passes and relative humidity for an Ampex 797A magnetic recording tape, using the Radicon Technique. This is a very complicated technique which gives a good indication of the rate of head wear, assuming that all the head material is deposited on the tape, and none is lost. It gives however, no detail on the spread of the wear across the face of the head, or on the resultant wear pattern.

2.4.5 Thin film method.

An alternative method for the determination of the abrasivity of a magnetic tape and its head wearing capabilities has been reported on by several workers, notably Jorgensen, Cash and Page, and Williams.

The thin film method involves coating a thin film layer of metal onto the surface of a curved object with a similar shape to that of a real recording head. The electrical resistance of thin films is inversely proportional to their thicknesses. As the film wears due to the action of the passing tape the resistance changes and is monitored on a chart recorder. The resistance reading can be calibrated to give a dynamic reading of the wear rate. It could be argued however, that the results obtained by this method cannot be extrapolated to real situations, as the metallurgical
Figure 2.13  Head wear of Ampex 797A recording tape measured using the radicon technique [19].
structure of the deposited films are different from the core material found in actual heads.

2.4.6 Other methods.

Other methods of measuring the tape abrasivity which are discussed in the literature on the subject of head wear include the Ampex Test Element\textsuperscript{23}. This technique involves passing tape at high pressure over the sharp edge of a square cross section bar, made from an Al-Fe alloy. This wears a pad, with a width proportional to the tape wear.

The Spin Physics Test uses a round test element, similar to that of the Ampex element, fabricated from Al-Fe. It is perfectly polished and contains several diamond shaped indentations in its surface. As the tape is passed over this element the wear reduces the measurable dimensions of the indentations, and thus offers a measure of the wear. It is questionable as to what effect the indentations have on localised hardening of the test piece.

The final measurement technique discussed is the Philips-Ball Test Method\textsuperscript{24}. In this test a small steel ball bearing is pressed at a high pressure onto the back surface of the tape, see figure (2.14). When the tape is moved between the ball and the head, a round indentation wears into the head surface. The diameter of the indentation is proportional to the abrasive properties of the tape at high pressures. The pressures used are much greater than would normally be found in a recording system, and so the relevance of the wear results given by this method are clearly questionable.

2.5 Discussion.

In the section above various methods which have over the years been used to measure the abrasivity of magnetic tapes, and their effect on the wear of recording heads has been reviewed. These are methods that are still being used by government agencies, as well as tape and head manufacturers.\textsuperscript{25}
Figure 2.14 Ball test method of wear measurement [24].
The problem with most of these techniques is the fact that they do not employ actual heads for the wear tests. The majority are content with dummy heads, rods, bars or thin film coatings all made from a large range of materials. While these tests no doubt give an excellent indication of the wear of, say, a brass shim when subjected to a number of passes by a specific tape under specific conditions, there is no evidence to link these results to the wear of a recording head from a domestic video-recorder under normal running conditions.

Although, as shown several of the publications have highlighted some useful information regarding the interaction of magnetic tape with certain metallic objects, some serious flaws exist in these tests to justify relating the results to that of video-head wear. Any difference in the physical properties of the test pieces, to that of a real recording head, or any change in the test conditions, such as the use of excessive contact pressures, will invalidate the results.

There is an obvious need for a new and reliable measurement technique, which is not a simulated test, but one which uses an actual recording head where the wear pattern can be measured across the whole face, and not only at specific points. Ideally, the test also needs to have the ability to measure the wear with great precision and sensitivity and to reduce the time taken between measurements.
References.


CHAPTER THREE

DIFFERENTIAL PHASE-QUADRATURE SURFACE-PROFILING INTERFEROMETER.

3.1 Introduction.

The system described in this chapter was developed to measure the surface profile of a video-head. It employs a non-contact interferometric profiling technique which requires no surface preparation; it can work with as little as 10% reflectivity and is insensitive to variations in reflectivity within a scan. It profiles in a linear fashion with no theoretical limitation to the profile length, and by the use of suitable x-y scanners it can profile areas. The system can measure small ($\lambda/4$) discrete steps as well as continuous smooth surfaces. The optical arrangement provides high stability against environmental vibrations and no independent reference or calibration surface is required.

The interference phase calculation and sample scanning is performed by a PC computer, with the surface profile available in near real time.

Height sensitivity is of the order of 1nm and is governed by the 8-bit resolution of the analogue to digital converter used. The lateral resolution is controlled by the spot sizes of the two laser beams and their separation. In the experiments described here this is approximately 10µm.

3.2 Review.

It was decided that due to the particular geometry of the video-head (see section 3.3.3) that a scanning optical probe profilometer was the most appropriate method to measure the surface profile of a video-head.

Several different profiling techniques have been developed over the years. Mechanical profilometers have been used$^1$. This technique can give a height sensitivity down to a few tenths of a nanometre with lateral
resolutions of below 1μm. However, mechanical methods are not always suitable as the stylus applies extremely high local pressures which can result in the surface being scratched, consequently affecting the measurement. For this reason considerable effort has been applied in recent years to design non-contact optical profiling systems.

Optical heterodyne interferometry is the basis for systems which use two concentric laser beams of slightly different frequencies. These are brought to focus onto the sample's surface with different spot sizes. On reflection from the surface the beams are made to interfere and any change in the resulting beat phase is related to variations in the surface height within the smaller focal spot to that of the weighted average within the larger focal spot as the sample is scanned. These systems have excellent height sensitivities, better than a few tenths of nanometres, and good lateral resolutions of a few microns when scanning a randomly rough surface. Limitations are imposed when measuring discrete steps, as the average height, measured by the larger laser spot, is decreased as the spot begins to overlap the edge. This gives an incorrect value for the surface height measured by the tightly focused laser beam.

Another optical heterodyne system published by Sommargren, uses a Wollaston prism to produce two focused but slightly separated laser spots on the surface of the sample. The sample is then made to rotate, with one of the focused spots at the centre of rotation. Unfortunately this system can only produce circular profiles whose radius is fixed by the particular Wollaston prism and objective lens used; typically this is 100μm.

3.3 Principle of operation.

The basic principle of operation of this surface profiling system is differential phase quadrature interferometry. In order to explain this, it is useful to briefly discuss conventional interferometry and its limitations.

3.3.1 Conventional interferometry.

Interference occurs when coherent radiation is split, and its
components are forced to follow more than one path from its source before recombination. The vibrating medium at this point is subjected to the combined superposed effect of the two vibrations and under suitable conditions this leads to stationary waves which appear as light and dark bands which are called interference fringes.

A traditional basis for classifying interferometers is by the method used to divide the light into separate beams. If the radiation from the source passes through one of several apertures, it is said to be separated into beams by division of wavefront. These beams are made up of radiation that has left the source in different directions. Examples of this type are Young's slits, Lloyd's mirror and the Rayleigh refractometer. If the beams are made up of radiation that has left the source in the same direction but is then separated by a beam-splitter, there is said to be division of amplitude. Interferometers of this type include Mach-Zender and Michelson. The interferometer developed in this chapter is of the latter type.

Before discussing specific arrangements of the interferometer system, a derivation of the simple interference equation is shown.

In this representation, the light displacement is described by a complex number

\[ U(r,t) = u_0 \exp(i(\omega t + \varphi)) \]  

\( \ldots (3.1) \)

where

- \( U(r,t) \) complex amplitude of the light.
- \( u_0 \) maximum amplitude of the wave.
- \( \omega \) angular frequency.
- \( t \) time.
- \( \varphi \) initial phase value.

The basic form of a Michelson interferometer is shown in figure (3.1a). Coherent light is emitted from a laser and divided by a beam-splitter. The two components of the beam are made to reflect off a sample and a reference mirror. On reflection the two beams are recombined and illuminate a photo-detector.
Figure 3.1a The basic form of a Michelson interferometer.

\[ I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos \left( \frac{4\pi d}{\lambda} \right) \]

Figure 3.1b Shows the intensity of the interferometer output signal varying in a sinusoidal manner with the change in interferometer arm length differential.
The reference and sample light waves represented by complex amplitudes $U_r, U_s$ are given by

$$U_r = u_r \exp i(\omega t + \varphi_r)$$
$$U_s = u_s \exp i(\omega t + \varphi_s)$$

The total complex amplitude is given by

$$U_T = u_r \exp i(\omega t + \varphi_r) + u_s \exp i(\omega t + \varphi_s) \quad \text{(3.2)}$$

This may be written as

$$U_T = (u_r \exp i \varphi_r + u_s \exp i \varphi_s) \exp i \omega t \quad \text{(3.3)}$$

Since the intensity of a wave is proportional to its amplitude squared, $(u_0)^2$, the intensity is readily found by multiplying the expression for a wave by its complex conjugate donated by $^*$. 

$$I \propto U_T U_T^* = u_r^2 + u_s^2 + u_r u_s \{\exp i(\varphi_r - \varphi_s) + \exp - i(\varphi_r - \varphi_s)\} \quad \text{(3.4)}$$

Using the relationship

$$(\exp ix + \exp - ix) = 2 \cos x$$

equation (3.4) becomes

$$U_T U_T^* = u_r^2 + u_s^2 + 2 u_r u_s \cos (\varphi_r - \varphi_s).$$

Hence

$$I_T = I_r + I_s + 2 I_r I_s \cos (\varphi_r - \varphi_s) \quad \text{(3.5)}$$

The phase of interference varies with the optical path length difference in the arms of the interferometer. Expressing this in terms of optical path length:

$$\varphi_r - \varphi_s = 2 \pi n_0 \Delta x / \lambda \quad \text{(3.6)}$$
where
\[ n_0 = \text{refractive index (≈ 1 in air)} \]
\[ \Delta x = \text{arm length difference} \]
\[ \lambda = \text{wavelength of laser light} \]

As light in a Michelson interferometer performs a double passage of each arm, the path length difference is multiplied by two, as shown above. From this it is clear that the intensity detected by the photo-diode varies in a sinusoidal manner with the change in the interferometer arm length differential, as seen in figure (3.1b).

The sensitivity of a conventional interferometer to surface displacements thus varies between zero and a maximum, depending on the relative optical path difference between the two interfering arms. In order to obtain maximum sensitivity to small displacements, the optical path difference \((\delta)\) should be fixed at

\[(n + \frac{1}{2}) \times \frac{\lambda}{4}\]

\(n = \text{integer}\)

In general, this is only possible with active control of the reference arm length\(^9\).

3.3.2 Phase quadrature.

An alternative method to overcome the above problem, proposed by Peck and Wendell-Obetz\(^{10}\) and Wilkomerson\(^{11}\) is termed phase quadrature. This was originally proposed to remove ambiguity with respect to the direction of surface movement when measuring large \((>\lambda/4)\) displacements with an interferometer.

The various optical components of a quadrature interferometer are illustrated in figure (3.2a). It is essentially two interferometers which use the same optical path. The two interferometer signals are derived from two orthogonally polarised components of light. These are produced by a polarising beam-splitter which splits orthogonal polarisations down the two arms of the interferometer. The recombined beam is then split by a conventional beam-splitter. The outputs of the beam splitter consist of two orthogonal polarisations, the phase relationship between them depending on
Figure 3.2a The optical components of the phase quadrature interferometer.

\[ a = \sin \left( \frac{4\pi d}{\lambda} \right) \]

\[ b = \cos \left( \frac{4\pi d}{\lambda} \right) \]

Figure 3.2b Curve showing the signals on the two outputs of the interferometer are in phase quadrature. This means that the sensitivity is never zero for one or other of the outputs.
the relative optical path difference of the arms of the interferometer. A λ/4 birefringent retardation plate is introduced into one of the outputs of the interferometer and orientated so as to produce a relative retardation of \(\pi/2\). This has the effect of producing a phase shift of 90° in one of the arms. The interference is detected by passing the beams through sampling polarizers orientated at 45° to the beam polarizations, and onto the surface of two photo-diodes.

The signal at one of the photo-detectors will be of the form; see equations (3.5) & (3.6):

\[
\text{Diode 1} \propto 1 + \gamma \cos \left(4\pi \frac{\Delta x}{\lambda}\right)
\]

where

\[
\Delta x = \text{Optical path length difference of interferometer arms.}
\]

\[
\lambda = \text{wavelength of light from laser.}
\]

\[
\gamma = \text{Degree of coherence between the interfering beams, with a value between 0 and 1.}
\]

As the \(\lambda/4\) retardation plate produces a \(\pi/2\) phase change in one of the polarizations relative to the other the response of the second photo-detector will be:

\[
\text{Diode 2} \propto 1 + \gamma \cos \left(4\pi \frac{\Delta x}{\lambda} + \pi/2\right)
\]

or

\[
\text{Diode 2} \propto 1 + \gamma \sin \left(4\pi \frac{\Delta x}{\lambda}\right)
\]

The signals on the two photo-detectors are thus in phase quadrature. This means that the sensitivity is never zero for one or other of the outputs, see figure (3.2b).

The interference term of the outputs can be found by dividing the outputs and taking the arc tangent. Details of this are given in section 3.5.3.

3.3.3 Video-head geometry.

The interferometer described in this chapter is a form of 'Talysurf' where the optical probe replaces the mechanical stylus.
In order to make the translation implied by figure (3.3), the geometry of the laser beam/video-head/drum must be looked at. Simplistically, profiling of the head can be visualised by figure (3.4). The drum is arranged on a shaft which can be rotated about an axis which coincides with the laser interferometer sensing beams and at right angles to it. However, the typical head/drum arrangement is as shown in figure (3.5). The video-head drum has a radius of ~30mm, whereas the radius of the head itself is smaller at ~8mm. This difference in radii poses a problem in rotating the drum through its centre of curvature. A few simple calculations show the problem. When the centre of curvature of both the head and the drum are aligned, the far extremes of the head, B, B', are 0.1 mm below the level of point A. This implies that as the drum rotates from B through A to B' the interferometer must count approximately 300 \([0.1\text{mm/} \lambda/2]\) fringes up and then down.

Furthermore, the fringes do not occur linearly in space, as shown diagrammatically in figure (3.6a). The situation is complicated by optical considerations. Although the theoretical spot size at the focus of the interferometric objective is ~1\(\mu\)m for the laser beam probe, in practice it is rather difficult to achieve less than say 5\(\mu\)m and 10\(\mu\)m is a realistic figure to aim at. It follows that the optical path length across the spot will change by several wavelengths at the extremities of the head, meaning that the fringe visibility is considerably reduced as shown in figure (3.6b). Any profiling technique must therefore remove the contour of the head.

By rotating the head on the drum by about the centre of curvature of the head, it is possible to remove the curvature by presenting an essentially flat and normal surface to the interferometer probe beam. A translation mechanism is required to enable alignment of the axis of rotation of the drum/head assembly with the centre of the head.

3.4 Experimental arrangement.

The experimental arrangement of apparatus is shown in figure (3.7). The output from a low-power (10mW) cw linearly-polarised HeNe laser is expanded to a diameter of 8mm and passed through a polarising cube
Figure 3.3 Basic principle of noncontact profiling method.

Figure 3.4 Principle of measurement applied to curved sample.
Figure 3.5 Typical video head/drum configuration.

Figure 3.6 a) The spread of the detected interference fringes across the head due to the curvature, b) taking into account the reduced visibility of the fringes due to the finite size of the probe beam.
Figure 3.7 The experimental arrangement of the system.
beam-splitter (P.B.S.). The orientation of the electric vector of the incident beam is set to give an output beam intensity ratio in the reflected and transmitted arms of approximately 1:1. The reflected s-polarised component passes parallel to the optical axis of the objective lens O1, (f=85mm) and is brought to a focus onto the sample's surface on the optical axis. The transmitted p-polarised component is reflected off the mirrored surface of a right angled prism with a slight angle $\Delta \beta$ to the optical axis of the objective lens, and is hence brought to a focus on the sample surface at a point separated from the focus of the s-polarised component by $(f \times \Delta \beta)$, figure (3.8).

The two slightly displaced orthogonally polarised focused beams form the basis of the differential phase quadrature interferometer.

The two reflected beams from the test surface pass back through the objective lens O1, and through a cube beamsplitter (B.S.1). This beamsplitter gives two reflected beams with intensities proportional to the reflectivities of the sample surface in the two focal regions. Diodes D1 and D2 monitor the separate beam intensities $I_1$ and $I_2$ respectively

\[ I_1 = A I_a \]
\[ I_2 = B I_b \]  \hspace{1cm} (3.7)

A & B indicate the fraction of total intensity sampled and $I_a$ and $I_b$ are the intensities of the two beams returning from the sample surface.

The mica $\lambda/4$ retardation plate on the output is orientated to lie in the polarisation plane of one of the beams and thus shifts the relative phase of the output beam detected by diode D3 by $\pi/2$.

Sampling polarisers P1 & P2 are orientated at roughly 45° to the beam polarisations. Assuming plane wave interference. The resulting intensity at diodes D3 and D4 can thus be written as shown:

\[ I_3 = A_3 I_a + B_3 I_b + \gamma A_3 I_a B_3 I_b \sin(\theta_a - \theta_b) \]
\[ I_4 = A_4 I_a + B_4 I_b + \gamma A_4 I_a B_4 I_b \cos(\theta_a - \theta_b) \]  \hspace{1cm} (3.8)
spot separation = \( f \times \beta \)

phase angle \( \theta = \frac{2\pi(2d)}{\lambda} \)

8-bit converter \( \therefore \theta = \frac{2\pi}{256} \)

\( \lambda = 633 \text{ nm} \)

\( \therefore d \approx 1 \text{ nm} \)

Figure 3.8 The interferometer arrangement.
The factors $A3, B3, A4$ and $B4$ represent the fractions of intensity sampled by the polarisers $P1$ and $P2$, and $\gamma$ is the degree of coherence between the beams.

The variation in the phase of interference $\Theta$ relates to the difference in the optical path length experienced by the reference and sample beams, where $\Delta z$ is the profile measurement, and since the interferometer is a double pass system, this may be expressed as:

\[(\Theta_a - \Theta_b) = \left( \kappa(2\Delta z) + \phi_{ab} \right) - \Theta \quad \ldots \ldots (3.9)\]

$\Delta z$ = height differential due to surface structure.
$\phi_{ab}$ = constant offset phase due to optics adjustment.
$\kappa$ = wave number of laser light $(2\pi/\lambda)$.
$\lambda$ = optical wavelength.

Effects due to $\phi_{ab}$ are unimportant as it gives a constant gradient to the data which can be simply removed at the final stages of processing.

It should be noted that $\Theta$ depends only on $\Delta z$ and not on the total shift of the sample in the z-direction. Hence this gives the system a good immunity to environmental disturbances.

3.4.1 System operation.

In operation the computer samples the interference of the two phases monitored by the two photo-diodes $D3$ & $D4$, figure (3.7). In addition diodes $D1$ and $D2$, measure the intensity of the reflected beams. This provides the computer with the necessary information to calculate the phase of interference and hence the gradient of the surface $(\delta z/\delta x)_i$ between the probe beams, figure (3.8). The sample is then scanned horizontally a distance equal to the probe beam separation $(\delta x)$ and the gradient of this adjacent surface region is calculated. This process is repeated until the desired line length on the surface has been recorded. The discrete surface profile is then deduced from this sequence of data by the expression

\[z(k\delta x) = \Sigma_{i=1}^{k} (\delta z/\delta x)_i \delta x \quad (k = \text{integer})\]
A sample area can be profiled by measuring a series of lines in a raster fashion.

3.5 Data acquisition and signal processing.

The computer used for control and data processing is an Amstrad PC1520, with a 20Mb hard-disc. All software is written using 'Turbo Pascal'.

All four photo-diodes are simultaneously sampled by 'sample and hold' amplifiers under computer control. The amplifier signals are sequentially digitised by an 8-bit analogue-to-digital converter and sent to the computer.

The electronic circuit designed and built to perform these tasks, shown in figures (3.9a) and (3.9b), can be divided into two halves. The data sampling & analogue-to-digital voltage conversion part and the digital computer interfacing part. These two parts are described below.

3.5.1 Sample & analogue-to-digital voltage conversion.

Analysis of the phase of interference requires the conversion of the optical signal into an electrical voltage or current. The device which performs this is the silicon photodiode. As seen in figure (3.10) the photodiode is operated in a photoconductive mode, with the load resistor $R_f$ in the negative feedback loop of the operational amplifier (A1), so that the bias voltage across the photodiode remains independent of photo-current. The incident light on the photodiode promotes electrons from the valence to the conduction band of the semiconductor, which gives rise to an external current flow. This photo-current passing through the resistor $R_f$ produces a voltage $V_{fd} = I_{pd} \times R_f$ at the input of the amplifier which is directly proportional to the incident light irradiance. The amplifier then buffers the signal to the next stage of the circuit. Due to the fact that the photodiode was remote from the amplifier (A1), pick-up noise was measured on the input terminal of (A1). This noise was reduced below the 1 in 256 digitisation limit of the A/D converter, by using a shielded and earthed cable for the connection.
Figure 3.9a  Schematic diagram of the 'sample, hold & convert' electronic circuit.
Figure 3.9b Detailed circuit diagram of the 'sample, hold & convert' electronics.
Figure 3.10 Diagram showing the BPX-65 photo-diode operated in a photoconductive mode with the load resistor in the negative feedback loop of the operational amplifier.
The detectors used are the BPX 65 photodiodes, which have a stated responsivity of 18.3 A/W when illuminated by light of wavelength 633nm. If the laser delivers a maximum of ≈ 0.5 mW to the photodiode, the photocurrent generated \( I_{pd} \) will be ≈ 90\( \mu \)A. The 8-bit analogue-to-digital converter ZN 427 used has an internal voltage reference of 2.56 Volts. To match this, the photodiode would need the load resistor \( R_f \) to be 2.56V/90\( \mu \)A = 28 kΩ. The load resistor \( R_f \) was a 0 - 50kΩ variable resistor to enable the gain of the amplifier to be adjusted.

The outputs of the 074 operational amplifiers (A1) are used to charge-up four 100pF capacitors. The charging period is determined by the length of time the quad bilateral switch 4066BE remains closed. This, in turn, is controlled by the pulse output by the monostable multivibrator 74121, whose pulse length is set by the external capacitor and resistor used. Initially the pulse length was set to be 2\( \mu \)s long, i.e.sampling over 2\( \mu \)s, and was therefore designed to filter out noise slower than 500kHz. It was found however that due to ringing in the amplifier output, it was necessary to increase this to 20\( \mu \)s, to allow the ringing to settle. This sampling time still removed the effect of noise below 50kHz, which gave the system good immunity from environmental disturbances.

The second set of 074 Op Amps which operate in a voltage follower mode, to remove need for inversion, buffer the signals held by the 100pF capacitors through to the ZN427E analogue-to-digital converter. The capacitors are 'read' by sequentially closing the switches of the 4051BE 8-channel analogue multiplexer. A resistor \( R_c \) (3.9kΩ) is used to protect the A/D converter input from spikes.

3.5.2 Computer Interfacing and data transfer.

The second part of the circuit is involved with sending the digitally converted-voltage signal to the computer. It also receives instructions from the computer for the control and timing of the operation of the electronic circuit.

All information exchanged between the computer and the A/D converter circuit goes via the interface card. This is a commercially available interface card produced by Eltime ltd, Essex. It was designed to
control their video-frame store, using an IBM PC computer, and used in the work in chapters 1 & 5. Due to the availability of the interface card, it was decided to design and build the converter circuit to be electronically compatible with it.

The active pins on the output of the interface card consist of eight data lines (D0 – D7), three address lines (A0 – A2), a clock pulse (Φ0) and a read/write signal (R/W). The addresses sent down pins A0 – A2 are used to control the 3-line to 8-line decoder 74 138, which in turn sends out a pulse on receipt of the appropriate code to different parts of the circuit for control purposes. Pin 11 of the 74 138 is not connected, and the code used to activate it is used in another circuit designed to drive a stepper motor, see section (3.6.1). An octal bus transceiver 74 245 is used to buffer all signals on the data lines to and from the interface card, whose direction of data flow is set by the R/W signal. Signals are sent down three of the data lines from the computer to an octal D-type flip-flop, which in turn relays them to the 8-channel analogue multiplexer 4051 BE. The code carried by these signals are used to sequentially close the switches of the 4051 BE, and thus transfer the analogue voltage held by the 100pF capacitors to the A/D converter chip ZN 427E.

On completion of the A/D conversion, the ZN 427E chip outputs a TTL level signal to indicate the end of conversion. This signal is fed to a transistor BC 548; this sends a 5V signal to the HEX bus driver 74 367, which outputs a signal to the octal buffer to be read by the computer. The end of conversion signal also activates the NAND gate Schmitt Trigger 74 132 which sends a clock pulse to the A/D converter.

It was found that to limit noise in the circuit, it was necessary to have separate power supplies with separate earths for the chips that dealt in analogue voltages to those dealing with the digital logic signals14. The power-supply for the analogue chips contained +5 volts and -5 volts terminals. The digital chips supply gave an output of +5 volts only. The earths of the two halves of the circuit were finally connected together at the earth pin of the A/D converter chip. Figure (3.11) shows the circuit diagrams of the two stabilised power supplies.
Figure 3.11 Circuit diagram of the two stabilised power supplies used to power the analogue and digital parts of 'sample, hold & convert' electronic circuit.
3.5.3 Signal processing.

It is necessary to calculate the interference phase given by equation (3.9), from the values at the photo-diodes represented by equation (3.8). Before this can be done it is necessary to deduce the response of the photo-diodes to both reference and sample arm contributions, such that the coefficients $A_3$, $A_4$, $B_3$ & $B_4$ in equation 3.8 can be normalised away. This is achieved by a calibration routine which is performed only once after the polarisers $P_1$ & $P_2$ have been adjusted. The beams are independently monitored and the required ratios deduced.

With the sample beam blocked:

$$R_3 = A_3/A - I_3/I_1 \quad \& \quad R_4 = A_4/A - I_4/I_1$$

and, with the reference beam blocked:

$$S_3 = B_3/B - I'_3/I_2 \quad \& \quad S_4 = B_4/B - I'_4/I_2$$

given these ratios the reference and sample powers at diodes 3 & 4, can be deduced by the outputs of diodes 1 & 2 in the sampled data by $I_bB_3-I_1S_3$ etc..

The recombined interference term is thus given by:

$$\gamma \sin \theta = \left[ I_3 - (R_3I_1 + S_3I_2) \right] / 2 \sqrt{R_3I_1 + S_3I_2}$$

$$\gamma \cos \theta = \left[ I_4 - (R_4I_1 + S_4I_2) \right] / 2 \sqrt{R_4I_1 + S_4I_2}$$

The interference phase $\theta$ is then deduced by dividing the above expressions and taking the arc tangent.

3.5.4 Removal of $2\pi$ discontinuities.

The resulting data string for a line profile is a set of phase values in the range $-\pi$ to $+\pi$. If a result is close to the range limit it is possible to get spurious $2\pi$ discontinuities between consecutive data samples due to the wrap round from $-\pi$ to $+\pi$ or vice versa. To avoid this
problem, the computer is programmed to assume that $|\Theta_1 - \Theta_{1+1}| < \pi$ and hence it adds or subtracts $2\pi$ accordingly to produce a continuous profile. This limits the height of the discrete steps that can be measured to be $<\lambda/4$.

3.6 Alignment and translation stage.

The angle ($\Delta\phi$) is set by projecting the two near-parallel unfocused probe beams a large distance onto a screen. By slightly rotating the mirrored right angle prism, the desired angle may be selected. The spot separation can thus be adjusted to any suitable value, typically 10-30$\mu$m.

The system was tested initially by profiling flat samples for ease of alignment and translation, with the view to converting it to profiling the curved video-head once the system was proven.

The sample to be profiled is mounted on a precision bearing translator stage, which is driven by a stepper motor controlled by the computer, described in section 3.6.1. The step size is matched to the spot separation selected above.

The focal plane is found by driving the translation stage and sample in a direction parallel to the laser beams. The focal plane is found when the reflected beams emerge as parallel beams from the objective lens. This is confirmed by projecting them over a large distance onto a screen.

The sample surface has to be carefully aligned for motion perpendicular to the laser probe beams to avoid any defocusing of the beams on long scans. Any defocusing by an amount exceeding the depth of focus of the objective lens, or random sample tilt due to environmental noise or a bad translation stage, will change the value of $\phi_{ab}$ during a scan and cause unrecoverable errors; see equation (3.9).

The sample tilt errors could be greatly reduced by making the probe spot separation as small as possible as the height error due to sample tilt equals the spot separation $\delta x$ multiplied by the tilt angle $\alpha$. For a spot separation of 10$\mu$m and a random sample tilt of 0.1$\mu$rad, a 1nm error is expected.
The orientation of the $\lambda/4$ retardation plate needs to be set accurately to avoid introducing sinusoidal phase errors. A $1/2$ degree error in the orientation of the phase plate will give rise to a $0.5\%$ phase error on the output. The alignment was achieved by slowly adjusting the plate whilst constantly monitoring the signals on photo-diodes D3 & D4. By taking the square root of the sum of squares of the signals a value of unity was measured when the retardation plate was giving an exact $\pi/2$ phase change.

3.6.1 Stepper-motor driver.

The motor used to drive the precision translator is an RS 332-947 4-phase stepper motor with a $7.5^\circ$ step angle. This is driven by a SAA1027 stepper-motor driver.

As stated in section 3.5.2, pin 11 of the 3 line to 8 line decoder 74 138 chip is not connected. By using a Y connector on the ribbon cable which connects the interface-card to the sample & A/D converter circuit, it is possible to use the same interface card to address the stepper-motor driver circuit.

Figure (3.12) shows the circuit diagram of the motor driver. To drive the motor, the computer sends the binary number 772 address to the port, see diagram (3.9b). This activates pin 11 on the 74 138 chip, which enables the voltage translator 4104. The numbers sent to the relevant port address activate the appropriate data lines, which after conversion to 12 volts, clock, and change the direction of, the motor's motion.

3.7 Errors and sensitivities of system measurements.

3.7.1 Main sources of measurement errors.

The photodiode is the first element in the detection system. Thus it is important to determine the errors introduced by it.

A diode under reverse bias exhibits a voltage-dependent capacitance caused by the variation in stored charge at the junction ($C_j$).
Figure 3.12 Circuit diagram of electronics used to interface to a computer, and control the stepper motor driver.
This capacitance, together with the load resistor ($R_L$), act as a low pass filter. The electrical bandwidth, $\Delta f_{el}$, is defined as the frequency range over which the output is above $(1/2)^{\frac{1}{3}}$ of its maximum, and is given by\textsuperscript{15}:

$$\Delta f_{el} = \frac{1}{2\pi R_L C_j}$$

The junction capacitance is given as 6pF\textsuperscript{16} when operated at a reverse bias of 5 volts. The load resistor was set at 28kΩ (see section 3.5.1). This gives a bandwidth of about 1 MHz.

The voltage generated across the load resistor will consist of voltage due to signal, and that due to noise. The noise is due to two main effects.

1. Johnson or thermal noise in the resistor.

   This noise arises due to the thermal agitation of charge carriers within a conductor. The random nature of this motion results in a fluctuating voltage appearing across the conductor. It is also known as Nyquist or black-body noise and is given by\textsuperscript{16}:

   $$v_t = (4kTBR)^{\frac{1}{2}}$$

   where $k$ is Boltzmann's constant, $T$ is the absolute temperature, $B$ is the bandwidth of the system and $R$ is the resistance of the conductor.

   Substituting in appropriate values at 20°C gives:

   $$v_t = 2 \times 10^{-6} \text{ Volts}.$$ 

2. Shot noise.

   Shot noise is encountered whenever there is a charge flow and arises directly from the discrete nature of the electronic charge. Thus when a charge flows past any point in a circuit, the arrival rate will fluctuate slightly. As a result, the photocurrent has a mean value about which random noise fluctuations occur. Also known as white noise, it is given by\textsuperscript{16}.
\[ i_s = (2eI_aB)^\frac{1}{2} \]

where \( e \) is the electronic charge, \( I_a \) is the average photocurrent and \( B \) is the bandwidth of the system.

Substituting appropriate values gives:

\[ i_s = 5 \times 10^{-9} \text{amps}. \]

With a load resistance of 2.8\( \text{k\Omega} \), conversion to volts gives:

\[ v_s = 150 \times 10^{-6} \text{volts}. \]

Another possible source of noise is the amplifier. The type chosen for this system is a low noise BIFET op amp (074). This amp has a quoted\(^7\) equivalent input noise voltage of \( 18\text{nV/(Hz)}^\frac{1}{2} \). This corresponds to a noise voltage in the system of:

\[ v_{\text{amp}} = 1.8 \times 10^{-11} \text{volts}. \]

It is clear that for the photo-diode, Shot noise is the dominant source of noise in the system.

3.7.2 Limitations and sensitivity of system.

The ultimate resolution limit is determined by the spot separation, and the measured height at a particular surface point is actually a weighted average over the spot size. The sensitivity is limited by shot noise in the photo-diodes and the accuracy of the \( \lambda/4 \) phase plates, which will introduce sinusoidal phase errors. The signal-to-noise ratio for the photo-diodes used in this system is estimated to be approximately 7000/1. This value is much larger than the 256/1, 8-bit digitising limit, imposed by the analogue to digital converter used. It would be possible to use a 12-bit digitiser to give the system sensitivities of less than a tenth of a nanometre while still being below the photo-diode shot-noise limit. However a \( 1/2 \) degree error in the \( \lambda/4 \) retardation plate, gives a phase error of approximately 0.5%; hence great care needs to be taken when aligning this component. In this experiment the retardation plate was set to better than a \( 1/2 \) degree, giving an estimated phase error of 200/1.
3.7.3 Defocus effects.

The system gives excellent results provided the depth of focus of the objective lens is not exceeded. Figure (3.13) shows the phase errors introduced by stepping the sample towards and away from the objective lens. As the two adjacent spots are continuously monitoring the same positions on the sample, one would expect the path length difference to remain constant by translating the sample in a direction parallel to the laser beams. This path length change may be explained by various effects. It is first necessary to consider the effect any defocusing of the sample has on the lateral displacement of the reflected beams.

It is clear from figure (3.14) that due to the fact that the two beams impinge on the sample at different angles their reflections will undergo lateral displacements of differing amounts on defocusing of the sample. Hence they will strike the objective lens and emerge in a plane parallel to that taken by the rays emerging from the focus, but with a slight lateral displacement, with the beam nearest the edge of the lens displaced more than the one nearer the centre.

It can be shown that for a 3μm defocusing of the sample, the outside beam emerges with a 1.7μm displacement. This on its own will not affect the phase measurement, as this displacement is less than 0.01% of the diameter of the return beam. Therefore the required superposition of the return beams is not affected.

3.7.3a The wedge effect.

To produce the two slightly separated focused spots on the sample the incident laser beams have a slight angle between them (≈ 0.1μrad). For the two recombined beams to emerge from the last cube beam splitter completely overlapped, there needs to exist a slight wedge angle (50μrad) in the mirrored right angled prise, see figure (3.7). As the return beams are displaced on reflection due to defocusing, the inside beam will strike the wedge at a different position, thus changing the optical path length. It is estimated that a 10μm defocus will only result in a 0.1mm change in the optical path length. This may be ignored.
Figure 3.13 The phase errors introduced on the output signals caused by defocusing of the sample.
Figure 3.14 diagram showing the lateral displacements of the return laser beams on defocusing of the sample.
3.7.3b Change of curvature of wavefront effect.

As seen in figure (3.8), on reflection, the two beams emerge from the objective lens at different positions. This is not a problem when the sample is in focus, as light originating at the focal plane of a lens will emerge with a plane wavefront across the whole face of the lens. However, light originating from a point beyond the focal plane of a lens emerges from the front of the lens with a curved wavefront. A very simplified mathematical model was developed to calculate the amount of defocusing required in the lens system to produce a measurable phase change between two interfering rays of light, emanating from a point source and passing through fixed points on the objective lens.

The complex field $U'_1(x,y)$ across a plane immediately behind the lens is related to the complex field $U_1(x,y)$ incident on a plane immediately in front of the lens by:

$$U'_1(x,y) = t_1(x,y) \cdot U_1(x,y)$$

where $t_1(x,y)$ is the multiplication phase transformation representation of the lens.

It has been shown by Goodman\(^{18}\) that:

$$U'_1(x,y) = \exp[jkn\Delta_0] \cdot \exp[-jk/2f (x^2+y^2)]$$

where $n$ is the refractive index of the lens glass, and $\Delta_0$ is known as the Thickness function of the lens.

The first term is simply a constant phase delay. The second is interpreted as a quadratic approximation to a spherical wave. If $y=0$ (assume a one dimensional system), the expression becomes:

$$\exp -jk/2f(x^2)$$

The expression for the difference in phase between $d$ & $f$, figure (3.14), becomes:

$$\exp jk(x^2/2d) \cdot \exp-jk(x^2/2f)$$
where \( d=f+\Delta l \), and \((1/f - 1/d) \approx (\Delta l/f^2)\) so this becomes:

\[
\exp \left[ jk\left( x_1^2 - x_2^2 \right)/(2f^2) \right]
\]

If the outside beam emerges from the objective lens at a distance \( x_1 = 15 \text{ mm} \), and the inside beam at \( x_2 = 7 \text{ mm} \), from the optical axis of the objective lens, the difference in the phase of rays emerging from these two points on the lens can be expressed by:

\[
k\Delta l/(2f^2) \left( x_1^2 - x_2^2 \right)
\]

If we assume the phase difference is \( \pi/3 \) of a wave i.e. \( \approx 100 \text{ nm} \) optical-path-length difference, \( k \) is the wave number \( = 2\pi/\lambda \) and \( f \) is the focal length of the objective lens, 85\text{ mm}, we can calculate \( \Delta l \) the defocus needed to produce a 100\text{ nm} phase change in the output beams. This is found to be \( \approx 9 \mu\text{m} \).

This figure is of the correct order of magnitude with the figure found by experiment, see graph in figure (3.13).

It should be stated however that this is a very simplified model. It does not take into account the fact that the two focus positions are not point sources emanating from the optical axis of the lens, and that the reflected beams occupy a finite area of the lens, whereas the model calculates the phase difference at specific points on the lens. It also does not take into account the movement of the beams across the face of the lens on defocusing. It does however give an order of magnitude figure for the amount of defocus required to produce a known phase change, which serves to back-up the experimental results, and helps to explain the cause of the phase change on defocusing of the sample.

A method that could be used to measure and correct for the defocusing of the sample, employing a quadrant split photodiode and a feedback servo system is proposed in chapter 7. This technique is similar to that found in compact disc players for maintaining the focus of the laser probe beam.
Although, this proposed method is highly suitable for correcting the defocus of the sample, it was not pursued with further, as a decision had been made to abandon the use of the system as a means of measuring the wear on video-heads, due to the prohibitive cost and complexity of the mechanical system needed to move and manipulate the position of the video-head.

3.8 Results.

Two types of sample were used to test the performance of the system. The first was a discrete step produced by depositing a thin film of aluminium on a glass microscope slide which had previously had half its surface masked off. The whole surface of the slide was then coated with another thin layer. This step was measured and the step height was compared to that obtained using a Talystep stylus machine from Rank Precision Industries Ltd, figure (3.15). It must be noted that the line scan across the step using the two different techniques was not necessarily exactly over the same part of the step, but it was as close as possible. There is however excellent agreement in the general shape and height of the step.

The second sample was a poor quality mirror. This surface was scanned in both the forward and reverse directions to test the reproducibility of the system. The results shown in figures (3.16a) & (3.16b) demonstrate the system's excellent reproducibility. This would seem to indicate that the translation stage motion is highly consistent.

The stability of the system was checked by attempting a scan with the stepper motor driver disabled, and for a duration equal to that taken on scanning the poor quality mirror. Since all the data is taken between the same two points, the plot should be that of a straight line. The results in figure (3.16c) demonstrate that the system is immune to normal environmental disturbances.

It is shown that this system is capable of performing measurements on both continuous surfaces and discrete steps. There is a limitation on the measurement of surface structure in that the difference in height between the two probe beams Δz must be < λ/4 to get a true profile. This is not usually a problem on very smooth surfaces.
Figure 3.15 Line profile across a thin-film edge using:

a) Differential phase quadrature surface profiling interferometer, and
b) Rank precision Industries Ltd. Talystep stylus machine.
Figure 3.16 Surface profile of an ordinary quality mirror in the:
    a) forward direction and b) reverse direction, to demonstrate the reproducibility of the system. Also shown is a scan c) with the translation stage disabled to demonstrate immunity against environmental disturbances.
3.9 Conclusions.

Although the differential phase-quadrature interferometer was designed, built and tested with the measurement of video-head wear in mind, it was finally concluded that due to the complexity and precision of the mechanics which would be required to hold and manipulate the position and movements of the video-head, and with the addition of a servo system to correct for defocusing, the cost would be prohibitive.

The differential phase-quadrature interferometer can nevertheless be successfully utilised as a tool for general surface profile applications. Sensitivities in the order of 1nm can be obtained, limited by the 8-bit analogue-to-digital converter used, provided that the orientation of the $\lambda/4$ mica retardation plate is correctly set. Spatial resolution of the system is limited by the focal spot size and spot separation, typically 10$\mu$m in both cases. Care however must be taken to ensure that the sample is maintained within the depth of focus of the two probe beams.
References.


CHAPTER FOUR.

REVIEW OF PHASE SHIFTING INTERFEROMETRY.

4.1 Introduction.

Phase shifting interferometry (P.S.I) is a recently developed technique, which has found wide use in the measurement of surface figure, surface roughness, and the metrology of various objects and surfaces in a range of applications\textsuperscript{1,2,3}.

In P.S.I. the phase difference between the two interfering beams is varied in a known manner, and measurements are made of the intensity distribution across the pupil corresponding to at least three or more different phase shifts.

Phase-measurement techniques have been applied to a large range of interferometer systems such as Twyman-Green, Mach-Zehnder, Linnik and Mirau among others.

In order to review phase-shifting interferometry (P.S.I), it is necessary to understand what phase measuring interferometry is and how it differs from conventional interferometry, in terms of it's use in the measurement of the shape of objects, by measuring the relative heights of the wavefront of light reflected off the surface. In conventional interferometry, the observer views the entire interferogram and performs a complex evaluation process in order to determine an estimation of the shape of the object which is producing the interference pattern. This is usually done by measuring the position of the centre of the interference fringes, and by the use of some form of interpolation obtain the necessary data points between these fringe centres, in order to obtain a uniform grid of data for analysis. This interpolation process can introduce error into the results, also it is usually difficult to determine the centre of an interference fringe, to better than \(1/10\)th of a fringe. In order to increase the number of data points across the field of view, more fringes can be generated by introducing tilt into the interferogram. This however, reduces the accuracy as it makes the fringes narrower and hence makes the task of determining the position of their centres more difficult. The
accuracy of the whole process depends greatly on the skill and experience of the operator to properly select the number of fringes needed to completely sample the object, process out the effects of intensity and contrast variations, and estimate the amount and polarity of a feature's fringe deviation. By polarity it is meant the determination of the increase or decrease of the height of a feature on a surface relative to another point, as the interference fringe pattern produced by a pimple on a surface, is almost identical to one generated by a crater of similar dimensions.

In phase measuring interferometry, instead of viewing the interference pattern as a whole, a discrete set of independent estimations of the wavefront under test is produced at specific, well-defined geometrical positions (detector sample points) such that the pupil sampling is determined by the detector coverage and not by the number of fringes in the interference pattern. The resultant array of wavefront phase estimates can be presented to the operator for evaluation or processed further in order to analytically describe the shape of the object under test. The algorithms which produce this array of phase information have the added advantage of being independent of irradiance and contrast variations, and being able to automatically determine the polarity of a feature. Although operator skill and experience is still important, it is the ability of phase-measurement interferometry to accurately and repeatably acquire the independent wavefront phase estimates that determine its utility.

4.2 Methods of phase shifting or modulation.

In order to determine the phase of a wavefront, a temporal phase modulation, or relative phase shift between the reference and measurement arms of an interferometer, is introduced into the measurement. By measuring the interferogram intensity at specific locations, as the phase is shifted, the phase of the wavefront can be determined with the aid of electronics or a computer.

Perhaps the most common and straightforward method of producing a phase shift in an interferometer, is by the positioning of a mirror, pushed by a piezo-electric transducer, in the reference beam arm of the
interferometer\textsuperscript{4,5}. Many types of piezo-electric translators are available, which can drive a mirror linearly over \( \mu \text{m} \) range. These are driven by a high voltage power-supply, which can produce a ramped voltage of zero to several hundred volts.

Other techniques are available, (or have been used), such as a rotating half-wave plate or, equivalently, a rotating quarter-wave plate used in double pass, in an interferometer with polarisation isolation, to produce a frequency shift at twice its rotation frequency. (A rotation of 45° will produce a phase shift of \( \pi/2 \))\textsuperscript{6,7,8}. Using reasonable rotation rates, it is not possible to obtain frequency shifts much larger than 1 or 2 kHz, employing this method. For greater frequency shifts Shagam and Wyant\textsuperscript{9} have proposed a system which comprises of a series of rotating half-wave plates with a stationary half-wave plate between consecutive rotating half-wave plates. The resulting frequency shift is \( N\omega/\pi \), where \( N \) is the number of rotating half-wave plates and \( \omega \) is angular rotation frequency of each rotating half-wave plate.

Another phase modulation technique continuously moves a diffraction grating in one arm of an interferometer. Diffraction gratings shift the frequency of the \( N^\text{th} \) order by an amount \( NVf \), where \( V \) is the velocity component of the grating perpendicular to the grating lines and \( f \) is the spatial frequency of the grating. Note that, for a grating, the frequency shift is independent of wavelength. Instead of using a grating, the same effect can be produced by passing one of the arms of the interferometer through an acoustic-optic Bragg cell. In the Bragg cell, the travelling acoustic wave acts as the moving grating. The frequency shift of the first diffracted order is equal to the frequency used to drive the Bragg cell, independent of the wavelength of the light\textsuperscript{10}.

Another method to obtain a frequency shift is to use Zeeman splitting of laser radiation\textsuperscript{11}. This is achieved by placing a stabilised laser source in an axial magnetic field to yield two laser lines, having different frequencies. These two frequencies are circularly polarised with opposite senses and differ in frequency by several megahertz.
4.3 Methods of phase detection or measurement.

The techniques used for the detection and measurement of phase can be divided into two main categories, electronic and analytical. With the advent of solid-state charge-coupled-device (C.C.D.) array detectors, and the availability of powerful desk-top computers, the analytical methods and their algorithms have become the predominant means for determining the phase of interference. However, for completeness, various electronic methods will be briefly reviewed first.

One method of electronic phase detection, described by Wyant\textsuperscript{10}, is termed zero-crossing, as it utilizes the detection of a modulated test signal passing through a zero phase value in relation to a modulated reference signal. In this technique a clock starts when the test signal passes through zero volts, and stops when the reference signal passes through zero volts. The phase difference is simply the time measured between crossings divided by the period of the reference signal. To increase the accuracy of this system, the sinusoidal signals are usually greatly amplified to yield a square wave to improve the zero crossing detection.

Another technique that has been commented on by several researchers is termed phase-lock or a.c. interferometry\textsuperscript{12,13,14,15}. In phase-lock techniques the signal in the reference arm is modulated sinusoidally, producing temporally-modulated interference terms with cosine of sine and sine of sine dependences. The detected optical signal will contain terms with odd and even order harmonics of the phase modulation frequency. A second, lower frequency phase shifter is used to change the path between the two beams in the interferometer until the odd order harmonics, including the fundamental, disappear. When high order harmonics are filtered out, the resulting electrical signal is directly proportional to the optical phase modulo $2\pi$.

The third electronic method for determining optical phase discussed here is one which employs up-down counters\textsuperscript{10}. The output of a photo-diode which monitors the test signal is connected to the up terminal of an up-down counter. A second photo diode, monitoring the intensity of the two interfering beams, is connected to the down terminal of the
counter. This second signal is the reference signal. If both the reference and test signals are of the same frequency, the up-down counter will register a zero. If however, the test frequency increases, which would occur when the test detector scans through several fringes, the up-down counter will give an output equal to the number of fringes the test detector scans through. Up-down counters are usually used in conjunction with frequency multipliers in order to measure the phase in units of less than a fringe.

All the above techniques have the disadvantage that to produce a two-dimensional phase map, the area of interest needs to be scanned using x-y translators. A detector array cannot be used. The up-down counter technique has the added disadvantage that it only measures changes in phase rather than absolute phase values.

4.4 Phase measurement algorithms.

The use of whole field techniques to detect and measure optical phases has several advantages over electronic methods. Modern solid-state C.C.D. arrays can be used as detectors to make measurements simultaneously at a very large number of points covering the interference pattern. Creath\textsuperscript{16} has developed a system which can sample an area using a 1.4 million element array. With the advent of large and fast computers and framestores, it is possible to make measurements rapidly and store the data. The effects of vibrations and air currents can be minimised by averaging a number of observations. Similarly, errors due to the interferometer optics can be eliminated by subtracting readings made without a test piece, or with a standard, from readings made with a test piece\textsuperscript{3,17}.

4.4.1 Sampling requirements.

Fringe modulation is a fundamental problem in all phase-measuring techniques\textsuperscript{2}. When an interference fringe is sampled by a detector array, the output consists of discrete voltages representing the incident intensity, integrated over the exposure time and averaged over the area of
the detector element. As the relative phase is shifted between the reference and object beams, the incident intensity on the detector element changes, as shown in figure (4.1). If the interference data are sampled at the Nyquist frequency, such that there are two detector elements for each fringe (each half wave of optical path difference), then the wave-front can be correctly reconstructed. If, however, the fringe pattern is not adequately sampled, the wave-front can not be correctly reconstructed. As long as there is less than half a fringe per detector, the intensity will be modulated. Figure [4.1d] shows a detector element viewing one fringe, hence there will be no modulation.

The modulation at specific points of the interferogram can be calculated from the phase-shifted data. If the modulation of a point is less than a predetermined threshold due to the sampling requirements just described, or indeed, due to extraneous scattered signals incident upon the detector array, then that data point should be ignored.

As the detector size influences the recorded fringe modulation, to ensure high intensity modulation, the size of the detector elements should be small compared to the fringe spacing. This may be accomplished, in a high density fringe situation, by imaging a mask of small apertures, separated by the spacing of the detector elements, onto the face of the detector array.

4.4.2 General phase measurement techniques.

To construct a wavefront map, the phase of the wavefront can be measured directly by changing the relative phase between two beams in an interferometer. This modulates the intensity of fringes incident upon a detector array. While the phase is shifted, several sets of fringe intensities are recorded. A wavefront map is then constructed by simple calculations inside a computer. Phase mapping techniques are faster than digitising fringes and produce more accurate results.

The most common way to shift phase is by translating a reference mirror with a piezoelectric transducer. This technique works very well, but it requires a translator whose motion is linear with respect to the applied
Figure 4.1  A) Sufficient sampling with point detectors showing high modulation of the intensity as the interferometers relative phase is shifted by $\pi/2$. B) Undersampling with point detectors showing aliasing of fringe pattern. C) Sufficient sampling with finite-sized detectors with $<\frac{1}{2}$ fringe over the detector area. D) Undersampling with finite sized detectors showing no intensity modulation. Figure taken from [18].
voltage. As the phase is shifted in the interferometer, the detector array will integrate the fringe intensity data over a change in relative phase of the two interfering beams. Greivenkamp\textsuperscript{19} has shown that one set of intensities can be written as:

\[
I_1(x,y) = \frac{1}{\Delta} \int_I I_0(x,y)[1 + \gamma_0 \cos(\phi(x,y) + \alpha(t)) \alpha_1 - \Delta/2) \] \ldots (1)

where \( I_0(x,y) \) is the average intensity at each detector point, \( \gamma_0 \) is the modulation of the fringe pattern, \( \alpha_1 \) is the average value of the relative phase shift for the \( i \)th exposure, and \( \phi(x,y) \) is the phase of the desired wavefront. After this integration is performed, the recorded intensity is:

\[
I_1(x,y) = I_0(x,y)[1 + \gamma_0 \text{sinc}(\Delta/2) \cos[\phi(x,y) + \alpha_1] \] \ldots (2)

where \( \text{sinc}(\Delta/2) = \sin(\Delta/2)/(\Delta/2) \). The only difference between integrating the phase and stepping the phase is a reduction in the modulation of the interference fringes after detection. If the phase shift were stepped (\( \Delta = 0 \)) and not integrated, the sinc function would have a value of one. Therefore, phase-stepping is a simplification of the integrating-bucket technique. At the other extreme, if \( \Delta = 2\pi \), there would be no modulation of the intensity. Since this technique relies upon a modulation of the intensities as the phase is shifted, the phase shift per exposure should be between 0 and \( \pi \). To reconstruct a wavefront, a minimum of three sets of recorded fringe data are needed because of the three unknowns \( I_0, \gamma, \) and \( \phi \).

4.4.3 Three-bucket or three-step technique.

One way to reconstruct a wavefront is to use a method described by Wyant et al\textsuperscript{20}, called the three-step technique, in which the detector array transfers the detected intensity values of the interferogram into computer memory. The phase of the reference beam is then shifted by 90° and
the detector array 'read out' again. The phase difference is shifted by a further 90° and the detector array 'read out' for a third time. The intensity described by equation (2), for the three frames of data, corresponding to \( \alpha_1 = \pi/4, 3\pi/4 \) and \( 5\pi/4 \), is given by:

\[
I_1(x,y) = I_0(x,y)[1 + \gamma \cos[\phi(x,y) + \pi/4] ]
= I_0(x,y)[1 + \gamma(\cos\phi(x,y) - \sin\phi(x,y))/2] \ldots (3)
\]

\[
I_2(x,y) = I_0(x,y)[1 + \gamma[\cos\phi(x,y) + 3\pi/4] ]
= I_0(x,y)[1 + \gamma(\cos\phi(x,y) - \sin\phi(x,y))/2] \ldots (4)
\]

\[
I_3(x,y) = I_0(x,y)[1 + \gamma\cos[\phi(x,y) + 5\pi/4] ]
= I_0(x,y)[1 + \gamma(\cos\phi(x,y) + \sin\phi(x,y))/2] \ldots (5)
\]

where \( \gamma = \gamma_0 \text{sinc}(\Delta/2) \).

It is clear that for an integration period \( \Delta \) of \( \pi/2 \), the modulation of the interference fringes \( \gamma \) equals \( 0.9\gamma_0 \). Thus, linearly ramping the phase shifter while taking measurements will have little effect on reducing the modulation of the interference fringes, and thus compares favourably to stepping the phase shifter discrete amounts, and allowing time to settle, before reading the data.

The phase \( \phi(x,y) \) at each point is obtained by manipulation of equations (3),(4) & (5) so that:

\[
\phi(x,y) = \tan^{-1}\left[ \frac{I_3(x,y) - I_2(x,y)}{I_1(x,y) - I_2(x,y)} \right] \ldots (6)
\]

Besides a reduction in intensity modulation due to sampling of the detector, at the same time as shifting the phase, the finite size of the detector elements and any extraneous scattered signals incident on the detector array will also act to reduce the intensity modulation. To make reliable phase measurements, the intensity must modulate at each point sufficiently so as to give a correct phase measurement. The value of the intensity modulation can be calculated from the intensity data given, using the equation:

\[
\gamma(x,y) = \frac{[I_1(x,y) - I_2(x,y)]^2 + [I_2(x,y) - I_3(x,y)]^2}{2I_0} \frac{1}{2I_0}
\]

119
This expression can be used to determine if a particular data point will yield an accurate phase measurement or should be ignored.

An alternative method proposed by Hariharan et al\(^\text{21}\) involves finding the phase profile, by taking three frames separated by phase steps of \(2\pi/3\).

For a phase shift other than \(\pi/2\) or \(3\pi/2\), the phase profile can be calculated using\(^\text{22}\):

\[
\Phi(x,y) = \tan^{-1}\left(\frac{\left\{1 - \cos(\alpha)\right\}/\sin(\alpha)\left\{[I_1(x,y) - I_3(x,y)]/[2I_2(x,y) - I_1(x,y) - I_3(x,y)]\right\}}{1 + \gamma\cos[\Phi(x,y) + \alpha]} - I_0(x,y)[1 - \gamma\sin[\Phi(x,y)]\right\}]
\]

where phase shifts of \(-\alpha\), \(0\), and \(\alpha\) are assumed. There are other permutations of these equations, but the foregoing are the most commonly used.

### 4.4.4 Four-bucket or four-step techniques

Another method which is commonly used to calculate the phase profile of a set of fringes is the four-bucket or four-step technique\(^4\). This method is similar in principle to the three-frame technique described in the section above, except this method involves four recorded intensity measurements as shown below:

\[I_1(x,y) = I_0(x,y)[1 + \gamma\cos[\Phi(x,y) + (0)]]\]
\[= I_0(x,y)[1 + \gamma\cos[\Phi(x,y)]] \quad \text{... (7)}\]

\[I_2(x,y) = I_0(x,y)[1 + \gamma\cos[\Phi(x,y) + (\pi/2)]]\]
\[= I_0(x,y)[1 - \gamma\sin[\Phi(x,y)]] \quad \text{... (8)}\]

\[I_3(x,y) = I_0(x,y)[1 + \gamma\cos[\Phi(x,y) + (\pi)]]\]
\[= I_0(x,y)[1 - \gamma\cos[\Phi(x,y)]] \quad \text{... (9)}\]

\[I_4(x,y) = I_0(x,y)[1 + \gamma\cos[\Phi(x,y) + (3\pi/2)]]\]
\[= I_0(x,y)[1 + \gamma\sin[\Phi(x,y)]] \quad \text{... (10)}\]

where \(\alpha = 0, \pi/2, \pi\) and \(3\pi/2\); Using equations (7), (8), (9) & (10) it is
possible to find the expression needed to measure the phase at a particular position given by:

\[ \phi(x,y) = \tan^{-1}\left[ \frac{I_4(x,y) - I_2(x,y)}{I_1(x,y) - I_3(x,y)} \right] \]  \hspace{1cm} (11)

As in the case of the three-step technique, the recorded fringe modulation \( \gamma \) will equal \( \gamma_0 \) for discrete phase steps, and \( (0.9)\gamma_0 \) for an integrating-bucket where the phase is integrated over \( \Delta-(\pi/2) \).

4.4.5 Carré technique.

The phase measurement algorithms using the three and four frame methods described above assume that the induced phase shift is known, either by measuring the phase shift after every step, or by calibrating the phase shifter.

An alternative method proposed by Carré 23, originally for length measurement, calculates the phase of the wavefront independently of the amount of phase shift. Shifting the phase an amount \( \alpha \) between four successive frames, yields the intensity equations shown by:

\[
\begin{align*}
I_1(x,y) &= I_0(x,y)[1 + \gamma \cos(\phi(x,y) - (3\alpha/2))] \\
I_2(x,y) &= I_0(x,y)[1 + \gamma \cos(\phi(x,y) - (\alpha/2))] \\
I_3(x,y) &= I_0(x,y)[1 + \gamma \cos(\phi(x,y) + (\alpha/2))] \\
I_4(x,y) &= I_0(x,y)[1 + \gamma \cos(\phi(x,y) + (3\alpha/2))]
\end{align*}
\]

where the phase shift \( \alpha \) between frames is assumed to be linear. From these equations \( \alpha \) can be found using:

\[
\tan(\alpha/2)(x,y) = \frac{\{3[I_2(x,y) - I_3(x,y)] - [I_1(x,y) - I_4(x,y)]\}}{\{I_2(x,y) - I_3(x,y)\} + [I_1(x,y) - I_4(x,y)]} \] \hspace{1cm} (12)

and the phase at each point is given by:

\[
\tan\phi(x,y) = \tan(\alpha/2)(x,y)[I_2(x,y) + I_3(x,y)] + [I_1(x,y) - I_4(x,y)]] \\
/[[I_2(x,y) + I_3(x,y)] - [I_1(x,y) + I_4(x,y)]] \] \hspace{1cm} (13)
Combining equations (12) & (13) gives:

\[
\tan \Phi = \frac{[3(12 - 13) - (11 - 14)][(12 - 13) + (11 - 14)]^2}{[(12 + 13) - (11 + 14)]}
\]

The Carré technique has the advantage of working when a linear phase shift is induced in a converging or diverging beam where the amount of phase shift across the width of the beam is not constant in contrast to the three and four-frame methods which require collimated beams.

4.4.6 Synchronous detection.

Synchronous detection of interference fringes is an early technique used for phase measurements, and which is equivalent to synchronous-detection in communication theory\textsuperscript{17}. If \(I_I\) is the intensity of the interference pattern obtained by shifting the phase, \(N\) times, by an amount \(\alpha_i\) such that:

\[\alpha_i = i \cdot 2\pi / N, \text{ with } i = (1 \ldots N).\]

then the phase \(\Phi\) can be measured using the equation below:

\[
\tan \Phi(x,y) = \frac{\sum_{n=1}^{N} [I_I(x,y) \sin(\alpha_i)]}{\sum_{n=1}^{N} [I_I(x,y) \cos(\alpha_i)]}
\]

In this case, the resulting intensity \(I_I\) is multiplied by cosine and sine signals of the same frequency, averaged over several cycles, and divided to give the tangent of the phase difference \(\Phi\).

4.5 Removal of phase ambiguities.

All the methods used for measuring the value of the phase angle \(\Phi\), described in section 4.4 will return a value in the range \((-\pi/2 \ldots +\pi/2)\), due to the nature of arctangent calculations.

In order for the phase angle \(\Phi\) to be modulo \(2\pi\), the denominator
and numerator, which are proportional to \( \cos \theta \) and \( \sin \theta \) respectively of the \( \arctan(\theta) \) expression, need to be monitored, and according to their signs the phases corrected to cover all four quadrants\(^{22} \). This method will yield \( 2\pi \) modulo values of \( \theta \) originating from all the methods used to determine the phase angle, apart from the Carré\(^{2} \) technique. Using this method, it is not sufficient to simply examine the signs of the numerators and denominators. It is necessary to monitor the signs of quantities proportional to \( \sin \theta \) and \( \cos \theta \). One such set of quantities is given by\(^{2} \):

\[
\begin{align*}
(l_2-l_1) &= [2\eta \gamma \sin \alpha] \sin \theta \\
(l_2+l_3)-(l_1+l_4) &= [2\eta \gamma \cos \alpha \sin^2 \alpha] \cos \theta
\end{align*}
\]

Once all the data are in modulo \( 2\pi \) format, the wavefront will have \( 2\pi \) discontinuities which have to be removed. This can be done with a routine which compares the phase of the signal from two adjacent detector elements (pixels). If the phase difference between these signals is greater than \( \pi \), then \( 2\pi \) is added or subtracted to give a difference smaller than \( \pi \). For this routine to correctly remove the \( 2\pi \) discontinuities, the phase difference between two adjacent pixels must not exceed \( \pi \). This limits the number of fringes measurable across a surface, and thus limits the ability of the system to accurately measure profiles of steep surfaces, as these cause closely spaced fringes.

A method that has been suggested to overcome this limitation is termed two-wavelength phase shifting interferometry. This involves obtaining two sets of interference fringes by the use of two different illuminating wavelengths. Full details of this technique and its benefits are described in chapter 7. An alternative method however is described below.

4.5.1 Extended unambiguous range interferometry.

This technique was suggested by Strand and Katzir\(^{24} \), and is based on the fact that the fringes generated with wide-bandwidth sources have a different appearance according to their order of interference. As can be seen from the general expression for the irradiance of a quasi-monochromatic two-beam interferogram:
where $I_1$ and $I_2$ are the irradiances of the reference and object beams respectively. $\psi$ is the degree of coherence between the two beams, and has a value of $(0 < \psi)$, for an optical path difference of the order of the coherent length of the source, to a value of $(\psi < 1)$, for very small path differences. It is a function of the optical path difference and describes a slowly varying envelope of the interference term, as seen in figure (4.2). Shifting the reference phase by $\pi/2$ radians produces an irradiance in phase quadrature to that defined by equation (14). After accounting for the differing intensities $I_1$ and $I_2$, simple algebraic steps yield the normalised quadrature interference terms$^{24}$:

$$I_1(x,y) = I_1(x,y) + I_2(x,y) + 2I_1I_2\psi(x,y;\phi) \cos \phi \quad \text{(14)}$$

$$\psi(\phi) \cos \phi(x,y);$$

$$\psi(\phi + \pi/2) \sin \phi(x,y);$$

These normalised interference terms can be used as coordinates to read the corresponding value of the OPD from a 2-D look-up table, which is essentially a digital representation of the spiral curve shown in fig (4.2).

Strand and Katzir$^{24}$ have demonstrated a system using this technique which provides absolute phase measurement over a range of several wavelengths on a point-by-point basis. A disadvantage of this method is that due to the full range being digitised as one scale, it is less sensitive than a system which digitises its range in units of $2\pi$.

4.6 Phase profile to surface contour.

Now that the wavefront phase profile is known, the surface shape $h(x,y)$ can be determined. For phases measured modulo $2\pi$, the relevant expression is:

$$h(x,y) = \Phi(x,y) \cdot \lambda / 2\pi(\cos \theta + \cos \theta') \quad \text{(15)}$$

where $\lambda$ is the wavelength of illumination, and $\theta$ and $\theta'$ are the angles of illumination and viewing with respect to the surface normal.
Figure 4.2 Graphic representation of the extended range algorithm for a wideband light source, the points with Cartesian coordinates given by \( m(\phi) \cos \phi, m(\phi+\pi/2) \sin \phi \) trace out a spiral as \( \phi \) goes from 0 in either a positive or negative direction. The radius of the spiral decreases in accordance with \( m \), the temporal coherence of the source. The spiral shown is for positive OPD; negative OPD would give an opposite handed spiral which intersects the positive spiral. The spiral is unambiguous over a range determined by the source temporal coherence length [28].
interferometers such as a Twyman-Green, which operate at near normal incidence, and are double-pass interferometers, equation (15) simplifies to:

\[ h(x,y) - \Phi(x,y) \frac{\lambda}{4\pi}. \]

It should be noted that the above expression is based on a simple interface model with air or vacuum above a homogeneous half space of material below, with a sharp dividing line at the junction. Lange discusses corrections to this expression which appear when one considers more complicated models.

4.7 Error analysis.

Interferometry is a potentially good method for making precision measurements. However, it suffers from errors caused by effects such as thermal expansion, air turbulence and mechanical vibrations. These may be reduced by placing the equipment on a vibration-isolated table, and enclosing the beam paths. Phase-shifting interferometry further reduces these effects by collecting and storing the data very rapidly. It is thus capable of making rapid and highly accurate measurements with sensitivities of the order of \((\lambda/100)\).

The limiting factors in its accuracy are thus mainly:

1) Errors caused by miscalibration of, and nonlinearities in, the phase shifter movements.

2) Noise and non-linearities in the detectors;

4.7.1 Phase shifter errors.

The calibration of the phase shifter for proper phase shifting between frames is very important if good phase measurements are to be obtained using phase shifting interferometry. The most common way to vary the phase difference between the two interfering beams in an interferometer is to mount the reference mirror on a piezoelectric transducer and drive
the mirror, by changing the voltage applied to it. Two fundamental problems associated with the motion of the piezo-electrically-driven reference mirror are firstly, the unknown sensitivity of the piezo-electric material, such as PZT, where the reference mirror may not move to the correct position to introduce an expected amount of phase shift (e.g. 90°) unless a careful calibration has been done first, and secondly the nonlinear effects of the PZT, which includes hysteresis and thermal drifts. Since the equations used for phase calculations are based on an assumption that the PZT is linear in its motion, this problem can introduce a considerable error if the nonlinearity is large. Both problems will generate a periodic phase error which has twice the spatial frequency of the interference fringes.

The problem of unknown sensitivity and non-linearities of the PZT can be overcome in several ways. The algorithm proposed by Carré for calculating the phase by a phase stepping method is independent of the amount of phase shift. It simply assumes that the phase shift between steps is linear. A PZT calibration-insensitive algorithm, similar to that of Carré, was suggested by Cheng and Wyant, but is, however, intended for the four-bucket technique. Using this method, and phase-shifting the phase angle from 60° to 160°, averaged phase errors of (∆/3600), were observed. Schwider et al proposed a method of reducing calibration and non-linearity errors by averaging two sets of data taken with a phase difference of π/2. This is termed the 3x3 method, and is discussed in more detail in chapter 5.

4.7.2 Detector noise and non-linearities.

The detector introduces intensity measurement errors caused by electronic noise such as shot noise and amplifier noise. Fortunately, because random phase errors exhibit a zero mean distribution, their effect on the phase measurement may be reduced by acquiring and averaging many independent data sets. The causes and effects of the above mentioned electronic noises are discussed in detail in chapter (3).

Non-linear responses from the detector array, caused by the odd and even rows of the array detector being read by different shift
registers, can cause phase errors. The differing response and sensitivities of individual pixels within a detector array, due to inhomogeneous doping, can also cause phase errors.

Creath\(^29\) considered a second-order detector nonlinearity, which was expressed mathematically as:

\[ I' = I + \epsilon I^2, \]

where \( I \) & \( I' \) are the incident and detected optical irradiances respectively, substituting the detected irradiances into the interference equation. Equation (2) thus becomes:

\[
\begin{align*}
1' &= I_0 \left[ 1 + \gamma \cos(\Phi + \alpha) \right] + \epsilon I_0^2 \left[ 1 + 2 \gamma \cos(\Phi + \alpha) + \gamma^2 \cos^2(\Phi + \alpha) \right], \quad (16) \\
1' &= I_0 \left[ 1 + \epsilon I_0 \right] + I_0 \left[ 1 + 2 \epsilon I_0 \right] \gamma \cos(\Phi + \alpha) + \epsilon I_0^2 \gamma^2 \cos^2(\Phi + \alpha), \quad (17) \\
1' &= I_0' + I_0' \gamma \cos(\Phi + \alpha) + \left( \epsilon/2 \right) \left[ I_0 \gamma \right]^2 \left[ 1 + \cos(2(\Phi + \alpha)) \right], \quad (18)
\end{align*}
\]

where \( \alpha \) is the phase shift between exposures. The non-linear term in equation (18) will cause phase errors. By substituting equation (18) into the expression for measuring the phase by the four step technique, equation (11), the \( 2\Phi \) dependent terms will cancel in the numerator and denominator. It is found that, when substituting equation (18) into the expression for measuring the phase by the three step method, equation (6), the non-linearities add and cause large phase errors. This would imply that in order to remove any effects of second order detector non-linearities, a minimum of four frames need to be sampled. Stetson and Brohinsky\(^30\) have discussed the effect of detector non-linearities on the various phase stepping techniques, and have published a table containing the degree of non-linearity causing phase errors for different numbers of phase steps per period, figure (4.3).

It can be seen from the table given in figure (4.3), that all even order non-linearities are removed by using the four step method. This finding is disputed by Kinnstaetter et al\(^31\), who derive an expression containing contributions from the third and fourth orders in the intensity as a forefactor of \( \sin 3\Phi \) and \( \cos 3\Phi \) in the numerator and denominator,
Table 1: Harmonics due to detection nonlinearities

<table>
<thead>
<tr>
<th>Number of Buckets</th>
<th>Harmonic Order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 4.3 This table shows the order of detection nonlinearity errors which affect the phase measurement for small number of phase steps, [33].
respectively. Using this expression, they confirm the phase measurement to be free from distortions if only second-order non-linearities are present when using a four-step technique. They find, however, that fourth-order terms do influence the phase to be measured, but their effect is proportional to the third harmonic of the phase term.

Creath\textsuperscript{29} has produced a computer simulation of phase errors due to phase-shifter miscalibration and non-linearity, as well as detector non-linearities, as applied to various measurement algorithms for comparison purposes. The simulation has shown that even with relatively large system errors (10\%), phase measurements well within $\lambda/20$ can be obtained. The Carré algorithm is the best to use for phase-shifting errors, and the four-step technique is best for eliminating effects due to second and third order detection non-linearities.

Bruning\textsuperscript{17} has stated that phase or wavefront errors will decrease as $(N)^{-1/2}$, where $N$ is the number of divisions per fringe. Hence it is clear that a four-step technique is better at removing phase errors, than a three-step technique, and a five-step system better than a four.

4.8 Conclusion.

In this chapter, an attempt has been made to review the material written in the various open literature journals on the subject of phase-shifting interferometry.

Phase shifting interferometry has been defined as a method of determining a wave-front phase profile by actively introducing a known phase shift into one of the arms of an interferometer, whilst sampling the intensity at specific discrete locations on the wave-front.

The various methods which have been employed to produce a known phase shift have been discussed.

The electronic techniques which have been developed to measure the phase of a detected signal have also been reviewed. These methods have been shown to be at a disadvantage when compared to analytical methods for
producing a phase profile of a two-dimensional wavefront.

Several of the most common analytical phase-detection methods, and their algorithms have been discussed and compared. It has been shown how the phase discontinuities are removed and the measured phase converted to absolute height values. The limitations of these techniques as applied to the measurements of profiles of steep gradients have been highlighted, and modifications to overcome these problems, such as the use of two wavelength interferometry and extended unambiguous interferometry, have been discussed.

Finally, the many factors which limit the accuracy and sensitivity of phase-shifting interferometry were reviewed, and several methods used to overcome or limit these effects have been discussed.
References.


CHAPTER FIVE.

PHASE-SHIFTING INTERFEROMETRY FOR THE MEASUREMENT OF VIDEO-HEAD WEAR.

5.1 Introduction.

In this chapter the development, testing and use of a measurement instrument based on the idea of phase-shifting interferometry, is described specifically for the measurement of video-recorder head wear.

Initially, other phase-shifting interferometer arrangements which are available commercially, and that have been used for measurements on video-heads, are discussed. The limitations of these systems, when profiling large areas of highly curved surfaces, such as video-heads, are highlighted.

The Twyman-Green optical arrangement is proposed as a method to overcome the problem of the sample's curvature. The implementation of this method as a practical phase-shifting interferometer system is described. The principle is first demonstrated for a steel ball-bearing with curvatures similar to that of a typical video-head, and later on actual heads. The configuration is then used to determine the different radii of curvature of the video-heads. These values are then used to determine the required optical system needed to image the video-head onto the surface of a CCD image sensor at the correct magnification. A brief review of the phase shifting algorithm which is used in the system to measure the surface profile is given.

A description of the electronic circuit designed and built for driving the PZT transducer high-voltage power supply is given, together with details of the interfacing to the controlling PC computer. The method of calibration and linearity testing of the phase-shifter motion is discussed.

Results showing the wear of typical video recorder heads due to the passage of tape across their faces are given. The results show the
initial interference fringe patterns on the unworn head and the subsequent processed phase profile. The same is shown for the worn head. The difference in the phase profile between the two cases, after correcting for mis-positioning errors, is a measure of the wear. The wear is then represented by a variety of means, including a contour map of the wear pattern across the face, a numerical value of the displaced volume and multi-perspective 3-D images. A computer routine which was written to calibrate the absolute dimensions of the head is also shown. Also given are results which show the workings of a smoothing routine which was developed to remove or reduce the effect of errors in the phase profile. The ultimate sensitivity and spatial resolution of the system are discussed.

The purpose of the work in this chapter is to find and develop a measurement technique which may be successfully and routinely used in the assessment of video-head wear. The results given in this chapter are primarily used to illustrate the capabilities of the system, and do not claim to be a comprehensive set of video-head wear measurements. A complete analysis of video-head wear under differing conditions would be a separate research project in itself. The few results obtained with the system however, are believed to be of great importance in the study of video-head wear, and the interaction of magnetic tape with heads. The developed system is an important tool in these investigations, and is currently in use at the research laboratories of a major international video-tape manufacturer, by whom this project was commissioned.

5.2 Wyko-Zygo phase-shifting interferometer.

The optical profilometer described by Wyant et al\textsuperscript{1} uses an interference microscope, modified to incorporate an optical phase-shifting device and an image sensor, which provides the capability of accurate surface-height measurements.

The system is based on a Mirau interferometer\textsuperscript{2}. This interferometer is an attachment to a long-working-distance microscope objective and operates in the reflection mode. Figure (5.1) shows a schematic diagram of the optical system. Light emitted from a tungsten lamp which is then reflected from the object under test, interferes with the
Figure 5.1 Phase-modulated Mirau interferometer.
light reflected from the internal reference surface. The resulting interference fringes may be viewed through the eyepiece of the microscope. The interference pattern is also recorded by the image sensor which may be read by the computer. Deviations within the fringe pattern show areas where the test surface differs from the internal reference. The Mirau interferometer objective is mounted on a piezoelectric transducer (PZT) which is moved under control of the computer. By shifting the PZT the computer automatically introduces a phase shift in the interferometer and simultaneously records the interference pattern produced. As the PZT is moved, the fringe pattern is displaced and the additional information obtained allows the computer to quantify the surface height deviations.

This provides the basis of a system available commercially from the Wyko Corporation, Arizona. The Wyko and other similar systems have been applied very successfully to the profiling of various surfaces and objects, including video-head profiles, as seen in figures (5.2) & (5.3). Figure (5.4a) is an example of a surface profile of a typical video-head taken by a Wyko type interferometer at the 3M research laboratory, Gorseinon. However, even when using a relatively low magnification objective (×10), the area which may be sensibly sampled is only ≈100μm². Thus one of the limitations of the Wyko interferometer system is its inability to measure relatively large area surface profiles (>1mm²) of highly curved objects (radius < 20mm). This is due to the fact that the Mirau objective contains a flat reference surface, so that the curved wavefront, which is produced by the reflection of a plane wavefront at the curved surface of the head, produces circular fringes on interference with the flat reference wavefront. These circular fringes increase in density in a direction away from the centre, as seen in figure (5.4b). This limits the sensitivity of the system, as the fringe visibility is considerably reduced. An added problem with a highly curved surface is the difficulty in recollimating the reflected light for imaging purposes without the using very large aperture lenses, which reduce the magnification of the imaging system.

5.3 Twyman-Green interferometer arrangement.

An alternative optical arrangement which may be employed to
Figure 5.2 Three-dimensional view of video-head in the region of the gap taken from [5].

Figure 5.3 Three-dimensional view of a video-head in the region of the gap. This image shows misaligned head parts, [4].
Figure 5.4a  Surface profile of small area of head taken by a WYKO type interferometer at the 3M Research Laboratories, Swansea [6].

Figure 5.4b  One of the set of interference fringes processed to produce the profile seen in figure 5.4a above. Note the increase in fringe density away from the centre, due to the curvature of the head.
overcome the problems described in the section above, is a Twyman-Green interferometer. This set-up was developed originally to test the quality of optical lenses and/or prisms\textsuperscript{7}, and is the basis of the optical arrangement used in the system described in this chapter for the surface profiling of highly curved surfaces, such as video-heads.

5.3.1 Optical arrangement.

It was initially decided to tackle the problems involved with the curvature of the sample by looking at the interference produced by light reflected off two 1 inch diameter steel ball-bearings. The optical arrangement employed is shown in figure (5.5). Light from a HeNe laser emitting at a wavelength of 633nm is expanded by use of a x10 beam expander/collimator. A convex lens with a focal length of 50mm is positioned so as to produce virtual focal points at the centre of curvature of the two ball-bearings. This ensures that the curvature of the incident wavefront matches that of the ball-bearings, and that the reflected light is returned along its original path so as to be collected by a 32mm focal-length objective lens and imaged onto the surface of a CCD image sensor. A polarising beam splitter is used to split the incident beam into two orthogonally polarised components, which are then reflected off the surface of the ball-bearings. A λ/4 retardation plate is placed in each arm of the interferometer so as to rotate the plane of polarisation of the beams by 180° on its double pass. This ensures that all the initial beam is incident upon the image sensor, and that no light is returned back towards the laser, to produce the effect of cavity detuning. The polarising plate on the output acts as an analyser by filtering out any unwanted reflections and increasing the fringe contrast.

The interference fringes produced by the reflected wavefronts off the surface of the ball-bearings are shown in figure (5.6a), where the field of view is 2mm. The circular fringes shown are due to the fact that the surfaces are not equidistant with respect to the focusing lenses. They are therefore illuminated differently with the consequence that the reflected beams emerge with unlike cone angles. This results in wavefronts with different curvatures, and circular fringes on interference. As the optical path-length difference is reduced the fringes widen, as seen in
Figure 5.5 The optical arrangement employed to observe the interference of light reflected off two 1 inch diameter steel ball-bearings.
Figure 5.6 The fringes produced by the interference of light reflected off the surface of the two steel ball-bearings. a) shows narrow fringes due to the unequal optical path lengths in the arms of the interferometer, b) shows the same fringes but with the arm lengths more closely matched.
figure (5.6b). Finally on matching the path lengths and the curvatures of the reflected beams, a single fringe is obtained across the field of view. This pattern, shown in figure (5.7a), represents the surface imperfections of the two ball-bearings together with any aberrations due to the optics used. Figure (5.7b) shows the same information but with the path-length of one of the arms of the interferometer changed by \( \pi \) radians. If figure (5.7a) represents the \( \cos \) of the interference then figure (5.7b) is the negative or \(-\cos\). It is clear that each pattern is the negative of the other.

5.3.2 Matching the curvature of a video-head to a curved reference.

Once the curvature of the wavefronts reflected off two 1 inch diameter steel ball-bearings had been successfully matched, it was decided to replace one of the ball-bearings with a video-head. The resultant interference fringes are shown in figures (5.8a) & (5.8b). The fringes shown in figure (5.8b) are the broader of the two; this is due to the fact that the curvature of the reflected wavefronts in this case are closely matched than those in (5.8a). By further matching the curvatures it is possible to obtain a single fringe across the field of view in the \( x \)-direction, i.e along the length of video-head, but not in the \( y \)-direction. This is due to the head having different radii of curvature in the two perpendicular dimensions.

In order to remove the problem of different radii of curvature of the video-head it is necessary to have a reference wavefront with a curvature which matches the wavefront reflected off the test video-head in both the \( x \) and \( y \) directions. The ideal reference surface would be another video-head with similar surface dimensions. It is shown in figure (5.8c) that excellent interference fringes are obtained by combining the wavefronts of laser light reflected off two different video-heads. The fringes shown are due to the slight misalignment of the two heads. By slight adjustment this effect may be removed, and as seen in figure (5.8d), a single fringe may be obtained across the whole face of the illuminated section of the video-head.
The arm lengths of the interferometer matched to produce a single fringe across the field of view. The pattern shown in a) is mainly due to the combined imperfections of the balls' surface, b) shows the same information but with one of the balls moved in the focal plane of the lens by roughly $\lambda/4$ to produce a negative image of that shown in figure 5.7a.
Figure 5.8 a) shows the interference fringes produced by replacing one of the ball-bearings shown in figure 5.5 with a video-head, b) shows the same but with the head moved towards the lens to match the curvatures of the reflected wavefronts more carefully.
Figure 5.8 c) shows fringes produced by interfering light reflected off the surface of two video-heads, d) shows the same but with one head slightly tilted to reduce the number of fringes across the surface.
In this experiment it was found that only about two-thirds of the surface of the head was being imaged. Either the light reflected off the extreme edges, which include the unworn reference strips, was not being collected by the imaging optics, or only two thirds of the head surface was being illuminated. In order to correctly illuminate and image the whole surface of the video-head together, it is important for such parameters as the curvature and physical dimensions of the head to be known.

5.3.3 Measurement of curvature of video-heads.

To accurately measure the curvature of the video-head, a Twyman-Green interferometer was constructed. The arrangement is shown in figure(5.9). A collimated beam of laser light is split by a cube beam splitter with one arm being reflected off a reference plane mirror, and the other component focused by a objective lens onto the surface of the video-head. On reflection this ray only emerges from the lens with a plane wavefront if the centre of curvature of the head coincides with the focal plane of the incident ray. By adjusting the distance between the objective lens and the surface of the video-head it is possible to obtain a plane wavefront which had been reflected off the curved surface of the head, and which, on recombination with the reference plane wavefront produce a single fringe across the face of the video-head, as seen in figure (5.10). The curvature of the video-head in the longer dimension was measured to be 12mm. This was found by subtracting the distance from the lens to the surface of the head from the known focal length of the lens (50mm).

In making this measurement it was assumed that the video-head curvature was constant across the whole face of the head. It is shown later that this assumption was invalid and indeed the 12mm value for the curvature of the head was only true for the part of the head's surface being illuminated, and other parts of the head had curvatures that slightly departed from the 12mm curvature value.

The curvature in the other direction of the head was found by using the same method with the lens to surface distance adjusted to give a single fringe in the y direction of the head. The curvature was measured to be 6mm. In this measurement the objective lens was replaced by a shorter
Figure 5.9 A Twyman-Green type interferometer
constructed to measure the curvature
of a video-head in the x & y directions.
Figure 5.10 Interference fringes from a video-head with the curvature effectively removed in the x direction.
focal length lens (16mm) for greater accuracy.

To confirm these curvature results, the approximation of the rule of intersecting chords can be used:

\[ h = \frac{l^2}{2r} \]

where \( h \) is the maximum height of the curved surface with respect to the chord connecting the edges of the curved head, figure (5.11), \( r \) is the radius of curvature of the head, and \( l \) is the length of half the chord connecting the two edges of the head.

The chord \( l \) represents half the width of the video-head (100 \( \mu m \)). The height \( h \) resulting from the curvature of 12mm across 200 \( \mu m \) gives 416 nm. see figure (5.12). The height \( h' \) due to the measured radius of 6mm is found to be 833 nm. Hence the difference in the two heights across the width of the head is 417 nm. This represents the curvature in the \( y \) direction of the head when the curvature in the \( x \) direction is being matched with a 12mm curved wavefront to give a single fringe. The height 417 nm corresponds to about 3 HeNe beam interference fringes across the width of the video-head. This is confirmed as shown in figure (5.10).

5.3.4 Imaging of the video-head.

In order to successfully image the surface of the video-head, an optical arrangement is required which will fully illuminate the whole surface of the video-head, re-collimate all the reflected light and ensure the surface of the head is in focus with the correct magnification.

The arrangement used is one employing a doublet system as shown in figure (5.13). An expanded (x10) parallel beam of light from a HeNe laser is passed through a cube beam splitter and focused by a 35mm f-2.4 SLR multi-element camera lens onto the centre of curvature of the video-head. This ensures that the laser beam meets the surface of the head at a right angle, so that all the reflected light is collected by the same objective lens and passed back towards the beam splitter as a collimated beam. The diameter of the incident beam is 10mm, and the head is positioned
Figure 5.11 Illustration of the Chord Rule where
\[ h(2r-h) = l^2, \text{ if } r \gg h : h = l^2/2r. \]

\[
\text{Chord rule, } \quad h(2r-h) = l^2 \\
\text{if } r \gg h : h = l^2/2r.
\]

\[ h=338 \text{ nm} \quad h=833 \text{ nm} \]

\[ 12 \text{ mm radius.} \]

Wavelength to match curvature of head in longer dimension.

Figure 5.12 Schematic diagram of the interference shown in Figure 5.10 where the wavefront of light matches the curvature of the head in the long direction.
Figure 5.13 The optical arrangement of the doublet system used to illuminate and image the surface of the video-head.
23mm from the lens. This distance is the focal length of the lens (35mm) minus the radius of curvature of the head (12mm). This then gives a beam diameter of 4.5mm at the plane of the head, which is sufficient to illuminate the 3mm long video-head.

To ensure that the surface of the video-head is in focus on the surface of the CCD detector, the correct distance between the objective camera lens and the 75mm flat microscope lens needs to be calculated. As the surface of the head is 23mm (u) away from the 35mm focal length lens (f), using:

\[
\frac{1}{f} = \frac{1}{u} + \frac{1}{v}
\]

gives \( v = -67 \text{mm} \). The 75mm microscope objective is positioned 160mm away from the CCD detector, which is the standard microscope tube length, and the distance for which the aberrations are corrected. Again using the standard equation above and setting (v) to be 160mm and (f) to be 75mm, it can be shown that (u) is 141mm. The 35mm camera lens produces a virtual image 67mm away, so the separation of the two lenses needs to be (141mm - 67mm = 74mm), ignoring for the moment the thickness of the multi-element camera lens. This separation is quite adequate to position a 20mm cube beam-splitter, and allow small movement of the microscope objective for fine focusing.

Finally the magnification of the optical system can be calculated by considering the ratio v/u for the two lenses. The camera lens gives a magnification of \( \frac{67}{23} = \times2.9 \) and the microscope objective gives a magnification of \( \frac{160}{141} = \times1.1 \). The total magnification of the system is thus \( 2.9 \times 1.1 = \times3.3 \). With the video-head being roughly 2.5mm long, this will give an image size of about 8mm. As the dimensions of the CCD chip are 10mmx10mm this appears to be a suitable magnification.

The 75mm lens is a flat microscope objective lens, hence the front surface of the lens can be considered to be its principal plane and all measurements can be made from this plane. The multi-element camera lens however has two principle planes, a front and back principle plane. The position of these planes can be easily found by passing a parallel beam of light through the lens in both directions and recording the positions of
the focal points. These points will occur 35mm away from the front and back principle planes of the complex lens, and all measurements concerning the lens are made from these planes. On the particular lens used in this system, the separation between the two principle planes was found to be 20mm. This distance can be ignored for any calculations of divergence or convergence of rays going through a multi-element lens system.

Figures (5.14a) & (5.14b) show the illuminated surface of a typical video-head adjacent to a 0.5mm ruled grating showing its dimensions to be roughly 2.5mm long and 200μm wide. Two images are needed with different orientations of the grating, as the video system has an aspect ratio of 4:3 in the x and y dimensions across the screen.

5.3.5 Matching the curvature of the video-head to a plane reference.

Although it was shown that good interference fringes were obtainable by the use of the reflection off a second video-head as a reference wavefront, section 5.3.2, it soon became clear that this particular configuration introduced problems of alignment, in particular those concerning the removal and replacement of the head, and their corrections. It is a much simpler task to measure and remove the effects of tilt and defocus of a spherical wavefront on its interference with a plane wavefront to the case of the interference of two slightly different spherical wavefronts.

It was decided that the optical arrangement, containing a plane reference mirror, used to measure the curvature of the heads, (section 5.3.3), would be employed to measure the profile of the video-head surface.

Although the surface of the video-head was initially thought to form part of a perfect sphere, it has been shown that the two dimensions of the head have different radii of curvature. It is further shown on viewing the interference pattern across the entire surface of the head, using the correct imaging optics, that there is an area in the centre of the head with a larger radius of curvature. This 'flattened' area is thought to be
Figure 5.14 Shows the illuminated surface of the video-head adjacent to a 0.5mm ruled grating to measure its dimensions, a) vertically, and b) horizontally.
produced as part of the manufacturing process, with the central area of the head, which contains the gap surrounded by the glass beads, ground down to a very fine surface finish.

The figures (5.15a) & (5.15b) show the interference fringes on a video-head produced by matching the curvature of roughly one half of the head with the spherical wavefront produced by the camera lens. This is done by slightly tilting the head to one side and then the other. Figure (5.15c) shows the same head set at an optimum position to give good broad interference fringes across the whole face of the video-head. These interference fringes demonstrate the departure of the surface from a perfect spherical shape. The fringes however, are perfectly adequate to work with.

### 5.4 Experimental arrangement.

The experimental arrangement is shown in figure (5.16). The optical configuration is as described in section (5.3.4), with the reference mirror mounted on a PZT transducer to produce the appropriate phase shifts. The transducer is driven by a high voltage power supply controlled by an Amstrad PC1512 microcomputer. The computer simultaneously controls a video frame-store which captures information imaged onto a CCD video-camera, and stores this information for processing, and the display of the surface profile results.

#### 5.4.1 Phase-shifting method.

Chapter 4 deals in detail with the theory and the many possible algorithms used in phase-shifting interferometry. The specific method used in the video-head wear measurement system given in this chapter is termed the 'Averaging three-and-three technique' and a description of the method is given below.

The accuracy of phase-shifting interferometry results rely on the correct calibration of the phase-shifting transducer. Miscalibration produces a sinusoidal error with a spatial frequency twice that of the

158
Figure 5.15 The interference fringes on a head produced by matching the curvature of about one half of the head's surface to that of the spherical wavefront produced by the objective lens. This is done by slightly tilting the head to one side a), and then b) to the other side, c) shows the same head set at an optimum position to give the broadest possible fringes.
Figure 5.16 The final experimental arrangement.
fringe pattern\textsuperscript{10}. Schweider et al\textsuperscript{9} proposed a technique of reducing errors by averaging two phase measurements taken with a relative phase-shift of 90° between the two measurements. By using successive π/2 phase shifts, four sets of intensity readings, A, B, C and D, may be obtained. Using the 'three-bucket method', A, B and C can generate one set of phase data while B, C and D can generate another set. The miscalibration error can thus be simply removed by averaging the two sets of data. This can be expressed mathematically as:

\[ \varphi(x, y) = \frac{1}{2} \left( \tan^{-1} \left( \frac{I_3(x, y) - I_2(x, y)}{I_1(x, y) - I_2(x, y)} \right) + \tan^{-1} \left( \frac{I_4(x, y) - I_3(x, y)}{I_2(x, y) - I_3(x, y)} \right) \right) \]

It was decided to employ this technique for its ability to average out errors, and for the relative simplicity of its calculation.

\textbf{5.4.2 PZT driving electronics.}

The PZT transducers are driven by a high voltage power supply unit (TecOptics FPZ-3). This unit contains 3 high voltage ramp generators which can be controlled externally via a BNC socket. (0 - 10 V input gives 0 - 1000 V output from the ramp generators).

In order to control the motion of the reference mirror using a computer, it was necessary to design and construct an electronic circuit which would convert a digital number output by the computer, via an interface card, into an analogue voltage which would be used to drive the PZT power supply.

The circuit diagram of the electronics is shown in figure (5.17). The arrangement was designed to be compatible with the Eltime interface card (described in chapter 1), and which is used by the computer to control the video frame-store. The output of the interface card consists of 8 data lines (D0-D7), 3 address lines (A0, A7 & A8) and a clock pulse line \( \varphi \). The address lines contain the offsets of the numbers sent by the computer, to control the various functions of the frame-store, to that of the base address number set by a DIL switch on the interface card, (HEX 300). Out of the possible 16 offset numbers, only 11 are actually used by the
Figure 5.17 The digital-to-analogue electronic converter circuit designed and built to computer control the piezo-electric transducer high voltage power supply.
frame-store registers. The offset +5 is one of the unused numbers, and is therefore chosen to control the PZT driver unit. This, together with the use of a Y ribbon cable connector, enables the computer to control two different sets of apparatus independently using only one interface card.

The address lines from the interface card are connected to pins (1-3) of the 74 138 3-line to 8-line decoder. On arrival of the offset (5) to these pins, together with a clock pulse output by the computer, the 74 138 activates pin 10. This signal is then used to enable the ZN-428 8-bit digital-to-analogue converter, which converts the digital number (0 - 255) held on the data lines (D0 - D7) into an analogue voltage at pin 5. The ZN 428 has a built-in reference voltage of 2.5 volts, with an output range of 0 - 2.55V. This analogue voltage output is then fed into a 741 Op-Amp operating in the non-inverting mode to give a gain of (×4) to convert this to (0 - 10 V). The Op-Amp is offset-nulled to give a 0 V output to correspond to a 0 V input. The gain of the amplifier is adjusted to give an output of 10 V when all the data lines are high (255) by adjusting the variable feedback resistor (R2). The gain of the Op-Amp is given by the expression below:

\[ \text{Gain (Av)} = 1 + \frac{R2}{R1} \]

The analogue voltage is then used to drive the PZT transducer high voltage supply, which drives the reference mirror forward to give the appropriate phase shift.

5.4.3 Linearity of the PZT power-supply system.

In order to test the linearity of the high voltage supply output with respect to the digital number generated by the computer, a simple computer program was written to output digital numbers representing (0 - 10 V) to the interface card. The corresponding analogue voltages output by the D/A converter circuit, together with the amplified high voltage signal output by the PZT driver supply, were monitored by a Fluke 8050A digital multimeter, and plotted in the table below:
It can be seen from the table above that, apart from the first 100 volts step, the PZT high voltage supply system has an output error of less than 3% to the appropriate digital number input. The D/A converter chip has a stated conversion accuracy of \( \pm \frac{1}{2} \) L.S.B., which corresponds to a 1/2% error on conversion\(^{12}\). The other effects are due to cross talk in the conversion circuit, and non-linearities in the operational amplifiers.

5.4.4 Calibration and linearity of PZT.

To calibrate the motion of the PZT mounted mirror as a function of applied voltage and to check the linearity of its motion, a Michelson interferometer was built with the reference mirror mounted on the PZT transducer. The two arms of the interferometer were made parallel to give a single fringe. The output signal of the interferometer was focused onto the surface of a silicon photo-diode. The measured intensity then fluctuated as the path length difference of the arms of the interferometer was changed by driving the PZT transducer forwards and backwards.

Figure (5.18) shows the photo-diode signal versus the voltage applied to the PZT transducer. The voltage was ramped from 0-1000V in the forward direction, then reversed and driven back down to 0 volts. The graph shows that at the start of the scan in each direction, the PZT mounted mirror suffers a slight non-linearity in its travel by view of the broader
This shows the output signal from a Michelson type interferometer, as the reference mirror is translated by driving the PZT transducer forwards and backwards.
spaced oscillations in these positions. This effect appears however to recede after the first 200 volts or so. It was thus decided to operate the PZT at its mid-operating position (500V), and always scan in the same direction to avoid any hysteresis effects\textsuperscript{13}.

From the graph of PZT motion versus applied voltage it was possible to measure the required voltage needed to produce a $\pi/2$ phase change in the interferometer. This was found to be approximately 55V. In order to confirm this figure, and to check the spread of the introduced phase shift across a finite area of the mirror, the output of the interferometer was imaged onto the surface of a CCD detector. The reference mirror was stepped five times so as to give a $\pi/2$ phase change on each step, and a frame of intensity data was stored in the framestore after each step. The exact phase shift ($\alpha$) at each of the detector points can be calculated using the expression shown below\textsuperscript{14}

$$
\alpha(x,y) = \cos^{-1} \left[ \frac{1/2 (l_5(x,y)-l_1(x,y))}{(l_4(x,y)-l_2(x,y))} \right]
$$

The results of this experiment are shown in figures (5.19a) - (5.19c), which show histograms of the distribution of phase shifts across roughly 10000 detector elements around the desired phase shift for a particular applied voltage. Three voltages were used to produce the phase-shift: 52V, 55V and 58V. It is clear that 55V is the correct voltage needed to produce a $90^\circ$ phase-shift. It is also clear that the narrow distribution of the histogram indicates a well calibrated phase shifter.

A more visual method of determining the phase shift is given in figures (5.20a) - (5.20d), which show interference fringes at various phase shifts. These shifts were obtained by moving the reference mirror of the interferometer by multiples of $90^\circ$, by applying 55 volts to the PZT transducer. If the phase-shifter is correctly calibrated, the fringe pattern after the fourth step ($360^\circ$) should exactly match the first fringe pattern. This is confirmed by comparing figures (5.21a) & (5.21d).

Care needs to be taken to allow the reference mirror time to settle after stepping and before taking an intensity measurement. This problem is illustrated in figure (5.21), which shows an oscilloscope trace of the photo-diode output on stepping the mirror by quickly applying 55V to
Distribution of phase shift by the application of 52 Volts

Figure 5.19a  Histogram of the distribution of phase shifts across roughly 10000 detector elements around the desired phase shift of 82.5°.
Distribution of phase shift by the application of 55 volts

Figure 5.19b Histogram of the distribution of phase shifts across roughly 10000 detector elements around the desired phase shift of 90°.
Distribution of phase shift by the application of 58 volts

Figure 5.10c Histogram of the distribution of phase shifts across roughly 10000 detector elements around the desired phase shift of 97.5°.
Figure 5.20 A more visual method of determining the phase shift. a), b), c) & d) are a set of interference fringes separated from each other by multiples of 90°, with a) 0°, b) 90°, c) 270°, & d) 360°.
Figure 5.21 Oscilloscope trace of photo-diode monitoring output of interferometer on stepping the reference mirror. This illustrates the settling time of the reference mirror.
the PZT transducer. The time scale is 50ms per division. It is clear that there needs to be a delay of approximately 150ms between stepping the mirror and taking an intensity reading. This allows the ringing to settle to acceptable levels.

5.5 Data transfer and processing.

All the intensity information relating to the interference fringes is captured by an EEV P4310 Charge-Coupled Device video-camera. The information is then transferred to an Eltime Image III video frame-store. An Amstrad PC-1512 micro-computer is used to access the individual pixels of information, perform the phase calculations, and display the results. The computer is also used for control and synchronisation of the various hardware elements of the measurement system. The various hardware components have all been described earlier in this thesis (Chapter 1), so will not be reviewed again.

5.5.1 Phase calculations.

Due to the nature of arc tangent calculations, the result of the phase calculation is given modulo $\pi$. This limits the unambiguous measureable optical path length difference between two adjacent pixels to less than $\lambda/8$. It is possible to double this value to $\lambda/4$ by converting the phase values modulo $\pi$ into modulo $2\pi$. To achieve this conversion, the signs of quantities proportional to $\sin \Phi$ and $\cos \Phi$ must be examined. The table given by Creath and shown below, explains how the phase is determined by examining the signs of these quantities and using absolute values in the numerator and denominator to yield a modulo $2\pi$ calculation.
Table (5.22)
Determination of the phase modulo $2\pi$.

<table>
<thead>
<tr>
<th>Numerator</th>
<th>Denominator</th>
<th>Adjusted phase</th>
<th>Range of phase values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(\sin \phi)$</td>
<td>$(\cos \phi)$</td>
<td>$\phi$</td>
<td>$0 - \pi/2$</td>
</tr>
<tr>
<td>positive</td>
<td>positive</td>
<td>$\phi$</td>
<td>$0 - \pi/2$</td>
</tr>
<tr>
<td>positive</td>
<td>negative</td>
<td>$\pi - \phi$</td>
<td>$\pi/2 - \pi$</td>
</tr>
<tr>
<td>negative</td>
<td>negative</td>
<td>$\pi + \phi$</td>
<td>$\pi - 3\pi/2$</td>
</tr>
<tr>
<td>negative</td>
<td>negative</td>
<td>$2\pi - \phi$</td>
<td>$3\pi/2 - 2\pi$</td>
</tr>
<tr>
<td>0</td>
<td>anything</td>
<td>$\pi$</td>
<td>$\pi$</td>
</tr>
<tr>
<td>positive</td>
<td>0</td>
<td>$\pi/2$</td>
<td>$\pi/2$</td>
</tr>
<tr>
<td>negative</td>
<td>0</td>
<td>$3\pi/2$</td>
<td>$3\pi/2$</td>
</tr>
</tbody>
</table>

5.5.2 Removal of $2\pi$ discontinuities.

As explained in the section above, the phase values are converted to modulo $2\pi$ leaving phase ambiguities which need to be resolved. The removal of these $2\pi$ phase discontinuities is accomplished by the comparison of adjacent pixels. When the phase difference between two adjacent pixels is greater than $\pi$ radian, a phase value of $2\pi$ is added or subtracted to make the difference less than $\pi$. For the reliable removal of the phase discontinuities, the phase difference between two adjacent pixels must not exceed $\pi$ (i.e. $\lambda/2$ in optical path). For this scheme to work, the CCD element spacing must be less than half the fringe spacing.

The result of this operation is illustrated in figures (5.23a) & (5.23b), which show the removal of the $2\pi$ discontinuities associated with the phase profile of a tilted mirror in the one arm of a Michelson interferometer. The phase values from a double pass type interferometer, once calculated, are simply converted into absolute height values using the expression below:

$$h(x,y) = \frac{\lambda}{4\pi} \times \phi(x,y)$$

where $\lambda$ is the wavelength of the laser light, and $\phi$ is the phase value.
Figure 5.23  a) The phase profile of a tilted plane mirror showing $2\pi$ steps. b) shows these steps removed.
Replacing the video-head.

The basis of this wear measurement system is the characterisation of the surface profile of a video-head in a before-and-after wear situation, and then the subtraction of the two results to yield a measure of the difference, or the wear.

An important requisite of this system is the ability to replace the video-head, in exactly the same position that it occupied before its removal, or to within an accuracy greater than the required measurement accuracy. This of course is unrealistic since the desired accuracy of the measurement is of the order of nanometres. It is found that in terms of lateral repositioning, however, it is possible with some practice to replace the head so that a particular point on the head occupies the same, or at worst an adjacent, pixel on the CCD detector. This task is helped by the writing of a computer routine which alternately displays the stored image of the unworn-head in the original position, and the captured image of the replaced head, on the video-monitor. The replaced head then undergoes fine adjustment until the flicker of the images due to misalignment disappear. The magnification of the optical system is approximately x3.5, with the separation of the individual pixels of the CCD element being 20μm. This requires a positioning accuracy of about 6μm which is clearly not impossible.

The system measures the profile of a head by determining the relative height of its surface to an accuracy of nanometres. It is in this dimension that replacing the head accurately becomes a problem. This repositioning error can be divided into two effects, i) video-head tilt, and ii) defocus. These effects and their effective removal are described below:

5.5.4.1 Tilt compensation.

Video-head tilt occurs when one side of the head is closer or further away from the lens than the other on repositioning of the head after removal for wear. This can occur in both the x or length dimension or in the y or width dimensions of the video-head. Although the surface of the head describes part of a circle, due to the very small angles involved, the
Tilt can be considered to be linear across the length and width of the head. The extreme edges of the video-head do not undergo any wear, as the video-tape does not come into contact with these parts. It is thus possible to use the relative heights of these unworn edges in the before-and-after situations to account for and compensate for any tilt in the head before subtraction.

The tilt is measured by taking an average 'grey-level' or height value in a 5x5 pixel square in three of the head's corners, and calculating the gradient between these corners to find the initial tilt of the head. On replacing the video-head, this calculation is repeated and any change in the gradient is subtracted from the surface height data to remove the effects of tilt.

Focus compensation.

Defocusing of the video-head in this context means the positioning of the head in a vertical plane nearer to or further away, from the objective lens which produces the curved wavefront to roughly match the curvature of the head. The effect of this defocusing is to produce the surface details superimposed on a convex or concave curve, depending on which side of focus the head is on. In order to remove the effect of the different focus position of the replaced head, to that of its original position, from the subtracted surface profile measurement, it is necessary to calculate the induced curvature, and thus remove it from the surface profile results.

The curvature superimposed on the wear measurement can be found by taking the coordinates of three points (A, B & C) on the unworn edges of the video-head, two from one edge and the third from the other edge. From the position of these three points it is possible to describe the circle which passes through them, figure (5.24). This is done by finding the gradients and the mid-points of two of the lines connecting the three points and thus obtaining the right angled bisector of these lines. The coordinates of the point where these two lines cross defines the centre of the circle. In practice, the x coordinate of the 3 points is simply the position of the point along the length of the head, while the y coordinate is the averaged 'grey-level' value or surface height of the head across the.
Unworn edges.

Centre of circle passing through three points.

Worn part of head.

Cursors moved to suitable position on unworn part of head.

Wear pattern on head.

Figure 5.24 Positioning of cursors on the unworn edges of a video-head to define the circle that is required to remove the defocus curvature superimposed on the wear pattern.
width of the head at this point. The values of these coordinates are found by moving three cursors, whose length matches the width of the video-head along the length of the head to suitable positions on the unworn edges, and taking the average value of the grey-levels, or relative surface heights, along these lines.

5.6 Results.

In this section results showing the wear of video recorder heads is given. The results show the initial interference fringe patterns on the unworn head, and the subsequent processed phase profile. The same is shown for the worn head. The difference in the phase profile of the two cases is a measure of the wear. The wear is then represented by a variety of means, including a contour map graph and multi-perspective 3-D images. The wear is also represented numerically as a displaced volume. The absolute dimension calibration routine is described in this section.

The repositioned head tilt error subtraction routine is demonstrated together with the defocus curvature removal algorithm. Also, results are given which show the workings of a smoothing routine which was developed to remove or reduce the effect of errors in the phase profile.

All the results shown in this section are thermal copies of a TV monitor image. They are recorded by a Mitsubishi video copier.

5.6.1 Video recording head phase-profile results.

Figures (5.25 a,b,c,d) show the set of four interference fringes on the surface of an unworn head mounted on a drum from a domestic VHS format video-recorder. The relative phases of the interference fringes are mutually separated by 90° to satisfy the requirements of the algorithm used to calculate the phase profile. The relative phase difference value is given in the bottom left hand corner of each image. It can be seen that there exists a 180° phase difference between figures (5.25a) and (5.25c) and figures (5.25b) and (5.25d). Hence they represent each others' negative.
Figure 5.25 a), b), c) & d) show the set of four phase-shifted interference fringes on the surface of an unworn VHS format video-recorder head.
In order to calculate the effective phase profile of the unworn head, a computer routine was written which enabled an operator to move two crosses to mark out the area of interest on the surface of the head. These cursors are moved to the top left and bottom right corner of the area to be processed. This procedure eliminates the problems associated with phase errors which arise at the very extreme edges of the head, where the fringes are very closely spaced due to the severe change of the head curvature in these areas. The coordinates of these cursors are saved in the computer memory, and are used to process the exact same area on the head after it has been removed, worn and replaced. Figure (5.26a) shows one of the set of interference fringes on the head with the two crosses defining the area to be processed. The algorithm used to calculate the phase profile requires the intensity values from specific locations, from each of the four set of interference fringes, across the face of the head. The computer obtains this information by accessing the data within the area specified by the cursors, and stored in the four frames of the video frame store, in a sequential fashion. On calculation of the phase value at a specific location, or pixel, it is then written into one of the stores, and the next pixel accessed. On completion of the area to be processed, the complete phase profile is displayed as seen in figure (5.26b). The phase steps are known as \(2\pi\) discontinuities and arise due to the nature of the arc tangent calculation in the phase calculation.

The \(2\pi\) steps in the phase profile can be removed as explained in section (5.5.2). The result is a continuous phase profile, with 64 grey levels. This is shown in figure (5.26c). The highest point of the surface is represented by the grey level 63, or white, and the lowest point represented by grey level 0, or black. The profile can be visualised more easily by reducing the number of grey levels representing the image, to give a contour map format. This is given by figure (5.26d), which shows 16 grey levels. Although, this effect is clear on the TV monitor, the thermal copier used to produce hard copies of the results only displays images with 16 grey levels. Hence the effect is lost although its effect is clear when viewing it in 3 dimensions, as shown next.

5.6.2 Representing results in three-dimensions.

To display the phase profile of the video-head and more
Figure 5.26 The various stages of processing the fringes shown in Figure 5.25. a) is one of the set of fringes, b) phase profile showing $2\pi$ steps, c) continuous phase profile of head in 64 grey levels, d) as in c) but in 16 grey levels for contour effect.
importantly to display the change in the phase profile, or the wear of a video-head, a computer routine was written to convert the intensity values, representing the relative heights of the head's surface into a 3 dimensional representation. This is done by plotting white pixels at points corresponding to positions on the head surface, but displaced in the y dimension by an amount proportional to the grey level, or surface height, onto a black background. For the correct 3 dimensional effect it is important not to plot the points which would be expected to be obscured by the shape of the profile. This is done by storing in an array the height values of the first plotted line, and thus only subsequently plotting points that have a greater y displacement than the point stored in the array for a particular x location. The value in the array is updated, each time a higher point is plotted. In order to give a perspective feel to the 3 dimensional plot, each line is plotted slightly displaced in the x dimension from the previous line.

Figure (5.27a) shows a 3 dimensional diagram of the surface profile of the unworn video-head, taken from the data in figure (5.26c). Figure (5.27b) shows a 3 dimensional plot of the contour map data taken from figure (5.26d).

A consequence of plotting the results in 3 dimensions is that if a feature of the phase profile occurs at a lower point, at the far side of the 3-D plot, it will not be seen. To avoid this, the computer routine is able to access the profile data from either of the two sides, and with the appropriate lateral shift, give four different perspectives of the profile. Figures (5.28 a,b,c) give the 3 other views that compliment the perspective given in figure (5.27a).

5.6.3 3 x 3 filtering.

The $2\pi$ steps are removed as explained in section (5.5.2). The correct removal of these steps relies on the fact that two adjacent pixels do not have a phase difference greater than $\pi$ ($\lambda/2$ in optical path difference). If this condition is violated then phase errors will occur. The continuous phase profile is calculated in two dimensions by removing the $2\pi$ steps and finding the continuous phase values of a line of pixels at the top edge of the defined area, and using these values as a reference.
Figure 5.27  a) A 3-D plot of the phase profile from the unworn video-head shown in figure 5.26c, b) is a 3-D plot of the 16 grey level image in figure 5.26d.
Figure 5.28 a), b) & c) give the three other views that complement the perspective given in figure 5.27a.
The rest of the area is then scanned with a series of vertical lines, to cover the area of interest. This method is adopted, rather than one continuous sweep, to reduce the effect of phase errors which could otherwise corrupt a large part of the phase profile. These errors would start at the initial error position through to the end of the scan since the error is carried through. In the adopted method, if a pixel is corrupted, then a line of corrupted data will cross the profile, from its source down to the bottom of the image, only corrupting a small amount of the image. The correct phase would then be resumed on the next adjacent vertical line. Figure (5.29a) shows the phase profile of a video-head, where the processing has been conducted near the extreme edges of the head and where the fringes are closely spaced. On removal of the $2\pi$ steps, a phase ambiguity near the top edge of the head causes a line of corrupted data to cross the profile, see figure (5.29b). Also shown are short corrupt lines of data at the bottom edge of the central region of the profile.

The effect of this corrupted data can be removed or reduced by a method of data filtering termed $3 \times 3$ median filtering\textsuperscript{16}. This technique works by sampling a $3 \times 3$ array of pixels, arranging their heights, represented by their grey levels, in ascending order and replott ing the median value in the central position of the square. The computer then samples the adjacent $3 \times 3$ array etc. A median filter is used instead of an averaging method, as a large phase error in one or more of the sampled data points, would produce an average value greatly removed from the median. It is noted that a border of width of one pixel surrounding the processed area is lost. Figure (5.29c) shows the phase profile given by figure (5.29b) with the phase errors removed. It can be seen that the vertical error line, although considerably reduced, is still evident. This is because three of the nine sampled points are obviously much lower than the median value, with the remaining 6 data points taken from the two sides of the corrupted data line. As the 9 data points are taken from an area with a relatively large gradient, it is obvious that the median value will be that of the points on the lower edge of the slope. Hence, a slight discontinuity in the slope is still present. The median filtering method is most appropriate for filtering large deviations from a relatively flat background.

The grey level data is stored in the frame store and the individual pixels are accessed by their universal cartesian coordinates,
Figure 5.29  a) This shows the phase profile of a video-head where the processing is conducted near the extreme edges of the head which have steep gradients and thus tight fringes, b) shows a line of corrupted data crossing the phase profile due to phase errors introduced when removing the 2π steps, c) shows the same profile after data smoothing.
In order to simplify the computer processing routine that identifies the $3 \times 3$ array and arranges them in ascending order of grey level, it was decided to establish a formula that would convert the position of each individual pixel of the array from its universal cartesian coordinates into a linear value in the range $0 - 8$ relative to its position in the $3 \times 3$ array. This was done by the equation shown below:

$$H = [3 \times (y + 1 - p)] + (x + l - q),$$

where $q$ & $p$ are the universal cartesian coordinates of the point at the centre of the $3 \times 3$ data array. The $x$ & $y$ values are the cartesian coordinates of the specific point within the array. The value $H$ will then be the position of the specific point in the range $0 - 8$ relative to its position in the array. This format greatly simplifies the computer routine which arranges the pixels of the array in ascending order, and replots the median value in the mid-point of the array.

Figure (5.30a) shows the 3-D plots of the phase profiles with the errors present, and figure (5.30b) shows it with the phase errors greatly reduced.

5.6.4 Results of the wear on a video-head.

In order to measure the wear on a video head, the head was replaced in a video-recorder and abrasive tape was passed over the face of the head for a period of 12 hours. On replacing the head in the measurement system, a different set of interference fringes were obtained, which corresponds to the difference in the surface topology, see figure (5.31a). The processed phase profile of the worn head is shown in figure (5.31b) with figure (5.31c) giving the same result with the $2\pi$ steps removed. Figures (5.32a) and (5.32b) show two 3-D perspectives of the surface profile of the worn head. These can be compared to the 3-D representations of the profiles of the unworn heads shown in figure (5.27a).

The two phase profiles, of the unworn and worn heads, can now be subtracted to yield the difference, or wear of the heads. Figure (5.33a) shows the phase difference between the two with $2\pi$ steps, and figure
Figure 5.30  a) 3-Dimensional plot of figure 5.29b clearly showing the line of corrupt data, b) the same profile after 3×3 data smoothing. The error has been reduced although not completely removed.
Figure 5.31 a), b) & c) show the various stages in measuring the phase profile of the head shown in figure 5.26 after 12 hours of wear by an abrasive tape.
Figure 5.32 Two different 3-D perspectives of the phase profile of the worn head given in figure 5.31.
Figure 5.33  a) Difference in phase profiles of head before and after wear showing $2\pi$ steps, b) steps removed to yield a continuous phase profile of wear.
(5.33b) shows the difference with the steps removed. Figure (5.34) shows the four 3-D perspectives of the wear. These figures clearly show the unworn, reference edges of the head surface, which are represented as flat areas. The change in shape has occurred in the central 2/3 of the head surface.

It is interesting to note the asymmetrical nature of the wear in this situation. The bottom edge of the head as viewed has undergone greater wear than the top edge, and the left lobe of the head has suffered a greater depth of wear with a steeper edge. This effect was observed on several different profiled heads. It is noted that the bottom-left corner of the head forms the trailing edge of both the rotation of the head across the face of the tape, left to right as viewed, and the trailing edge with regard the transport of the tape across the face of the head, top to bottom. It is suggested that the trailing edge suffers greater wear, as sharp particles and other debris of the wear scraped off the tape by the leading edge of the head accumulates throughout the length of the scan, and promotes grinding abrasion wear in the trailing edge region.

Due to the absence of any measurements of this type on actual video-recording heads in the literature, it is difficult to compare these results with those of other researchers.

The worn head was then removed and worn for a further 36 hours by a less abrasive tape. The results of this wear is shown in figures (5.35 a, b). It is clear that the maximum depth of the wear has not increased with proportion to time. This would suggest that the rate of wear in the first few hours in the life of the head is greatest. This would tend to confirm the findings of other researchers, see review in chapter 2. It is however noted that a different type of tape was used in the second wear test, and no account is taken of any misalignment of the replaced head in the video-recorder, where any tilt of the head could produce a different wear pattern on the surface.

As stated in the introduction, the purpose of this research is to find and develop a measurement technique that can be successfully and routinely used in the assessment of video-head wear. The results given in this chapter are primarily used to illustrate the capabilities of the
Figure 5.34 The four perspectives of the head wear pattern in 3-D including a figure for volume of material displaced.
Figure 5.35 2π-step phase profile and a 3-D perspective of the head shown in figure 5.34 after a further 36 hours of wear by a less abrasive tape.
system, and do not claim to be a comprehensive set of video-head wear measurements. A complete analysis of video-head wear under differing conditions would be a separate research project in itself.

5.6.5 Removal of defocus and tilt errors from the wear results.

As stated in sections 5.5.4.1 & 5.5.4.2, the validity of the wear measurement is dependent on the ability of the system to compensate for any misalignments in the repositioned head, after it has been removed for wear.

Figures (5.36a) and (5.36b) show the phase profiles of a video-head, with (5.36b) having a slight added tilt to the head compared to the (5.36a) profile. Figures (5.37a) and (5.37b) show the difference between these two profiles with the tilt compensating routine disabled. The tilt is clearly visible. Figures (5.38a) and (5.38b) shows the difference of the two profiles with the tilt compensated for. The noise level is shown to be below 30nm.

The curvature removal routine is demonstrated by figures (5.39a) and (5.39b). These show the curvature of a surface profile of a video-head removed. This is done by calculating the circle that passes through 3 points on the edges of the head, and hence its removal from the whole profile data, to yield a profile with flat edges.

5.6.6 Calibrating dimensions & calculating the displaced volume.

In each of the wear patterns shown earlier, a number is given which represents the volume of material removed from the head. This volume is calculated by summing the volumes occupied by each pixel of the wear. The \( z \) dimension is simply the height of the surface of the head relative to the unworn edges, in term of wavelength of light, in the area occupied by a particular pixel. The \( x \) and \( y \) values are the dimensions of the individual pixels multiplied by the correct magnification. The \( x \) and \( y \) dimensions are found by a calibration routine. This routine involves placing a two dimensional grating of known dimensions in the focal plane of the imaging
Figure 5.36 The phase profiles of a video-head with a) having a slight tilt added in the longer direction compared to that in b).
Figure 5.37 This shows the difference between figures 5.36a and 5.36b. The added tilt is clearly visible.
Figure 5.38  This is similar to figure 5.37 except the added tilt is removed by a compensating routine. The noise level is shown to be better than 30nm.
Figure 5.39 b) is the same phase profile as that shown in a) but with much of the curvature removed to give flat edges to the profile. This is done by a routine that finds the circle passing through 3 points on the edges and subtracts this curvature from the entire profile.

Figure 5.40 Image of grating with known dimensions showing cursors at specific points for calibrating dimensions of wear.
optics, in place of the head, prior to profiling. A computer routine was written which enables two cursors to be moved to specific points of the grid, figure (5.40). On the command of the computer the absolute distances between the cursors in the x and y dimensions are input. The computer is then able to count the number of pixels per unit length, and hence the dimensions of the individual pixels. These values are then used to accurately calculate the displaced volume of the wear.

5.7 Sensitivity and spatial resolution of the wear results.

The sensitivity of the system is ultimately determined by the 6-bit resolution of the video framestore. If the video-head describes part of a perfect sphere, the entire surface of the head is only covered by a single interference fringe. The measurement sensitivity is then be of the order of 1nm. But as is shown, the heads are typically covered by many interference fringes: 12 or more fringes in the particular case of worn heads. This, when averaged over 64 grey levels, gives a sensitivity of about 10 nm. On subtraction of two sets of data, this error is doubled.

Other factors which could degrade the measurement accuracy are such things as miscalibration of the phase shifter, electronic noise and nonlinearities in the detector, quantization and averaging errors, aberrations in the optics and irregularities in the reference mirror, vibrations and air turbulence. The last two effects are minimised by placing the apparatus on a vibration-isolated table, enclosing the laser beams paths, and sampling the data rapidly. Miscalibration of the phase shifter effects can be removed by correct calibration of the system. Errors due to the aberrations of the optics and the reference mirror are conveniently removed by virtue of the subtraction of two sets of data to yield the result. The 3 by 3 averaging algorithm used to calculate the phase profile is very good at reducing the effects of non-linearities in the phase-shifter motion.

The probable sensitivity of the system for the type of measurements shown in this chapter is nearer 30nm. This is shown up as the noise level, on subtracting two phase profiles of the same head, but with a
slight tilt between them, figure (5.38a).

The magnification of the optical imaging system is about $\times 3.5$. The dimensions of the CCD detector elements are $20\mu m \times 20\mu m$. This then gives a spatial resolution of the system of about $6\mu m$.

5.8 Conclusions.

In this chapter a system has been described which was developed and successfully used for the measurement of the wear of video-recorder heads. This optical non-contact system was designed to overcome the problems associated with measuring the change in the surface topology of highly curved surfaces, that other commercially available systems suffer from.

The system has been shown to have a height measurement sensitivity of better than 30nm, limited mainly by the 6-bit resolution of the video-frame store. The system is able to view the entire area of the curved head surface, $3mm \times 0.2mm$, with a spatial resolution of better than $6\mu m$.

Results of head wear given in this chapter clearly demonstrate the significance of this unique measurement system with regard to the study of video head/tape interactions, and is believed to be a major development in this field of research.
References.


6.1 Introduction.

The phase-shifting interferometer developed primarily for the measurement of video-head wear (chapter 5) was, with a little modification, found to be very useful for the characterisation of the surface topography of a range of other objects and surface structures. These include the laser-induced ripple structures formed on resolidified metals, dielectrics and semi-conductors. In this chapter results are presented which show the topography of the laser-induced periodic surface structures formed upon samples of germanium, which had been irradiated at a wavelength of 248nm, and fused quartz irradiated at 10.6μm. A slight modification to a theory which models the development of the surface contour during resolidification of a molten surface can be successfully employed to explain the observations on germanium. The profiles obtained on fused quartz can be satisfactorily explained by including the involvement of surface plasma oscillations.

The other application of the surface profiling system is the accurate measurement of depth profiles of craters on surfaces of samples produced by secondary ion mass spectrometry (SIMS). This is a well-established surface analysis technique used for the determination of the chemical composition of surfaces. In this chapter the modifications to the optical system required for imaging the SIMS craters are described, together with results showing the depth and profile of a typical crater on the surface of a silicon wafer.

Finally, the profiling system was again used in the configuration it was designed for, namely the measurement of the topology of highly curved surfaces. Results are given which show the surface detail of a 2mm square area of a 1 inch diameter steel-ball bearing to a sensitivity better than 10nm.
6.1.1 Laser induced ripple structures.

Many authors, as a result of performing investigations into the interaction of intense laser beams with solids, have noted the formation of well defined periodic damage patterns\textsuperscript{1-8}. It is generally accepted, as first proposed by Emmony et al\textsuperscript{2}, that the cause of these structures is interference between the incident laser light and some form of induced 'surface wave'. This results in an inhomogeneous deposition of energy at the sample surface. Attempts have been made to formally model the mechanism responsible for the formation of these laser induced ripples, notably by Emel'yanov et al\textsuperscript{7}, Temple and Soileau\textsuperscript{8}, and in particular by Sipe et al\textsuperscript{9,10,11}.

The theory developed by Sipe et al\textsuperscript{9} is, to the author's knowledge, the most rigorous and comprehensive attempt to date at developing a model of ripple formation.

It is inherent in this and all other theories of laser-induced ripple formation, that in order to form such structures there must be a periodic temperature profile induced on the surface. Previous work by Kerr et al\textsuperscript{12} has provided direct confirmation of this.

A model proposed by Emmony et al\textsuperscript{13} can be used to explain why, if such periodic heating leads to periodic melting of a surface, a permanently raised ripple pattern should develop upon resolidification. As its basis this theory assumes that all molten solids will, during the cooling process, be subjected to a redistribution so that potentially at least a new surface contour may be formed. This redistribution occurs as it is required by matter conservation if the liquid and solid phase densities for the material are different. It is therefore a basic requirement of this model that for the development of ripples, the liquid and solid phase densities must differ. If this is not so, then the development of a surface ripple is not possible by this method.

By making the assumptions that the liquid phase of the material has a very high surface tension and that thermal conduction at the liquid-solid interface is the dominant method of cooling the molten liquid, Emmony et al were successful in modelling this redistribution of matter for
periodically molten germanium samples. Germanium suffers an approximate 8% change in density as it resolidifies. Using algorithms developed in their paper it is possible to model the development of such profiles.

This chapter details some of the numerous ripple contours obtained on samples of germanium and fused quartz, and attempts have been made to explain their nature using the currently accepted theories.

6.1.2 Experimental work.

The experimental aspects of this work fall into two main categories. i) that which deals with the generation of the laser induced ripples, to which I am indebted to N.C.Kerr for, without whose efforts in generating the many samples that were available for measurement, this work would otherwise not be possible, and ii) the measurement system which was used to produce topographical data on these ripples, and hence allow a discussion and the proposal of models that could explain the mechanics of the ripple formation.

6.1.2a Ripple formation.

The generation of laser induced ripples has been extensively studied using the damage facility at Loughborough University\(^1\). This facility centres around a Lambda Physik model EMG 200 excimer laser which is operated with a krypton and fluorine gas mixture and produces nominal 1J pulses of 20ns duration at a wavelength of 248nm. A Laser Applications TEA CO\(_2\) laser is also available which gives a total pulse output of 5J at a wavelength of 10.6\(\mu\)m. The output consists of a 100ns pulse followed by a long tail associated with N\(_2\) in the laser gas mixture.

The surfaces used for ripple formation were polished germanium mirrors and fused quartz plates, which prior to irradiation had no special treatment except a wipe using methanol to remove dust etc.

Previous work\(^{14,15,16,17}\) had led colleagues at Loughborough to develop a tried and tested 'recipe' of laser fluence, angle of incidence and incident polarisation which enabled the rapid generation of ripples upon surfaces of germanium using the excimer laser and fused quartz using
the CO₂ laser. Full details of the experimental arrangement of apparatus used in the formation of ripples can be found in those references.

6.1.2b Ripple topography measurement.

The only aspect of the system's hardware to have changed significantly from that used in chapter 5 is the imaging optics, as seen in figure (6.1). The system has a height resolution of approximately 5 nm and a lateral resolution of approximately 1 μm. To obtain the required image resolution, a compound microscope arrangement was used which consisted of a well-corrected 20x long working distance microscope objective coupled to a large aperture camera lens which served both as an eye-piece for the system and to provide an additional 2x magnification. An identical lens to that in the test arm of the interferometer was placed in the reference arm to match the curvature of the reflected wavefront.

6.1.3 Results.

Profiles were obtained for various regions upon a rippled germanium surface, see figure (6.2a). These ripples are typical of many obtained but of particular value because they show two spacings which are spatially very close to each other in the same image. The ripples were produced by multiple p-polarised pulses from an excimer laser, operating at a wavelength of 248 nm, incident at an angle of 75°. The left-hand side of figure (6.2a) corresponds to the edge of the rippled site, the right hand side corresponds to the centre of the laser interaction area. The fluence at the centre of the beam interaction area is greater than at the edge due to variation of beam intensity across the laser beam spatial profile. Figures (6.2b) and (6.2c) show the respective surface profiles for these two areas on the sample.

Figure (6.2d) shows smaller (<1 μm) ripples obtained by Emmony et al.¹³ on Ge using a wavelength of 10.6 μm.

Figure (6.3a) shows ripples obtained upon a fused quartz surface irradiated with multiple pulses of s-polarised light from a CO₂ laser operating at 10.6 μm. The pulses were incident normal to the surface. Figure (6.3b) shows a typical surface profile.
Figure 6.1
Experimental arrangement for phaseshifting interferometry.
Figure 6.2 a) Laser-induced ripples on germanium. The ripples were produced by multiple p-polarised pulses from an excimer laser operating at a wavelength of 248 nm. The pulses were incident at an angle of 75°.
Figure 6.2  b) Surface profile taken on the left of figure 6.2a.  c) Surface profile taken on the right of figure 6.2a.
Figure 6.2  d) This shows small ripples (<1μm) obtained by Emmony et al [13] on Ge using a wavelength of 10.6 μm.
Figure 6.3  a) Laser induced ripples on fused quartz. The ripples were produced by multiple s-polarised pulses from a CO$_2$ laser operating at a wavelength of 10.6 µm. The pulses were incident normal to the surface. b) Surface profile taken in the centre of figure 6.3a.
Calculations using the algorithms developed by Emmony et al.\textsuperscript{13} with the values of the liquid (5600kgm\textsuperscript{-3}) and solid (5200kgm\textsuperscript{-3}) densities of germanium\textsuperscript{14} inserted result in the theoretical profile shown in figure (6.4).

6.1.4 Discussion.

The previous results of Emmony et al.\textsuperscript{13}, figure (6.2d), clearly show the 'sombrero' contour as predicted by calculation in figure (6.4). Here the surface is assumed to have been only periodically melted in a linear array and the ratio of the sunken depth to the raised height of the profile would experimentally appear to be exactly as expected from the theory. Unfortunately these ripples were too small to be resolved by the imaging optics of the phase-shifting interferometry system. For this reason measurements were made on a germanium sample which had been especially irradiated to produce ripples with a larger spacing.

Clearly both the spacing and depth together with the general contour of the ripples formed on germanium alter over the laser interaction area, as seen from figure (6.2a). In the centre of the area where the laser fluence was higher the ripples have a larger spacing and they are greater in depth, figure (6.2c). At the edge of the laser interaction area the ripples are closer together; they are not as deep and are more sinusoidal in profile with an indication of a flat region between them, figure (6.2b).

Neither the profile of the ripples at the edge of the interaction area nor that at the centre matches the sombrero contour expected from the theory in\textsuperscript{13} for the case of germanium where the liquid density is higher than that of the solid density. The extent of the dips in the theoretical profile, figure (6.4), appears approximately equal to the height of the raised centre and so should be resolved in a real ripple pattern by the apparatus. Very careful and numerous attempts failed to detect exactly the expected sombrero contour on this sample.

The qualitative explanation for the profile seen is shown schematically in figure (6.5). Here the assumption is made that the incident laser fluence was in fact high enough to melt the germanium surface completely, figure (6.5a). Since the molten surface is fluid it
Germanium resolidification profiles.

Figure 6.4 Theoretical resolidification profiles for germanium, solid density 5200 kg m$^{-3}$ and liquid density 5600 kg m$^{-3}$.
Figure 6.5 Proposed ripple formation mechanism for a liquid surface. a) Fully molten surface. Note how the depth of melting varies following the periodic temperature profile on the surface. c) Coolest points just solidify and therefore 'pin' the remaining molten material. d) Surface ripples to conserve mass which is upwards for germanium due to decrease of density on solidification.
flows and because of surface tension pulls down as shown to minimise its energy, figure (6.5b). However, the solid-liquid interface still retains the convoluted sinusoidal contour produced by the inhomogeneous deposition of energy at the surface, as implied by the theory of Sipe et al.\textsuperscript{9} Cooling of the molten material is predominantly by thermal conduction to the solid underneath, and so a solidification front proceeds upwards which retains the convoluted interface profile, figure (6.5c). As the resolidifying front reaches the surface of the germanium the front becomes pinned, thus effectively breaking the surface into isolated molten pools. Because of the slowly varying nature of the solidification front's profile and since the molten germanium wets the solid germanium surface (no contact angle exists) the regions between the molten pools can be wide and flat. As the resolidification proceeds in the pools and in order to conserve matter whilst allowing for the required change in density the surface of these pools is raised. This eventually results in the final surface contour, figure (6.5d).

The percentage change in density of the germanium as it resolidifies implies an approximate 2\% change in linear dimensions. Since the ripples are approximately 50nm in height this implies that during the laser pulse the melt extended to a depth of about 2.5\(\mu\)m. Calculations, following the algorithms developed in Carslaw and Jaeger\textsuperscript{21} and Ready\textsuperscript{22}, to a first approximation, imply that this is not unreasonable. Following Ready figure (6.6) shows the theoretical variation of temperature with depth below the surface for a germanium sample irradiated with a pulse from an excimer laser. Values for pulse energy and duration were used in the calculation to represent the conditions under which the original ripples were formed.

For the pulse duration used it can be assumed that all of the absorbed excimer laser light is converted immediately to heat since the time to convert the energy of electron-hole pair formation to heat is on the picosecond time scale. As a result the heating process can be treated by the usual heat-diffusion equation with the laser pulse as a heat source term. In addition, the skin depth of the laser pulse is much less than the depth to which heat can diffuse during the period of the pulse. It is therefore possible to assume a surface heating source where the optical and thermal properties of germanium are assumed to be independent of
Figure 6.6 Temperature versus depth for a germanium surface irradiated by a pulse from an excimer laser operating at 248 nm. The total pulse energy is 25 mJ, pulse width 25 ns, with a measured Gaussian spot radius of 1 mm, total absorption (i.e. 1 - R) of 0.35, and germanium density 5200 kg m\(^{-3}\), specific heat capacity 309.8 J kg\(^{-1}\) K\(^{-1}\), thermal conductivity 58.61 W m\(^{-1}\) K\(^{-1}\). The melting point of germanium is approximately 936°C. All values taken from [32].
temperature. The problem of laser melting has been addressed in more detail by numerous authors\textsuperscript{23,24,25} but it felt that to a first approximation the calculations in this chapter give a fair indication of the melting of the germanium sample, although predicted surface temperatures up to 10,000 °C are clearly unrealistic.

In the central region it is suggested that the ripples are deeper because these areas have simply melted more deeply than others.

Private communications\textsuperscript{23} lead to the conclusion that the magnitude of the decrease in density suffered by fused quartz as it melts is insignificant compared to that of the increase in density experienced by germanium. It is therefore hard to foresee a situation where the arguments put forward by Emmony et al in\textsuperscript{13} can be used to explain the development of the ripples seen to form on fused quartz. The fact that the fused quartz ripples are in fact greater in depth than those on germanium still further lead to the conclusion that a different process is at work in producing this surface contour.

In attempting to explain this process it was noted by Kerr\textsuperscript{15} that ripples did not form on fused quartz except when the laser fluence was in a very narrow band at which a plasma was generated in the air above the surface. At lower fluence no ripples formed for any number of laser pulses. Similarly at higher fluence the surface was ablated by the laser pulses. This observation for the necessity of a plasma is very similar to the observation of Isenor\textsuperscript{24}. Isenor observes that ripples do not form on Ni\textsubscript{x}P\textsubscript{1-x} in a vacuum under identical circumstances which do produce them in air. It is noted that the absence of air would prevent the creation of a plasma at the surface. Isenor suggests that plasma oscillations may be involved in the formation of the ripples.

During a laser pulse a critical density of electrons may exist close to the sample surface. At normal incidence the laser E-field would excite oscillations parallel to the surface. Surface irregularity would lead to laser light being scattered tangentially to the surface which could produce a standing wave pattern. The combined incident, reflected and diffracted waves could combine to give an E-field distribution capable of driving localised plasma oscillations. These could result in erosion of
material to form valleys at the surface or a displacement of material to form ridges. It is felt that a process such as this could explain the sinusoidal profiles measured for fused quartz, figure (6.3b).

Konov et al.\textsuperscript{25} have proposed a similar evaporative mechanism of ripple formation on a quartz target for the case of spatially modulated radiation. The energy for the formation of ripples is once again observed to correspond to that at which a jet of vapour is produced which indicates that there is a correlation between the formation of the surface relief on quartz and its vaporization. It is shown that a possible mechanism for the formation of ripples is the expulsion of vaporized material or the ablation of the liquid phase due to the vapour pressure gradient. In both cases the vaporization of the surface is a necessary condition.

6.1.5 Conclusions.

The orientation and spacing of all of the observed germanium ripple patterns can be explained in terms of the currently accepted theory for the inhomogeneous deposition of energy at a surface during ripple formation as developed by Sipe et al.

The theory of resolidification developed by Emmony et al can explain the general profile of the germanium ripples which form depending upon the relative densities of the material's solid and liquid phases if the surface is only periodically melted in a linear array.

If the surface is uniformly molten then the addition of a process involving pinning can be used to explain the different contours which develop. The exact profile which forms and its depth would seem to be dependent upon the fluence at the material surface during pulsed laser irradiation. Increased fluence would appear to lead to ripples with a greater size.

In order to explain the ripples which form on fused quartz the theory fails since this material suffers no significant density change on melting. It is felt that plasma oscillations may however be involved in the formation of these ripples, as proposed by Isenor.
6.2 Profiling of SIMS sputter craters.

Secondary ion mass spectroscopy, SIMS, is a well established method in surface diffusion processes for determination of the chemical composition of surfaces.

In SIMS a beam of low energy ions is used to bombard the sample; sufficient energy is transferred to the surface for atoms to be sputtered into the vacuum where the ionised material is analysed using a mass spectrometer. Electropositive elements produce positive secondary ions whilst electronegative elements give negative secondary ions. The technique is capable of detecting all elements and isotopes. The detection limits in most semiconductor materials are in the range $10^{13}$ to $10^{16}$ atoms cm$^{-3}$. In the depth profiling mode, it can be used to obtain concentration profiles by continuously eroding the sample surface and monitoring the sputtered ions as a function of depth (up to a few microns). A chemical profile of the surface is usually obtained by sputtering a rectangular crater in the sample surface. However, a serious problem is that of controlling the crater depth and profile shape, preferably in situ, while sputtering the surface.

The surface analysis group of Loughborough Consultants Ltd. use SIMS to perform a variety of analytical surface measurement on a range of materials. At present the crater depth is measured using a Vickers Instruments M41 interference microscope. This is a modified optical microscope employing a set of interference microscope objectives to give a range of magnifications. This method is reasonably accurate in determining the depth of SIMS craters which are a few microns deep. The craters typically have fairly shallow edges, which allows the fringes to be counted, to give a measure of the depth, as shown in figure (6.7a). Figure (6.7b) however, shows the interference image of a shallow crater, which illustrates the limitation of this method of measurement. The fringe separation corresponds to half the wavelength of light used, which in the Vickers system is 541nm. It is clear that it is difficult to determine accurately the deviation of the fringes to a value better than $1/10$th of a fringe. This corresponds to a measurement sensitivity of about 30nm. It is also fairly difficult to determine the general profile of the crater by this technique.
Figure 6.7 The depth of surface craters on silicon produced by the SIMS technique, measured using a Vickers Instruments M41 interference microscope. a) Shows a shallow crater and b) shows a deep crater.
Another method suggested by Makosch and Drollinger\textsuperscript{26} for measuring the profile of SIMS craters is by the use of a specially developed scanning differential ac interferometer. This method employs two probe laser beams which scan the sample surface and measure the optical path difference between the reflected light beams. The two laser spots are typically separated by about 400μm, with the one spot serving as a reference point scans over a flat non-sputtered part of the surface. Simultaneously, the other spot scans across the crater. This method offers potential high sensitivity measurements (<1nm). It does however, assume that the 'flat non-sputtered' part of the surface is indeed flat. It is also only capable of producing line profiles of the crater.

Another traditional method of profiling SIMS craters is by the use of a mechanical stylus instrument. This however, suffers from the constraints of only producing line profiles. There are added problems associated with dragging a sharp tip across the surface of a soft sample.

The phase-shifting interferometer offers the capability of non-contact, high sensitivity measurements (<10nm), together with the ability of profiling the entire crater and the surrounding surface simultaneously with high speed.

\textbf{6.2.1 Experimental.}

The principle of operation of the phase-shifting interferometer together with details of the various hardware components are described in earlier parts of this thesis, so will not be repeated here. The optical arrangement has been slightly modified, to that used for profiling the laser-induced ripple structures, in order to image a larger surface area of the sample with lower magnification. Figure (6.8) gives the optical arrangement used.

The interferometer is basically a Michelson Interferometer, with the silicon wafer containing the crater in the test arm, and a high quality (better than $\lambda/20$) plane mirror in the reference arm. The reference mirror is mounted on a PZT transducer, to provide the necessary phase shifts. A cube-beam splitter is used to split the HeNe 633nm wavelength laser beam into two components. On reflection and interference the recombined

222
Figure 6.8 Optical arrangement of the modified Michelson interferometer used to measure the shape and depth of the SIMS craters.
wavefront contains information related to the relative heights of the test sample's surface. This information is then imaged onto the surface of a CCD video camera. The video image is stored in a video frame-grabber, and the information processed by a PC computer. The imaging objective used is a standard 32mm focal length, flat mounted refracting microscope objective. It is connected to the video-camera by a 160mm tube. The stated working distance for this lens is 35mm, which is sufficient distance to image through the 15mm cube beam splitter. The magnification of the lens at these distances is x4. The typical lateral dimensions of the SIMS craters are normally set to be in the region of 25μm - 400μm². The craters measured in this work were about 400μm square, so a magnification of four would produce an image about 1.6mm square on the 10mm square CCD detector element.

The test sample and the reference mirror were adjusted so as to obtain a single interference fringe across the field of view to maximise the sensitivity. They were also set at a slight angle to the incoming laser beam, so part of the reflected beams did not pass back into the laser and cause cavity detuning. This method is an alternate to using a polarised beam-splitter and quarter-waveplates in the two arms of the interferometer, as used in the system described in chapter (5).

6.2.2 Results.

The results given in this section are measurements showing the depth and profile of a SIMS crater on a piece of silicon wafer. The depth of the crater is shown to be approximately 90nm. One of the four phase shifted interference images of the crater showing a single fringe across the field of view is shown in figure (6.9). On processing the phase shifted interference data in and around the area of the crater, a phase profile of the crater is found, as seen in figure (6.10a). After smoothing the data (see section 5.6.3) to remove phase errors caused by scatter from dust particles and ablation debris on the surface, the smoothed phase profile is given by figure (6.10b). The phase profile is a direct representation of the surface height. This information is stored in the computer memory, and can be accessed and displayed in various forms. Figure (6.11) shows a 3-D representation of a cross section through the crater. This clearly illustrates the relatively shallow edges of the crater, whilst giving a good overview of the shape of the crater. In addition to 3-D profiles, line
Figure 6.9 One of the four phase-shifted interference images of a crater with the sample surface parallel to the reference mirror giving a single interference fringe across the field of view.
Figure 6.10 Phase profile of the sample in the area of the crater obtained by processing the set of four phase-shifted interference images. The grey levels indicate the relative surface heights, with black indicating low and white indicating high. a) shows the phase profile with phase errors due to light scatter off the edges of the crater and ablation debris, b) shows the same result after data smoothing.
Figure 6.11  This shows a 3-Dimensional plot of a cross section through the crater given in figure 6.10.
profiles can be instantly obtained across any part of the crater, both in the x and y dimension, simply by moving a cursor across the monitor screen. Figures (6.12) & (6.13) are line profiles across the crater in the horizontal and vertical dimensions. These give a very accurate measure of the depth of the crater at any one point.

Figure (6.14) shows the interference micrograph of the same crater as given by the present method for determining its shape and depth. It serves to corroborate the value obtained for the depth of the crater, as the fringes appear to deviate approximately 1/3 of a fringe which corresponds to about 90nm. Although as stated earlier, this method has an associated high degree of uncertainty in its measurement of shallow craters.

It is very interesting to note that by studying the two line profiles in figures (6.12) & (6.13), there is clearly an area to the side of the crater that has had material removed from the surface to a maximum depth of about 40nm in a fairly irregular fashion. This is also visible in figure (6.9) as a lighter shaded area to the side of the rectangular crater, although this figure does not give a quantitative value for the surface height.

The ion beam used to produce a crater by the SIMS system on the silicon surface is that of positive molecular oxygen ions \((O_2^+)\). The direction of this ionised stream of atoms is controlled by a system of electrodes which perform the function of accelerator and electron-lens, so the beam is brought to pinpoint focus at the target surface. A system of magnetic coils is used to deflect the ion beam on its way, so as to produce a raster scan of the surface. This system is very similar to that employed in television sets to control the electron beam which illuminates the phosphor screen.

The ions are produced by a duo-plasmatron ion source. This is used to ionise molecular oxygen. However, not all the molecules become ionised, and the unionised molecules (neutrals) form an ambient level of oxygen in the vacuum chamber. On collision of the ion beam with these neutrals, they acquire sufficient momentum to be propelled towards the target. As these neutrals carry no electric charge, they are not deflected.
Figure 6.12 This shows a vertical line profile through the centre of the crater seen in figure 6.10.
Figure 6.13 This shows a horizontal line profile through the centre of the crater seen in figure 6.10.
Figure 6.14 Interference micrograph of the 90nm deep crater seen in figure 6.10, using the Vickers interference microscope method.
by the beam steering system. Hence, they impinge onto the surface of the sample in an irregular fashion, and act to ablate material from the sample surface. In order to extract the positively charged secondary ions emitted from the sample for analysis, the surface is held at a slight positively charged potential. This acts to slightly deflect the incident positive ion beam, which strikes the surface at an angle. The neutral molecules are not affected, and they follow the original beam trajectory and strike the sample at a point slightly displaced from the crater site, as seen in the figures shown.

6.2.3 Conclusion.

A comparison of the results from the existing method employed by the surface analysis group of Loughborough Consultants to measure the depth of shallow (<300 nm) SIMS craters, see figure (6.14), to those obtained using the phase-shifting interferometer, shown in this chapter, demonstrate the superiority of the latter technique. The system developed has the advantage of being fast, non-contact and highly sensitive with the output data being presented in a range of ways, depending on the requirements.

It is clear that without the higher sensitivity available, fine details such as the secondary ablated area, caused by the neutral molecules, would not have been detected.

The ability to accurately measure the depth and profile of these secondary ablated areas has generated a great interest from the surface analysis group at Loughborough Consultants, and there is a possibility of conducting further experiments in the future to study these secondary craters29.

6.3 Surface profile measurements of steel ball-bearings.

Steel ball bearings have many uses in a large range of engineering industries. They are primarily used as anti-friction bearings. In modern engineering, much attention is paid to the tolerances of their specified diameter, roundness, surface roughness, hardness and other
There is a comprehensive grading system administered by the Anti-Friction Bearing Manufacturers Association (A.F.B.M.A.) for this purpose.

The traditional method for measuring several surface parameters of steel ball-bearings is by use of stylus instruments, where the ball bearing is rotated whilst the stylus tip remains in contact with the surface of the ball. A typical example of such an instrument is a 'Talyrond', manufactured by Rank Taylor Hobson in Leicester. An example of a certificate of conformity to an (A.F.B.M.A.) standard for a 30mm ball bearing is shown in figure (6.15).30

The phase shifting interferometer can be used as an alternative method for measuring the surface topography of ball bearings, this non-contact method is particularly necessary on highly polished soft surfaces which might be scratched by the action of a stylus. The technique employs a Twyman-Green optical arrangement, which is used to match the shape of the incident wavefront to the curvature of the ball, as described in section (5.3.1).

The system used is similar in some respects to the Hilger & Watts Sphericity Interferometer31 (figure 6.16), an optical instrument which also uses the effect of the interference of light to compare the curvature of the reflecting spherical surface under test, with a curved reference surface.

In the Sphericity Interferometer, the light is reflected between the ball and a curved mirror in the instrument, and the resulting pattern can be related to curvature variations on the surface of the ball. This system however, only gives qualitative information, and cannot determine the polarity of surface deviations.

6.3.1 Results.

The interferometer was set up so as to nearly match the curvature of the ball bearing to the incident wavefront. This resulted in about 3 interference fringes across the field of view (2mm square), as seen in
CERTIFICATE OF CONFORMITY

BALL REF: Bally 1178/75-97/97

It is hereby certified that the above Tungsten Carbide Ball comprises:

94% Tungsten Carbide
6% Cobalt

Calibration by DCS approved Johansson S27W90E comparator
Talyrond 73 confirms:

Diameter and sphericity to within AERMAC "C"
\( +0.00025" \)
\( -0.00025" \)

Figure 6.15  An example of a certificate of conformity to an (AFEMA) standard for a 30mm ball-bearing.

234
Figure 6.16 Hilger & Watts sphericity interferometer [31].
figure (6.17a). Had this near matching of curvatures not been performed and instead a plane wavefront incident on the surface of the ball bearing, ignoring the problems associated with recollimating the scattered light, then a 1 inch diameter ball bearing would produce 1600 interference fringes across the same field of view. This curvature would simply swamp any small surface information, that needed to be measured.

Figure (6.17b) shows the interference fringes having been processed to produce a phase profile of the ball bearing surface, with 2π discontinuities evident, each step of which corresponds to \( \lambda/2 \). These discontinuities are due to the nature of arc tangent calculations. These are then removed (figure 6.17c) to produce a single fringe, or grey level gradient of the surface, which represents surface height. This information can then be displayed in a variety of ways. Figure (6.17d) shows the 64 grey levels banded into 12 levels, to give a contour map feel to the information. It is also possible to produce line profiles across any part of the image. Figure (6.18a) gives a line profile of the relative surface height across the middle of the field of view of the ball bearing.

By slightly adjusting the focus of the ball bearing to match the curvatures even more than as shown above, it is possible to virtually remove the curvature, and view the surface information with maximum sensitivity. Figure (6.18b) shows the same line profile as figure (6.18a) but with the curvature almost removed. The sensitivity of the measurement is better than 10nm.

6.3.2 Conclusions.

Although the stylus instruments give similar sensitivities, it has been shown that phase shifting interferometry is an alternative method, which is non-contact, fast and very simple to align, and can give height information over areas of the surface as opposed to single line profiles.

It should be noted that the accuracy of the phase information obtained is dependent on any aberrations or distortions of the imaging optics and the reference mirror of the interferometer. The optics used in this interferometer are however, chosen to be of high quality and used in the correct configurations so as to minimise any aberrations.
Figure 6.17 a) One of the set of four phase-shifted interference fringes obtained from the surface of a one inch diameter steel ball-bearing. b) the phase profile of the surface of the ball after processing the interference fringes showing $2\pi$ steps, c) the same phase profile with the steps removed and d) the 64 grey levels given in c) banded into 12 levels. This gives greater sensitivity to the eye to detect irregularities on the surface.

237
Figure 6.18a This gives a line profile through the middle of the field of view as seen in figure 6.17c. This illustrates the curvature of the surface of the ball-bearing.
Figure 6.18b This shows the same line profile to that given in Figure 6.18a but with much of the curvature removed for greater sensitivity of surface measurement. The ball-bearing in this figure shows a maximum surface irregularity of about 10 nm.
References.


[23] Private communication Thermal Syndicate Wallsend Newcastle upon Tyne.


[27] Loughborough Consultants ltd. Surface Analysis Group. SIMS special publication. AVS Loughborough University.


CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK.

This final chapter gives an overall conclusion together with some minor recommendations for possible improvements and future work.

There are two main aspects to the work covered in this thesis. The first deals with the characterisation of the ultra-violet laser-induced fluorescence of various doped glasses, with the aim of using them as uv-to-visible image converters. These glasses form the basis of a video-based system which was developed for the spatial-profiling of high-powered ultra-violet laser beams. The system is presently in routine use on Sprite, Europe's largest power ultra-violet laser, at the Rutherford Appleton laboratory. There appear to be no obvious ways of improving the system apart from introducing minor cosmetic changes to the software for a more sophisticated graphical display and representation of the results. At the start of this work, the system was considered to be a major improvement on other techniques used for beam profiling. Today, there do exist, at considerable expense, commercial systems which employ the same principle to that covered by the work in this thesis.

The other part of this thesis deals with the development of novel optical surface profiling instruments with the specific aim of measuring and characterising the wear patterns on video-recording heads. The first system presents a new approach to optical surface profiling using differential phase quadrature. The system is shown to be highly resistant to environmental disturbances and to have a measurement sensitivity of about 1nm. This figure could be improved by upgrading the 8-bit analogue-to-digital converter to 10 or 12 bits. The spatial resolution of the arrangement is of the order of 10μm, limited only by the spot size of the focused laser beam. Higher quality optics and larger diameter beams would significantly improve this figure. Anti-reflection coatings on the surface of the lenses and beam-splitters used in the system would reduce unwanted reflections and enable samples to be profiled with reflectivities lower than 10%. These all, however, only offer relatively minor improvements to the system.
A limitation of the system is the small operating range of its depth of focus (~3 μm), which necessitates careful alignment of the sample and the use of high precision translation stages. A significant improvement to the scanning probe interferometer would arise from a defocus correction system. It is suggested that the arrangement detailed below could be incorporated into the measurement system at a later stage.

7.1 Defocus correction system.

A possible method to overcome the defocusing of the sample is to employ a technique based on a system devised by Bricot et al. (1976). As illustrated in figure (7.1a), it uses a cylindrical lens and a quadrant photodetector to produce an error signal when the sample begins to move out of focus. The focal lengths and optical-path lengths between the sample and the objective lens, and those between the second focusing lens and the cylindrical lens are adjusted so that, when the laser beam is properly focused on the sample, the reflected light is also focused on the the cylindrical lens. The cylindrical lens then has no effect on the reflected beam and a circular spot reaches the quadrant photodetector. On the other hand, if the focus is incorrect, the reflected beam will not be focused on the cylindrical lens and an elliptical beam will fall onto the quadrant photodetector as indicated in figure (7.1b). The major axis of the ellipse will be vertical if the sample is too far from the lens and horizontal if the sample is too close. By summing the output of the photodiodes $P_1$ & $P_3$ and subtracting the sum of the output of photodiodes $P_2$ & $P_4$, an error signal may be produced. A servo loop connected to the coil of a loudspeaker would use the error signal to control the position of the sample with respect to the objective lens. This technique is commonly used in compact audio disc and video disc players to maintain the lens-to-medium distance within the depth of focus (~1 μm) of an objective lens.

The other optical system dealt with in this thesis is a phase-shifting interferometer based on a Twyman-Green optical arrangement. This configuration was adopted to overcome the problems involved with processing interference fringes that are produced by interfering a plane wavefront with light reflected off a highly curved surface, such as a video recording head. This problem limits the use of commercial phase-shifting instruments, which are mainly Mirau-interferometer based. The developed
Figure 7.1  a) Schematic diagram for the proposed method of monitoring, and correcting for, the defocusing of the sample, b) top view of the quadrant split photo-detector.
system was used to produce, to the author's knowledge, the first direct measurements of video-head wear patterns, and is now used at the 3M (UK) Ltd research laboratories, Swansea, for the assessment of head-wear and the measurement of the abrasivity of newly developed magnetic recording tape.

The system has a measurement sensitivity of about 10nm, again limited mainly by the electronic resolution (grey levels) of the video frame store used. This figure, as well as the 6µm spatial resolution of the system, could be improved by the use of equipment with better specifications. The system has been shown to accurately measure wear patterns on new VHS video heads to a depth greater than 2µm. Greater amounts of wear have been shown to produce higher surface slopes, as the shape of the curved head begins to depart from a spherical form. This produces narrow fringes and presents measurement difficulties, as phase ambiguities are created when two adjacent pixels measure phases that are separated by a value greater than \( \pi \). A method proposed to extend the useful measurement range of the system and provide the basis of future work, is given below

### 7.2 Two-wavelength phase-shifting interferometry

The \( \pi \) phase difference limit between adjacent pixels corresponds to one-half wave in the optical path difference. An obvious method to overcome this limitation and measure steeper gradients is to use a laser source with a longer wavelength. This, however, causes the error in the measurement, which is proportional to the wavelength used, to increase.

A method of reconstructing a wavefront, which employs two different wavelengths, has been proposed by several workers\(^2,3,4\). In this method the surface is illuminated by light of two different wavelengths \( \lambda_a \) and \( \lambda_b \). These may be produced from two different lasers or from the phase modulation of a single laser. Two different sets of phase information will thus be obtained. Subtracting these two sets of data will yield a set of phase information corresponding to an equivalent wavelength \( \lambda_{eq} \), which Wyant\(^5\) has shown \( \lambda_{eq} \) to be:

\[
\lambda_{eq} = \lambda_a \cdot \lambda_b / (\lambda_a - \lambda_b)
\]

The equivalent wavelength phase \( \phi_{eq} \) is shown by Creath et al to be\(^4\):
\[ \phi_{eq}(x,y) = \frac{2\pi \cdot OPD(x,y)}{\lambda_{eq}} = \phi_a(x,y) - \phi_b(x,y), \]

This new set of phase information is the same as if the surface had been measured using the equivalent wavelength \( \lambda_{eq} \). If \( \lambda_a \) and \( \lambda_b \) are only slightly different, then \( \lambda_{eq} \) will be large. This longer-wavelength set of data can then be used to remove phase ambiguities, in one of the sets of data obtained using the shorter wavelength \( \lambda_a \) or \( \lambda_b \), by determining the correct fringe orders.

The system has the benefit of the high sensitivity and low noise level associated with measuring with a short wavelength. It is also capable of determining the profiles of surfaces with steep gradients, by removing the phase ambiguities, using the much longer equivalent wavelength. Hence, a much larger operating dynamic range is achieved.

Finally, in order to measure the wear of a video-head it is necessary for the head to be removed from the recorder, and positioned in the measurement system. It is then replaced in the recorder for further wear. A question then arises as to what effect misalignments of the replaced head has on the subsequently generated wear pattern, as it is likely that a different relative position of tape to head will produce different wear patterns. It is suggested therefore that future work might involve incorporating the measurement system described in this thesis, with a recorder dedicated to wear measurements. This would enable in-situ readings to be taken, and remove the uncertainty involved with the possible misalignment of the replaced heads.
References.


